ORACLE

SPARC SUPERCLUSTER

An Oracle White Paper
December 2012

SPARC SuperCluster: Taking Reliability and Availability to the Next Level

ORACLE

Introduction

The Oracle SPARC SuperCluster T4-4 is a multi-purpose engineered system that has been designed, tested and integrated to run mission critical enterprise applications and rapidly deploy cloud services while delivering extreme efficiency, cost savings, and performance. It is well suited for multi-tier enterprise applications with Web, database, and application components. This versatility, along with powerful, bundled virtualization capabilities, makes it an ideal platform on which to consolidate large numbers of applications, databases, and middleware workloads, or to deploy complex, multiuser development, test, and deployment environments.

The reliability, availability, and serviceability (RAS) characteristics of the Oracle SPARC SuperCluster T4-4 are more than just reliable components. It is the combination of hardware and software RAS combined with advanced and integrated management and monitoring that enable the Oracle SPARC SuperCluster T4-4 to deliver mission-critical uptime and reliability.

Oracle SPARC SuperCluster T4-4 combines highly available and scalable technologies, such as Oracle Real Application Clusters (Oracle RAC), a feature of Oracle Database 11*g*, and Oracle Solaris Cluster software with industry-standard hardware. Its architecture enables a high degree of isolation between concurrently deployed applications, which may have varied security, reliability, and performance requirements.

The SPARC SuperCluster platform is architected with high availability (HA) from the ground up. It is enabled with Oracle RAC for Oracle Database HA and Oracle Solaris Cluster for application HA, and the combination delivers end-to-end HA. This software HA combined with hardware (server, storage, and network), which has no single point of failure (SPOF), makes the platform ideal for consolidation and for running mission-critical applications.

SPARC SuperCluster T4-4 provides an optimal solution for all database workloads, ranging from scan-intensive data warehouse applications to highly concurrent online transaction processing (OLTP) applications. With its combination of the Oracle Exadata Storage Server, Oracle Database software, and high-performance compute and network components, SPARC SuperCluster T4-4 delivers extreme performance in a highly available, highly secure environment.

For application environments that follow Oracle's best practices for highly scalability, no application architecture or design changes are required to benefit from SPARC SuperCluster T4-4. Customers can integrate SPARC SuperCluster T4-4 systems with their current data center infrastructure using the available 10 GbE ports in each SPARC T4-4 server within SPARC SuperