SPARC T-SERIES

An Oracle White Paper
October 2010

Maximizing Service Uptime with Oracle's SPARC T3-1, SPARC T3-2, SPARC T3-4 and SPARC T3-1B Servers

Introduction

IT services play a critical role in helping enterprises compete effectively in today's rapidly changing global markets. Organizations of every size and type increasingly depend on information technology to execute day-to-day operations, interact with customers, and generate revenue. Current trends toward nonstop business operations place pressure on IT departments to keep services available around the clock. As a result, the importance of systems' and software applications' reliability, availability, and serviceability (RAS) capabilities continues to escalate.

Reliability, availability, and serviceability are intricately related, and each factor is equally important to achieving maximum IT service uptime. The reliability features of a system work to minimize the frequency of faults and to ensure data integrity. Availability capabilities support continuous accessibility to IT services despite system faults or error events. Serviceability mechanisms foster short service cycles for component upgrades or repair. In some cases, designing for all three RAS factors at the same time can pose a challenge. For instance, maximizing redundancy can boost system availability but redundancy also adds to component count, lowering potential reliability levels. Determined to continuously improve RAS capabilities, Oracle uses carefully engineered metrics to aid design efforts to balance and optimize the RAS levels of each new generation its servers.

Oracle's SPARC T3-1, SPARC T3-2, SPARC T3-4, and SPARC T3-1B servers provide excellent RAS characteristics that are ideal for maximizing the uptime of business-critical IT services. Highly reliable parts and a relatively low total component count minimize the risk of system errors. The ability to configure multiple PCI Express (PCIe) root complexes, processors, memory DDR3 DIMMs, and I/O cards adds to resiliency. In addition, these servers include memory page retirement (MPR) and core and thread offlining capabilities, processor cache coherency, integrated disk RAID functions, and extensive ECC hardware protection—along with redundant hot-swap disks, power supplies, and fans. SPARC T3-1, SPARC T3-2, SPARC T3-4, and SPARC T3-1B servers also minimize heat generation, helping enterprises avoid faults related to environmental factors. An onboard service processor and Integrated Lights Out Management (ILOM) software ease administration.

Taking a comprehensive top-to-bottom approach, Oracle focuses on RAS at each and every platform layer. An intimate knowledge of technology at the processor, server, virtualization, operating system, and system management layers results in tightly integrated, extremely dependable platforms. Combining the energy-efficient SPARC T3-1, SPARC T3-2, SPARC T3-4, and SPARC T3-1B server platforms with Oracle VM Server for SPARC (previously called Sun Logical Domains), Oracle Solaris 10, and Oracle management tools further improves platform stability. By taking advantage of Oracle technology, organizations can create exceptional IT solutions that minimize the total cost of ownership (TCO), optimize asset utilization, and maximize IT service uptime levels.

The Critical Role of IT Services

IT services play a critical role in organizations of every kind—from large government agencies and Fortune 500 companies to small municipalities and startup ventures. In the extreme, the very existence of some organizations depends on information technology and Web connectivity. Electronic storefronts, social networking sites, and Voice over IP (VoIP) providers offer just a few examples. Traditional businesses also rely heavily on information technology. For example, manufacturers often use IT services to communicate and complete business-to-business transactions with supply chain partners—functions critical to effective operation, production, and distribution of products. Indeed, numerous organizations depend on IT services to expand revenue opportunities, execute critical business functions, and lower the cost of operations.

As network-centric technologies evolve, the importance of IT continues to escalate. Coupled with a rapid increase in new user and device connections, compelling network services and the emergence of Web 2.0 collaboration are driving demand for new IT services. Furthermore, end users often expect constant availability, accessibility, and responsiveness from these new IT services. As a result, service providers and IT departments are under intense pressure to minimize service interruptions.

Requirements for Dependable IT Services

Today pervasive dependence on electronic transactions and communication places IT services in the critical revenue path for most organizations. System downtime can carry financial consequences on the order of hundreds of thousands of dollars per minute of lost revenue. In addition, trends toward network-facing business operations mean that the consequences of downtime can often reach beyond financial loss to include damage to brand image, lower levels of customer satisfaction, and heightened potential for security risks and compliance failure. With such heavy reliance on IT and so much at stake, downtime in many environments is now unacceptable.

Whether these responsibilities are specifically stated in service-level agreements or simply implicit, IT departments and CIOs are often held responsible for meeting availability expectations. True availability of an IT service hinges on all contributing elements—processors, system platforms, operating systems, management tools, and more. With heightened demand for highly reliable, available, and serviceable IT products and solutions across every area of the enterprise, organizations are working to minimize both planned and unplanned downtime.

Pervasive Demand for Reliability, Availability, and Serviceability

The RAS features of IT systems are more critical to a wider range of organizations than ever. RAS is important to many projects, but budgets, workload sizes, and resulting architectural approaches can vary greatly. Large vertically scalable systems are known for extensive RAS features and are suitable for some IT service deployments. However, a monolithic scaling strategy is not appropriate or cost-effective for all projects. For example, many stateless network-centric IT services achieve maximum performance by spreading the workload across multiple systems and application instances in a horizontally scaled architecture. Although techniques that distribute network-centric applications across multiple systems take some pressure off individual server availability, continuous operation of each system remains important.

The growing trend toward adoption of virtualization technologies in horizontally scaled architectures is increasing the importance of RAS features in these environments. Virtualization technology helps organizations maximize utilization of IT assets by enabling multiple applications to run on a server. Increasingly powerful multithreaded and multicore processors, along with fine-grained virtualization capabilities, help organizations continuously raise the number of applications and services hosted on a single server. By combining powerful processors with virtualization technologies, organizations can consolidate multiple horizontally scaled applications onto a minimal number of servers (Figure 1). A virtualized, consolidated, scale-out architectural approach can maximize performance and efficiency while reducing administrative efforts and acquisition costs. As with any consolidation strategy, placing more applications on each server increases the importance of system RAS capabilities.

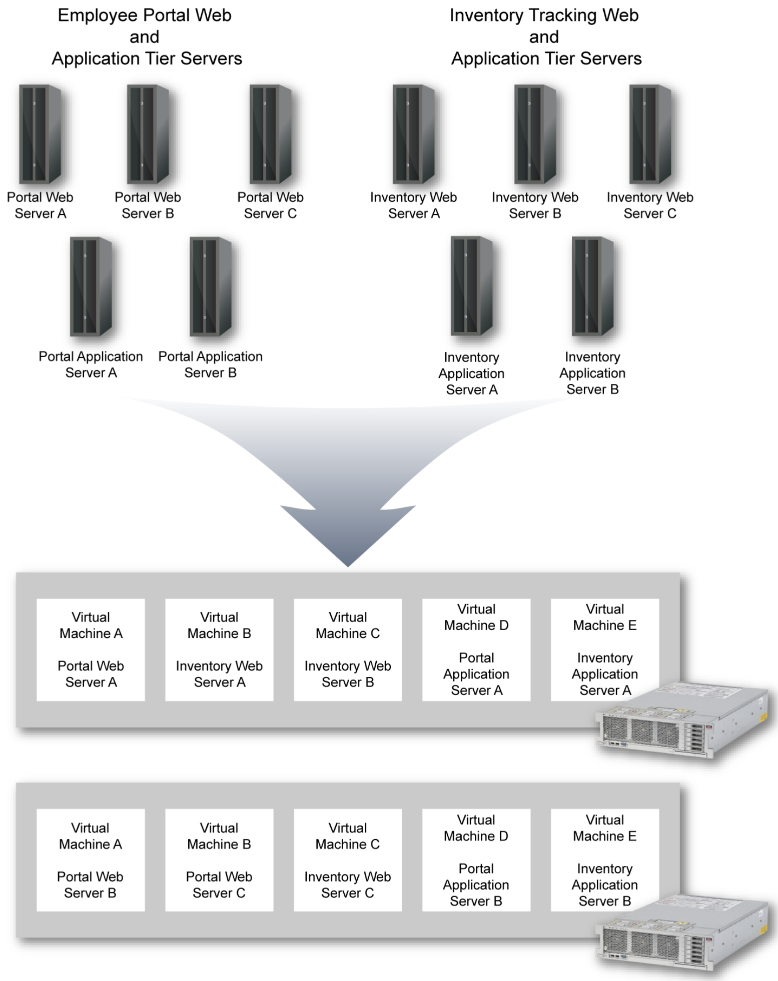


Figure 1. Virtualization technology facilitates radical server consolidation, dramatically increasing the importance of platform RAS features.

SPARC T3-1, SPARC T3-2, SPARC T3-4, and SPARC T3-1B Servers

The energy-efficient, high-performance SPARC T3-1, SPARC T3-2, SPARC T3-4, and SPARC T3-1B servers offer organizations a unique blend of vertical and horizontal scalability and demonstrate the performance and reliability traditionally reserved for midrange to higher-end systems (Figure 2). These servers represent the next wave of high-efficiency systems based on the fourth-generation of chip multithreading technology (CMT).

Figure 2. The newest members of the SPARC T-series family—Oracle's SPARC T3-1B blade server, SPARC T3-1 server, SPARC T3-2 server, and SPARC T3-4 server—include advanced RAS features to help organizations maximize IT service uptime.

SPARC T3-1, T3-2, T3-4, and T3-1B servers utilize Oracle's SPARC T3 processor, which has the capabilities of the successful UltraSPARC T2 and T2 Plus processors. These servers are ideal for meeting the demands of Web-scale businesses, creating virtualized and eco-efficient datacenters, and securing enterprise applications without sacrificing throughput. The servers incorporate the following key design elements to help organizations improve the dependability of IT services.

- **Reduced parts count** contributes to better overall stability and reliability of the platform.
- **Processor thread and core offlining** supports continuous system operations in the face of certain fault conditions.
- **Redundancy and hot-swap components** lays the foundation for system resiliency and increased serviceability.
- **Parity protection and error correction capabilities** detect and correct errors throughout the system and works to ensure data integrity.
- **Built-in RAID capabilities** foster greater system reliability and enhances data integrity.

- **System monitoring** measures temperatures, voltages, currents, and fan speeds—using a series of physical sensors—and gauges performance and systems management parameters through several logical sensors.
- **Onboard service processor and ILOM software** eases remote management and provides considerable administrative flexibility.
- **Superior energy efficiency** reduces heat dissipated into the datacenter, helping minimize susceptibility to faults due to thermal conditions.
- **Robust virtualization technology** facilitates fault isolation between applications and contributes additional RAS features to further improve IT service uptime.
- **Comprehensive fault management** provides proactive management of faults and error conditions across major elements, including the system, virtualization technology, and operating system layers

Table 1 details the features of Oracle's SPARC T3-1, T3-2, T3-4, and T3-1B servers.

TABLE 1. SPARC T3-1, T3-2, T3-4, AND T3-1B SERVER FEATURES

FEATURE	SPARC T3-1 SERVER	SPARC T3-2 SERVER	SPARC T3-4 SERVER	SPARC T3-1B BLADE SERVER
CPUs	• 16-core 1.65 GHz SPARC T3 processor	• Dual 16-core 1.65 GHz SPARC T3 processors	• Dual or Quad 16-core 1.65 GHz SPARC T3 processors	• 8- or 16-core 1.65 GHz SPARC T3 processor
Threads	• Up to 128	• Up to 256	• Up to 512	• Up to 128
Memory Capacity	• Up to 128 GB (8 GB DDR3 DIMMs)	• Up to 256 GB (8 GB DDR3 DIMMs)	• Up to 512 GB (8 GB DDR3 DIMMs)	• Up to 128 GB (8 GB DDR3 DIMMs)
Maximum Internal Disk Drives	• Up to 16 HDD (2.5-inch SAS2 300 GB disk drives), RAID 0/1, (5/5 + BBWC)	• Up to 6 HDD (2.5-inch SAS2S 300 GB disk drives), RAID 0/1	• Up to 8 HDD (2.5-inch SAS2 300 disk drives), RAID 0/1	• Up to 4 HDD (2.5-inch SAS2 300 GB disk drives), optional RAID Expansion Module
Video	• 1 VGA port	• 1 VGA port	• 1 VGA port	• 1 VGA port via dongle
Removable, Pluggable I/O	• Slimline DVD-R/W • 5 USB 2.0 ports	• Slimline DVD-R/W • 5 USB 2.0 port	• No DVD (Done via rKVMS) • 4 USB 2.0 ports	• No DVD • 3 USB 2.0 ports
PCI	• 6 x8 PCIe Gen2 slots	• 8 x8 PCIe Gen2 slots • 2 x4 PCIe Gen2 slots	• 16 EM x8 PCIe Gen2	• Optional Fabric Expansion Module for RAID 0 and 1, or 5/6 • Supports PCIe Gen2
Ethernet	• Four onboard Gigabit Ethernet ports (10/100/1000)	• Four onboard Gigabit Ethernet ports (10/100/1000)	• Four onboard Gigabit Ethernet ports (10/100/1000)	• Two onboard Gigabit Ethernet ports (10/100/1000)

	• Two 10 Gigabit Ethernet ports via XAUI combo slots (shared w/PCIe)	• Two 10 Gigabit Ethernet ports via XAUI combo slots (shared w/PCIe)	• Eight 10 Gigabit Ethernet ports via XAUI 2 QSFP Quad Connectors	• Two optional 10GB XAUI Ethernet ports expansion modules
Power supplies	• 2 hot-swappable AC 1200 W power supply units (N+1 redundancy)	• 2 hot-swappable AC 2000 W power supply units (N+1 redundancy)	• 4 hot-swappable AC 2060 W power supply units (N+N redundancy)	• Contained within SunBlade 6000 Modular System Chassis
Fans	• 6 hot-swappable fan modules, with counter-rotating fans per module, N+1 redundancy	• 6 hot-swappable fan trays, with counter-rotating fans per module, N+1 redundancy	• 5 hot-swappable fan modules, with counter-rotating fans per module, N+1 • redundancy	• Contained within SunBlade 6000 Modular System Chassis
Operating System	• Oracle Solaris 10 Update 8 + MU9, Solaris 10 Update 9	• Oracle Solaris 10 Update 8 + MU9, Solaris 10 Update 9	• Oracle Solaris 10 Update 8 + MU9, Solaris 10 Update 9	• Oracle Solaris 10 Update 8 + MU9, Solaris 10 Update 9

Designed for Reliability, Availability, and Serviceability

SPARC T3-1, T3-2, T3-4, and T3-1B servers help organizations maximize the uptime of IT services. The inclusion of redundant components and features that automate data integrity, fault isolation, and error correction improve the robustness of SPARC T3-1, T3-2, T3-4, and T3-1B servers. Online maintenance capabilities and simplified maintenance procedures help enterprises avoid the need for planned outages. Extensive fault management and self-healing capabilities reduce unplanned outages and recovery time.

The massively threaded, multisocket, system-on-a-chip (SoC) SPARC T3 processor directly contributes to the low component count of the SPARC T3-1, T3-2, T3-4, and T3-1B servers. Based on a 40 nm manufacturing process, the SPARC T3 processor combines advanced server functions on the processor die itself, including two eight-lane PCIe Generation 2 interfaces for low-latency data transfer, two 10GbE ports, a stream processing unit (SPU) on each core for wire speed cryptography and a fully pipelined floating point unit with a Fused Mul/Add (FMADD) instruction (Figure 3). The SPARC T3 processor also has multisocket capabilities, cache coherency links, and the ability to handle FB-DIMM single-link failover.

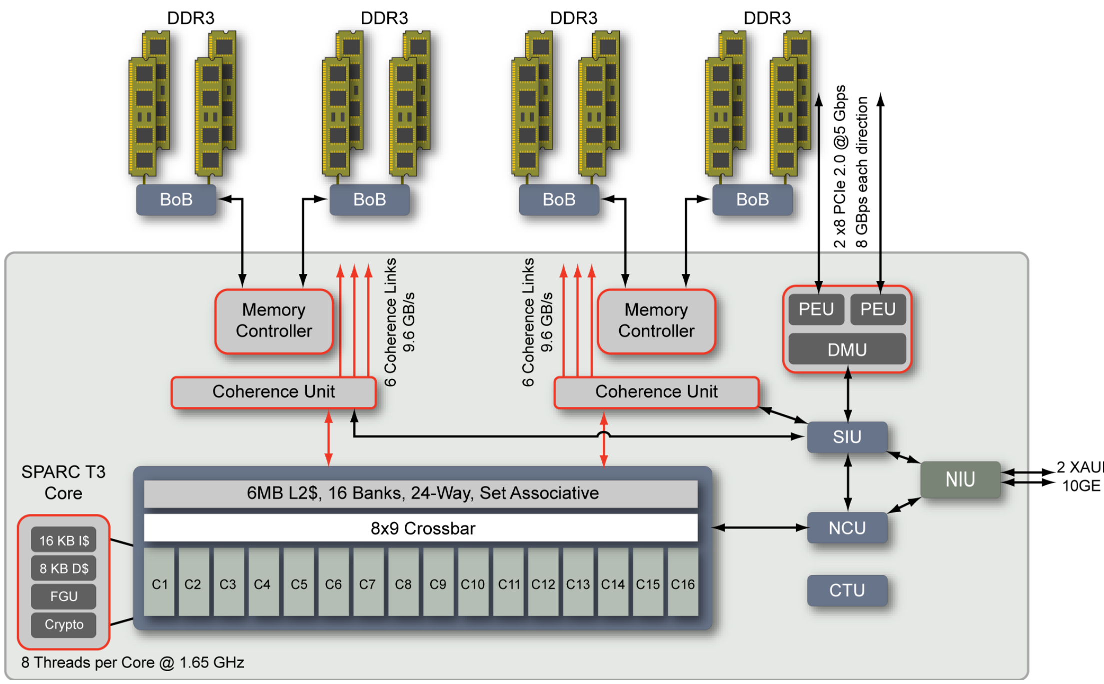


Figure 3. The SoC design of the multiprocessing SPARC T3 processor incorporates massive compute power, I/O, network, and cryptographic capabilities onto a single die.

The SPARC T3 processor includes the following key design elements.

- Support for up two or four-processor multiprocessor implementations
- Six coherency links
- Support for 128 simultaneous threads per processor
- Sixteen cores per processor each with two integer processing pipelines per core
- Eight threads per core
- Sixteen newly designed, fully pipelined floating-point units (one per core)
- Sixteen stream processing units (one per core, acting as cryptographic coprocessors operating in parallel with the core)
- On-chip caches and memory management
- Two On-chip PCIe Generation 2 I/O interfaces
- Two On-Chip 10 GbE XAUI ports
- Coherency links with FBDIMM-like single-link failover

The SoC design of the SPARC T3 processor reduces the need for additional ASICs to connect onboard components. As a result, SPARC T3-1, T3-2, T3-4, and T3-1B servers simply contain fewer parts and pins that can fail than many traditional servers with multiple single-threaded dual-core or quad-core processors. Figure 4 illustrates the drastic simplification in system design afforded by the SPARC T3 processor.

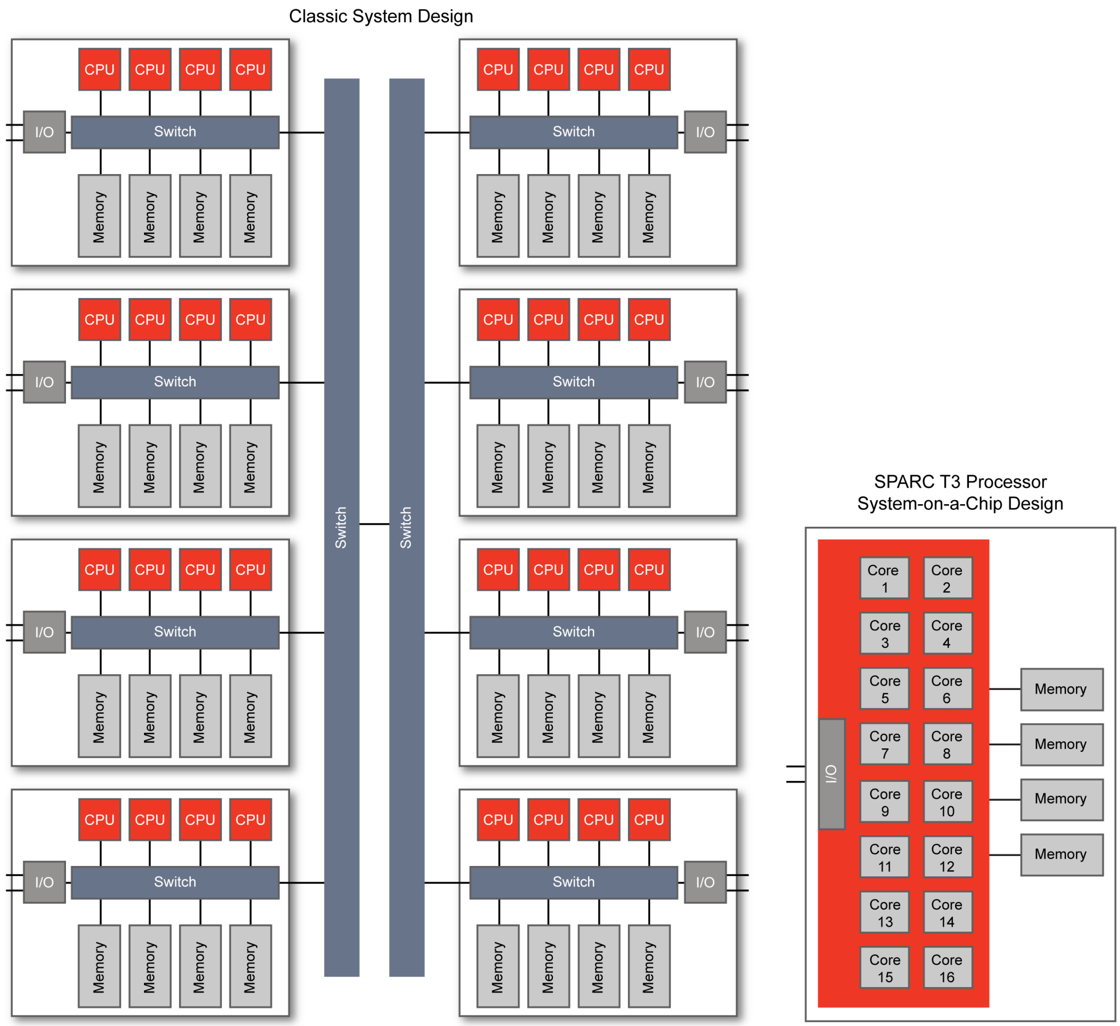


Figure 4. With 16 cores providing 128 threads, the SPARC T3 processor maximizes compute power and minimizes system component count at the same time, delivering greater reliability than systems with many more processors, multiple system boards, and far more complicated designs.

Improving Availability: Redundancy, Serviceability, and Self-Diagnosis

Minimizing the need for system interruptions in order to correct error conditions improves the availability of a system and the hosted IT service. SPARC T3-1, T3-2, T3-4, and T3-1B servers provide many design features to help enable completion of maintenance procedures without having an impact on continuous system operation.

Redundant Hot-Swap Components and RAID Capabilities

Oracle's SPARC T3 servers are built to achieve high levels of uptime and fast recovery from failures. Two PCIe root complex per processor and the ability to configure multiple CPUs (T3-2 and T3-4), memory DDR3 DIMMs, and I/O cards add to the resiliency of Oracle's SPARC T3-1, T3-2, T3-4, and T3-1B servers. Hot-swap and hot-plug chassis-mounted hard drives, fan units, and power supplies improve serviceability and availability. For systems configured with redundant components, administrators can use software commands to remove and replace disks, power supplies, and fan units while the system continues to operate.

The built-in RAID capabilities of the SPARC T3-1, T3-2, T3-4 and T3-1B servers provide data redundancy and increased performance at no additional cost. These servers support onboard hardware RAID 0/1 to enable striping or mirroring of data across any two internal drives. In addition, Oracle's SPARC T3-1, T3-2, T3-4, and T3-1B servers running Oracle Solaris can take advantage of software RAID capabilities. For example, Oracle Solaris ZFS can be used for internal or external storage devices, providing flexibility and redundancy beyond the chassis.

Hardware Protection—ECC Everywhere

System self-diagnosis, error correction, and parity checking features in high-end systems help organizations maximize IT service uptime. SPARC T3-1, T3-2, T3-4, and T3-1B servers include mainframe-class processor RAS features that enhance system uptime by maintaining data integrity across on-chip memory—a unique capability not found in other volume market systems.

The extensive data protection of the SPARC processor provides for self-diagnosis and corrective actions, facilitating continuous system operation in the face of error conditions. Parity protection is provided throughout the SPARC T3 processor, including the following elements.

- Instruction cache (I-cache) tags and data
- Data cache (D-cache) tags and data
- Internal L2 cache tags with ECC protection of data
- Instruction translation look-aside buffer (ITLB)
- Data translation look-aside buffer (DTLB)
- Modular arithmetic memory and store buffer addresses
- Coherency links

Utilizing a combination of hardware and software correction flows, SPARC T3-1, T3-2, T3-4, and T3-1B servers work to maximize data integrity. In the event of an error, hardware refetch is used for single-error correction (SEC) of the I-cache and D-cache. ECC protection is provided for integer register file (RF), floating-point RF, store buffer data, trap stack, and other internal arrays.

SPARC T3-1, T3-2, T3-4, and T3-1B servers use content-addressable memory (CAM) technology to implement hardware-based virtual address (VA) lookup, and physical address translation is implemented in RAM. Parity protection is provided for both VA lookup and physical address translation. Although parity protection for the physical address translation function is increasingly common, variable-page-size support makes accomplishing parity protection for hardware-based VA lookup much more challenging. Oracle's approach to providing parity protection for hardware-based VA lookup gives organizations added reliability not found in other platforms.

The Robust SPARC T3 Processor

SPARC T3-1, T3-2, T3-4, and T3-1B servers leverage the UltraSPARC T2 Plus processor's fault management capabilities. The SPARC T3 processor provides fault monitoring and recovery mechanisms for major elements, including the cores, threads, integrated I/O root complex, processor coherency links, and dedicated cryptographic and floating-point units (FPUs). To help speed maintenance efforts, fault messaging clearly identifies the field-replaceable units (FRUs) requiring attention, the system impact, and suggested corrective action.

- **Thread and core offlining.** The SPARC T3 processor includes the ability to diagnose CPU errors and can perform offline operations at the thread and core level with no impact on adjacent threads or cores or interruption of executing IT services. In the event of a fault within a processor resource that is shared across all threads, the entire core will be offlined to provide adequate fault isolation.
- **Coherency links.** ECC and parity-protected coherency links connect sockets within the SPARC T3 processor, helping ensure the integrity of cache data.
- **Memory.** The memory subsystem of the SPARC T3 processor offers diagnosis at the memory bank, DDR3 DIMM, and page level and provides page retirement capabilities. In addition, each memory channel includes multiple high-speed serial lanes in each direction and supports single-lane automatic failover, enabling continued operation if a lane fails due to excessive errors.
- **I/O subsystem.** The SPARC T3 processor provides fault management and diagnosis of the I/O subsystem and PCIe fabric. In addition to diagnosis of the root complex itself, the processor takes advantage of additional support in Oracle Solaris for PCIe fabric diagnosis.
- **Cryptographic units**. If a fault is diagnosed on one of the cryptographic units (SPUs), offlining capabilities will prevent cryptographic drivers from utilizing the failed device. At the same time, surviving cryptographic units on the other processor cores will remain available and processor threads will remain active.

Integrated Lights Out Management for Simplified Remote Serviceability

A delay in error notification lengthens the time required to resume operations. In addition, cumbersome system maintenance tasks can prolong system downtime and introduce new system configuration issues. ILOM 3.0 software and an embedded service processor on SPARC T3-1, T3-2, T3-4, and T3-1B servers ease remote system management, simplify administration, and speed maintenance tasks.

The service processor runs independently of the server, using the server's standby power. Therefore, ILOM firmware and software continue to function when the server operating system goes offline or when the server is powered off. Integrated Lights Out Manager monitors the following the SPARC T3-1, T3-2, T3-4, and T3-1B server conditions.

- CPU temperature conditions
- Hard drive status
- Enclosure thermal conditions
- Fan speed and status
- Power supply status
- Voltage conditions
- Oracle Solaris watchdog, boot timeouts, and automatic server restart events

Integrated Lights Out Manager enables administrators to monitor and control SPARC T3-1, T3-2, T3-4, and T3-1B servers over a dedicated Ethernet connection and supports Secure Shell (SSH), Web, and Integrated Platform Management Interface (IPMI) access. Integrated Lights Out Manager functionality also can be accessed through a dedicated serial port for connection to a terminal or a terminal server. The ILOM command-line and browser-based interfaces simplify remote administration of geographically distributed or physically inaccessible machines. In addition, Integrated Lights Out Manager provides remote execution of diagnostics that generally require physical proximity to the server serial port. Integrated Lights Out Manager can be configured to distribute e-mail alerts about hardware failures and warnings as well as other events related to the server.

Environmental Monitoring

Environmental elements play a critical role in the stability of system operation. SPARC T3-1, T3-2, T3-4, and T3-1B servers take a proactive approach to protecting internal components against extreme temperatures, lack of adequate airflow through the system, power supply failures, and hardware faults. They even provide the ability to report system power consumption. Furthermore, the Intelligent Fan Control (IFC) feature of SPARC T3-1, T3-2, T3-4 servers can help avoid fan speeds or frequencies that introduce unnecessary vibration. Minimizing vibration can reduce component wear.

Temperature sensors are located throughout SPARC T3-1, T3-2, T3-4, and T3-1B servers to monitor the ambient temperature of the system and internal components. Fan speeds automatically adjust to compensate for thermal conditions and actually slow whenever possible, reducing power consumption.

If the temperature observed by a sensor falls below a low-temperature threshold or rises above a high-temperature threshold, the monitoring subsystem software will generate a temperature warning alert. If the temperature condition persists and reaches a critical threshold, the monitoring subsystem will light the amber Service Required indicator on the front and back panel and initiate a system shutdown. The power subsystem is handled in a similar fashion: power supplies are monitored, and any fault is reported via the front and rear panel indicator lights. In the event of a failure of the system controller, a forced hardware shutdown will occur to protect the system from serious damage. Indicators remain lit after an automatic system shutdown to aid in problem diagnosis.

Isolating Faults with Oracle Virtualization Technologies

Consolidation of many applications on a single server is often necessary for economically hosting and managing the high number of IT services required within an enterprise. Oracle VM Server for SPARC and Oracle Solaris Containers technology provide the ability to virtualize the considerable system resources of SPARC T3-1, T3-2, T3-4, and T3-1B servers. Within the context of a consolidation strategy, virtualization technologies from Oracle can help isolate individual software application faults so they don't affect other IT services hosted on the same platform. In fact, using Oracle virtualization technologies can help organizations improve resource utilization while reducing downtime.

Oracle VM Server for SPARC

Any virtualization solution must be carefully examined for its contribution to the RAS of key applications and services. Implemented in both firmware and hardware, Oracle VM Server for SPARC fosters greater reliability than software-based virtualization solutions. In addition to robust fault isolation, Oracle VM Server for SPARC includes several options for creating redundancy within configurations, provides proactive fault management capabilities, and implements a unique I/O architecture that reduces the recovery time of guest domains.

Oracle VM Server for SPARC provides extensive resource isolation through firmware and hardware constructs. Each logical domain can be created, destroyed, reconfigured, and rebooted independently without requiring a power cycle of the server. Logical domains are also managed as entirely independent machines, with localized control of the following resources.

- Kernel, patches, and tuning parameters
- User accounts
- Disks
- Network interfaces
- MAC addresses
- IP addresses

Oracle VM Server for SPARC uses the SPARC hypervisor, a firmware layer included in the system firmware stack loaded on the motherboard, to virtualize machine hardware and decouple the link between the operating system and the hardware (Figure 5). As such, the number of virtual machines that can be created relies on the capabilities of the hypervisor, as opposed to the number of physical hardware devices installed in the system. Within SPARC T3-1, T3-2, T3-4, and T3-1B servers, as many as 128 logical domains can be established.[1] Each logical domain is a full virtual machine that runs an independent operating system instance or runtime environment and contains virtualized CPU, memory, storage, console, and cryptographic devices.

All logical domain instances rely on the fundamental technology constructs just described. However, several different roles exist for domains. Based on the context and use, a single logical domain can function in one or more of the following roles.

- The *control domain* governs domain creation and assignment of physical resources.
- The *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.
- The *I/O domain* controls direct physical access to input/output devices such as PCIe cards, storage units, and network devices.
- The *guest domain* utilizes virtual devices offered by service and I/O domains and operates under the management of the control domain.

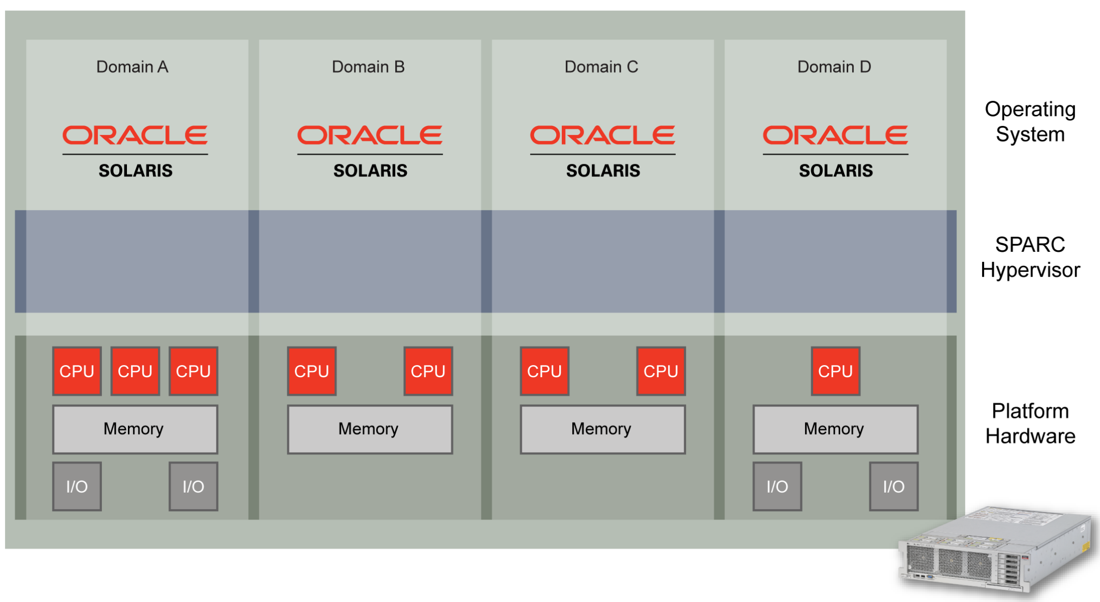


Figure 5. The SPARC T3 hypervisor provides isolation of virtual machine operating system instances and allocated hardware resources through Oracle VM Server for SPARC.

[1] The optimal number of logical domains appropriate for any particular system will vary according to the resources available and the mix of applications.

Logical Domain Isolation

Organizations often utilize virtualization technologies to host multiple business-critical applications on the same server. As a result, enterprises need assurance that software faults or maintenance events in one virtual machine will remain isolated and unable to affect the availability of IT services in other partitions.

Unfortunately, many virtualization solutions require a full system reset for rebooting a control or a service domain. In contrast, Oracle VM Server for SPARC supports reboot and reset of any logical domain independent of all other logical domains—even logical domains with direct control over physical hardware resources. Guest domains can be configured, started, and stopped independently, without the need to power-cycle the machine and without any impact on continuous operation of other domains. In addition, virtual I/O interfaces can connect and disconnect as necessary without affecting other domains on the same platform. Administrators can even dynamically add virtual CPUs and remove them from a logical domain while the operating system instance continues to execute, helping avoid the need for planned downtime.

Robust Virtualized I/O

Continuous communication with I/O devices proves critical to delivery of many IT services. Within Oracle VM Server for SPARC, an I/O domain is a logical domain that physically connects to I/O devices. I/O domains also generally act as service domains, sharing I/O access to other logical domains in the form of virtual devices. Oracle VM Server for SPARC provides several features and architectural elements to mitigate the impact of I/O and service domain fault conditions, helping provide uninterrupted access to I/O devices.

- **Masking I/O fault events.** The capabilities of Oracle VM Server for SPARC mitigate the impact of temporary loss of I/O connectivity. In response to a fault or maintenance operation on an I/O device, Oracle VM Server for SPARC interfaces with Oracle Solaris to gracefully suspend I/O operations to all affected virtual I/O devices. After recovery of the physical I/O device, virtual I/O devices can reconnect and resume I/O transactions at the point of suspension.
- **Redundant I/O paths.** Although I/O suspension provides a means of masking faults, many applications simply cannot afford any period of time without I/O service. Configuring redundant I/O paths and utilizing Oracle Solaris I/O multipathing software can help ensure continuous access to disk and network services despite the loss of a single I/O route. SPARC T3-1, T3-2, T3-4, and T3-1B servers are equipped with multiple PCIe root complexes, supporting the configuration of two or more service domains. Each service domain is capable of providing a unique path to a physical disk subsystem if additional host bus adapters (HBAs) are installed. This scenario is somewhat analogous to employing two HBAs in a traditional I/O connectivity model. Once initialized within a guest domain, Oracle Solaris I/O multipathing software provides a mechanism for automatic failover between the two paths in the event of a failure. In addition, Oracle Solaris I/O multipathing software includes a manual path switchover capability, enabling administrators to redirect I/O to an alternative path during reconfiguration or reboot of the primary I/O domain.

- **Clustering I/O domains.** Administrators can further increase the availability of virtualized I/O services by adopting clustering technology. Oracle VM Server for SPARC technology is certified for use with Oracle Solaris Cluster to provide for failover of I/O domains. Clustering I/O domains provides for automated failover of I/O device services, minimizing the time it takes to recover I/O connectivity and ultimately improving overall IT service availability.

Reducing Boot Times

After a fault or a maintenance event occurs, reducing the time required to bring systems back online is a key step toward maximizing IT service availability. Hosting an IT service in a logical domain can actually reduce recovery time. In the event of a planned or unplanned reboot of a logical domain, Oracle VM Server for SPARC enables an operating system instance to bypass time-consuming I/O bring-up procedures. Most domains utilize virtual I/O devices, delegating the burden of I/O bus ownership, probing buses for devices, and loading device drivers to an I/O domain. Domains that utilize virtual I/O devices contain no I/O bus topology to probe and no physical connection to I/O devices, speeding recovery by eliminating the need for time-consuming I/O initialization steps during boot of an operating system.

Oracle Solaris Containers

Whether used in a single system image of Oracle Solaris or within a domain, Oracle Solaris Containers can further isolate software applications and services, using flexible, software-defined boundaries. A breakthrough approach to virtualization and software partitioning, Containers technology enables creation of many private execution environments within a single instance of Oracle Solaris (Figure 6).

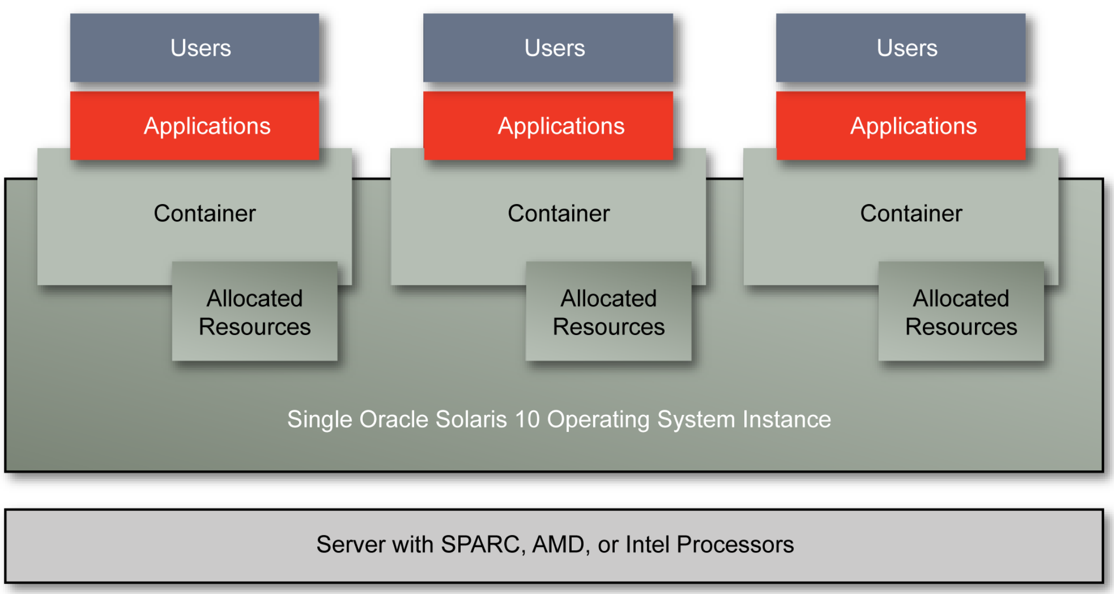


Figure 6. Oracle Solaris Containers use flexible software mechanisms to isolate applications.

Oracle Solaris Containers provide a complete, isolated, secure runtime environment for applications and enable granular management of system resources. Each Container can be managed independently with regard to users, device paths, CPU and memory resources, and networking. Dynamic resource reallocation capabilities enable unused system resources to shift among containers as needed, providing high-quality service to hosted applications. Resources can also be allocated or reserved for critical services to reduce contention for compute power with lower-priority workloads. The fine-grained resource management provided by Containers encourages efficient use of computing power while maintaining high levels of service availability.

Applications within containers are isolated, preventing activities in one container from monitoring or affecting processes running in another container. Even a super-user process cannot view or affect activity in other containers. Software fault and security isolation features of Oracle Solaris Containers also prohibit poorly behaved applications from having an impact on other containers. By utilizing Oracle Solaris Containers and Oracle VM Server for SPARC individually or in combination, organizations can better define and meet service levels by dynamically controlling application and resource priorities.

Comprehensive Focus on IT Service Uptime

Operating system and management tool choices heavily influence IT service uptime. As a part of a dedicated effort to help organizations build highly dependable solutions, Oracle works to integrate RAS at all levels of the platform. In fact, Oracle Solaris 10 includes a comprehensive fault management architecture known as Oracle Solaris predictive self-healing that governs error handling throughout the hardware, virtualization, and operating system layers of SPARC T3-1, T3-2, T3-4, and T3-1B servers. By utilizing Oracle Solaris and Oracle management tools in conjunction with these servers, organizations can further enhance the availability of hosted IT services.

Speeding Diagnosis with a Fault Management Architecture

When faults occur or maintenance is required, the amount of time required to return an IT service to an operational state becomes critical. Extending the fault management framework first introduced in Oracle Solaris, the predictive self-healing feature governs error events across all technology layers within SPARC T3-1, T3-2, T3-4, and T3-1B server platforms, including individual hardware components and Oracle VM Server for SPARC. By proactively diagnosing, isolating, and recovering from hardware as well as software failures, predictive self-healing technology provides meaningful information about faults, speeds resumption of IT services, and even helps prevent certain system failures.

Predictive Self-Healing

The predictive self-healing feature in Oracle Solaris proactively monitors and manages system components to help organizations achieve maximum availability of IT services. Predictive self-healing is an innovative capability that automatically diagnoses, isolates, and recovers from many hardware and application faults. As a result, business-critical applications and essential system services can continue uninterrupted in the event of software failures, major hardware component failures, and even software configuration problems.

The fault manager and service management facility in Oracle Solaris are the two main components in predictive self-healing. Fault manager receives data relating to hardware and software errors and automatically diagnoses the underlying problem. Once the problem is diagnosed, fault manager automatically responds by offlining faulty components and signaling the error condition to administrators (via console messages and external system and component status indicators on the front and rear of the chassis). The service management facility turns services—rather than processes—into first-class citizens, permitting automatic self-healing. Service descriptions for base services of Oracle Solaris include full dependency information for start, stop, and restart. Configuring user applications to run under the service management facility is relatively simple, helping organizations effectively manage faults for IT services hosted on SPARC T3-1, T3-2, T3-4, and T3-1B servers.

Predictive Self-Healing for SPARC T3-1, T3-2, T3-4, and T3-1B Servers

SPARC T3-1, T3-2, T3-4, and T3-1B servers leverage several predictive self-healing response agents already in Oracle Solaris, including syslog, CPU offline, MPR, and I/O retire agents. The ILOM software on the service processor provides an additional set of response agents. These agents include a dynamic field-replaceable unit ID (FRUID) agent that updates the faulted FRU with the error event and an LED agent that lights the appropriate system or component (FRU) status indicator for specific faults. Fault diagnosis and recovery status are synchronized between the logical domain and the service processor. Fault events on SPARC T3-1, T3-2, T3-4, and T3-1B servers include the FRU part and serial number of the faulted components.

The following capabilities are a few examples of the granularity and power of the predictive self-healing feature when implemented on the SPARC T3-1, T3-2, T3-4, and T3-1B servers.

- Within the SPARC T3 processor, the Level 1 (L1) and Level 2 (L2) caches, per-thread registers, TLBs, and cryptographic units include error reporting mechanisms that can be diagnosed by predictive self-healing technology.
- When a correctable memory error is encountered on SPARC T3-1, T3-2, T3-4, and T3-1B servers, the error is queued up for predictive self-healing diagnosis. Automatic retirement of memory pages as necessary and predictive analysis and self-healing of DDR3 DRAM errors help avoid potential system interruptions.

- Faults within SPARC T3-1, T3-2, T3-4, and T3-1B servers governed by the predictive self-healing feature include direct I/O devices with hardened and nonhardened drivers, as well as devices that are virtualized by the hypervisor.
- In the event that a single-lane failover of a coherency link is triggered within SPARC T3-1, T3-2, T3-4, and T3-1B server, the fault will be logged, the indicator light will be turned on, and the system will continue to operate.
- The PCIe switches and SAS/SATA controllers utilize hardened device drivers that generate reports whenever an error is detected by the error handlers. Errors are forwarded for complete analysis and diagnosis.

Predictive Self-Healing for Domains

For proper error handling within a virtualized environment, messages and alerts must extend to virtual machine instances. In addition to close integration with SPARC T3-1, T3-2, T3-4, and T3-1B server hardware, the predictive self-healing feature governs fault handling within domains. Taking a holistic system view, error conditions for devices abstracted by the hypervisor—processors, memory, console, and cryptographic devices—trap to the hypervisor, generating a standard error report that is sent to the service processor via the host-to-service-processor mailbox channel. A diagnosis engine provides complete analysis of each error report, and distributes alerts to each potentially affected OS instance. A knowledge base is available for the predictive self-healing feature in Oracle Solaris. Administrators can use the knowledge base to correlate error messages related to Oracle VM Server for SPARC execution to documented predefined corrective actions, shortening time to system recovery.

Reducing Downtime with Oracle Solaris

More than two decades of investment have contributed to making Oracle Solaris one of the most reliable operating systems. Key computing elements—operating system, networking, and user environment—combine within Oracle Solaris to provide a stable, high-quality foundation for execution of IT services. In fact, many organizations can point to systems running Oracle Solaris that execute continuously for months or years without the need for a restart.

Oracle Solaris is designed for availability. Built with a small, compact kernel, it limits the potential for operating system faults and subsequent platform downtime. In addition, Oracle Solaris establishes a clear distinction between the kernel, shared libraries, and applications to limit the impact of application failures. Furthermore, the ability to install most patches and other incremental software updates for Oracle Solaris without taking the system offline helps organizations increase uptime and ease serviceability. Ease-of-use features, including Web-based installation and a graphical process manager, can also help boost availability, reducing the risk of operator error and minimizing service times. Described in the sections that follow, several key features of Oracle Solaris can help organizations develop, deploy, and manage IT services with extreme reliability, relentless availability, and simplified serviceability.

Memory Page Retirement

In many systems, addressing both correctable and uncorrectable permanent memory errors requires server downtime. As a part of the predictive self-healing technology framework, the MPR capability works to isolate memory issues without system interruption. Diagnosis software in the fault manager facility examines memory correctable errors and uncorrectable errors detected by underlying hardware, on a continual basis. MPR retires memory pages containing correctable errors and relocatable clean pages containing uncorrectable errors without interrupting user applications. In addition, MPR can isolate relocatable dirty pages containing uncorrectable errors with limited impact on affected user processes and can avoid forcing an outage of an entire system.

Oracle Solaris File Systems

Reliable data subsystems are critical to creating highly available IT services. Organizations continue to depend on the UNIX file system within Oracle Solaris 10 to provide high-resiliency features, such as metadata logging, to protect against data corruption and speed recovery in the event of system failure. In addition, Oracle Solaris 10 now also features Oracle Solaris ZFS, a file system that offers a dramatic advance in data management with an innovative approach to data integrity.

Oracle Solaris ZFS provides increased protection against administrative error and delivers end-to-end data integrity elements such as 64-bit checksumming and comprehensive data updates. To ensure that the data on disk is self-consistent at all times, Oracle Solaris ZFS combines proven and cutting-edge technologies, such as copy-on-write and end-to-end checksumming. Data is always written to a new block on disk before the pointers to the data are changed and the write is committed. Because the file system is always consistent, time-consuming recovery procedures such as file system check (`fsck`) are not required if the system is shut down in an unclean manner. Copy-on-write also enables administrators to make consistent backups or roll data back to a known point in time.

Oracle Solaris 10 with Oracle Solaris ZFS is the only OS designed to provide end-to-end checksumming for all data. Oracle Solaris ZFS constantly reads and checks data to help ensure integrity, and if an error exists in a mirrored pool, the technology can automatically repair the corrupt data. This relentless vigilance on behalf of availability protects against costly and time-consuming data loss—even previously undetectable silent data corruption.

Oracle Solaris Flash Archive and Oracle Solaris Live Upgrade

In many cases, planned downtime accounts for the bulk of system interruptions each year. The flash archive feature in Oracle Solaris and Oracle Solaris Live Upgrade help organizations decrease requirements for planned downtime by providing efficient installation and upgrade operations. With flash archive and Oracle Solaris Live Upgrade, deployment or upgrades of SPARC T3-1, T3-2, T3-4, and T3-1B servers can complete in minutes.

- The flash archive facility helps IT organizations quickly install and update systems with an operating system configuration tailored to enterprise needs. The technology in flash archive provides tools to help system administrators build custom rapid-install images—including applications, patches, and parameters—that can be installed at a data rate close to the full speed of the hardware.

- Oracle Solaris Live Upgrade software provides mechanisms for upgrading and managing multiple on-disk instances of Oracle Solaris. This technology facilitates installation of a new operating system or patches on a running production system without taking a server offline. Downtime is required only to reboot the new configuration. One feature of Oracle Solaris Live Upgrade gives administrators the ability to quickly roll systems back to the initial state if desired.

Oracle Software for Efficient System Management

Technologies that automate management procedures can help prevent faults and provide efficiencies that help system administrators manage a greater number of servers. In addition to benefiting from the ILOM software, organizations can take advantage of powerful Oracle Enterprise Manager Ops Center. This sophisticated tool automates monitoring and administrative functions, lowering the administrative burden and reducing the chance of common errors. Oracle Enterprise Manager Ops Center automates complex software installation and configuration, helping ease operations.

Oracle Enterprise Manager Ops Center

Providing high availability levels in a rapidly changing IT infrastructure requires the ability to provision and re-provision servers efficiently while minimizing the opportunity for process errors. Oracle Enterprise Manager Ops Center automates complex software installation and configuration across heterogeneous network infrastructures, helping speed and simplify IT service deployments. Oracle Enterprise Manager Ops Center provides a comprehensive solution to simplify system infrastructure lifecycle management. By utilizing Oracle Enterprise Manager Ops Center, administrators can discover, provision, monitor, update, and manage hundreds of x64-based servers and servers based on Oracle's Sun SPARC architecture from a single console. Oracle Enterprise Manager Ops Center software also provides greater flexibility in allocating and reallocating resources as enterprise requirements shift.

Designed to simplify datacenter management tasks, Oracle Enterprise Manager Ops Center includes features that facilitate remote power control, operating system deployment and patching, system BIOS and firmware updates, event logging and notification, and hardware and operating system monitoring. In addition, the software features the ability to create logical groups of systems and perform actions across these groupings as easily as if it were performing actions on a single node. By providing fast and easy access to systems for monitoring and maintenance, Oracle Enterprise Manager Ops Center increases operational efficiency and minimizes downtime related to process errors.

Measuring Server Reliability, Availability, and Serviceability

Taking a holistic view of system operation, Oracle studies the interrelation between reliability, availability, and serviceability and works to optimize all three factors at the same time. By utilizing repeatable metrics during the design of new systems, Oracle can achieve continuous improvements over current and previous generations of Oracle servers and competitive systems. Given the importance of system availability, Oracle continues to invest and work actively to help establish industrywide standards for measurement.

Availability Benchmark Framework

Benchmarks often guide research-and-development efforts to enhance computer systems performance. As the RAS features of systems increase in importance, there is a natural desire on the part of engineers to benchmark availability levels. Unfortunately, industry-accepted standards are still a work in progress in this important area. The goal of adopting a methodical approach to improving system dependability and the lack of an industry standard for measurement led to the creation of an availability benchmark framework called R-Cubed [2]. Oracle's R-Cubed availability benchmark framework is applicable to a wide range of systems, and information about the methodology is published openly for reuse by the industry, academia, and the technical community at large.

In addition to providing traditional measures of server outage minutes per year, the R-Cubed availability benchmark framework reaches further to focus on overall dependability—the ability of a system to remain available and prevent service interruptions. As such, the R-Cubed framework accounts for system RAS levels by quantifying three key attributes.

- **Fault and maintenance rate** measures the frequency of system faults in a given period of time.
- **Robustness** accounts for the extent to which system operation degrades due to a fault, as well as the potential for completing repairs online.
- **Recover**y examines system characteristics to quantify the effort required to return to an operational state after a fault or maintenance event.

Although optimizing the three elements of the R-Cubed framework results in high levels of system dependability, increasing all three R-Cubed factors at the same time requires a careful approach. For example, designing a server with a low component count minimizes the fault and maintenance rate, increasing reliability. However, a system with a low component count most likely lacks the robustness provided by redundant parts. Without redundancy, single-component errors or faults can cause system outages—essentially placing reliability at odds with increased availability. Utilizing the R-Cubed framework helps Oracle take a balanced approach to maximizing overall system RAS.

R-Cubed Metrics

The R-Cubed availability framework incorporates the following measures to help system designers quantify the dependability of a system design and evaluate potential improvements.

- **Fault Robustness Benchmark-A (FRB-A).** Rewards systems in which faults do not cause disruption of service. Measurement is indicated by a numeric scalar between 1 and 100, inclusive, where 1 means that any single failure causes disruption and 100 means that no single failure causes a disruption. A system scores higher on FRB-A by optimizing cost and redundancy trade-offs—less reliable parts are made redundant, and more reliable parts are utilized where economical.
- **Maintenance Robustness Benchmark-A (MRB-A).** Quantifies the ability to perform maintenance without system disruption by utilizing a numeric scalar. A score of 1 means that all

[2] "R-Cubed (R3): Rate, Robustness, and Recovery—An Availability Benchmark Framework," labs.oracle.com/techrep/2002/abstract-109.html

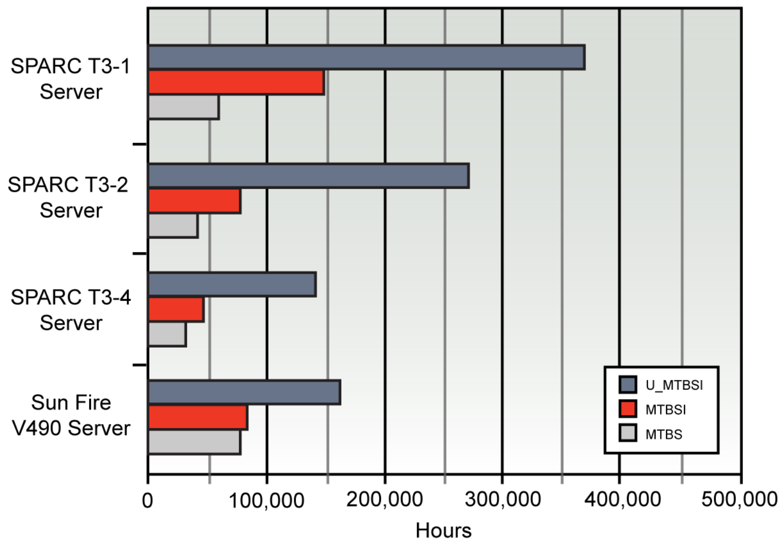
maintenance actions result in a system outage, whereas a score of 100 indicates that all field-replaceable units (FRUs) can be replaced without downtime. MRB-A scores are higher for systems with hot-swap capabilities. In general, small-form-factor servers do not score as well as larger servers designed for online hardware servicing.

- **Mean Time Between Services (MTBS).** Utilizes mean time between failures (MTBF) rates, isolating calculations to include only FRU failures that incur a service exception.
- **Mean Time Between System Interrupts (MTBSI).** Quantifies component failures that lead to system downtime and scheduled shutdowns for service actions. MTBSI considers scheduled and unscheduled system interruptions. In some cases, an unscheduled interruption results in a degraded mode of system operation and requires a scheduled interruption to replace failed components.
- **Unscheduled Mean Time Between System Interrupts (U_MTBSI).** Measures the rate of system interrupts that are caused by component failures. Unscheduled interruptions pose the most significant threat to predictable service delivery.
- **Availability.** Provides a traditional measure of the percentage of uptime achieved by a system per year.

As evidenced by the data in Table 2 and the bar charts in Figure 7, SPARC T3-1, T3-2, and T3-4 servers make great strides toward maximizing dependability, compared to earlier Sun systems such as the Sun Fire V490 server, which was released only a few years ago.

TABLE 2. R-CUBED METRICS SCORES OF ORACLE'S SPARC T3-1, T3-2, T3-4, AND SUN FIRE V490 SERVERS

SYSTEM	FAULT ROBUSTNESS BENCHMARK-A	MAINTENANCE ROBUSTNESS BENCHMARK-A	U_MTBSI	MTBSI	MTBS	AVAILABILITY
Oracle's SPARC T3-1 server	86.6	53.4	366972	148,526	56220	0.99998
Oracle's SPARC T3-2 server	89.1	62.3	269428	78090	34571	0.99996
Oracle's SPARC T3-4 server	85.1	27.3	143172	40612	27057	0.99993
Oracle's Sun Fire V490 server	58.7	32.77	161,804	83,539	77,905	0.99998



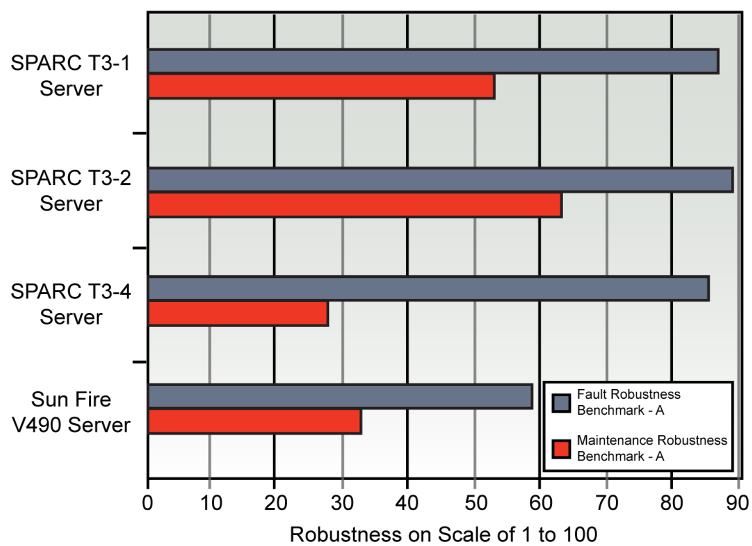


Figure 7. These graphs show the R-Cubed benchmark results for Oracle's T3-1, T3-2, and T3-4 servers and the Sun Fire V490 server.

Conclusion

Requirements for relentless availability of IT services are increasingly common. Deploying SPARC T3-1, T3-2, T3-4, and T3-1B servers can help support organizational efforts to achieve high IT service uptime goals. A low component count, extensive data integrity features, and superior energy efficiency promote reliability. Redundant components foster high levels of availability. ILOM, hot-swappable components, and self-healing features simplify serviceability.

As a part of a comprehensive design approach, the processor, virtualization technology, operating system, and software tools for SPARC T3-1, T3-2, T3-4, and T3-1B servers all provide innovative RAS features. Utilizing Oracle VM Server for SPARC and Oracle Solaris Containers technology to virtualize platforms helps organizations isolate IT services against failure while optimizing asset utilization. Oracle Solaris enables RAS features, such as MPR and Extended ECC protection, previously available only in large-scale systems. The extension of the predictive self-healing feature in Oracle Solaris to SPARC T3-1, T3-2, T3-4, and T3-1B platforms, the SPARC T3 processor, and Oracle VM Server for SPARC help facilitate rapid system recovery in the event of a fault. Oracle also provides additional tools such as Oracle Enterprise Manager Ops Center to help organizations prevent administrative errors, realize greater serviceability, and benefit from shorter recovery times.

Oracle brings enterprise expertise and innovation to the development of hardware and software products—each new generation of systems improves platform reliability, availability, and serviceability capabilities. As a result, Oracle's systems provide a strong foundation for organizations seeking to support nonstop IT service operations.

Maximizing Service Uptime with Oracle's
SPARC T3-1, SPARC T3-2, SPARC T3-4, and
SPARC T3-1B Servers
October 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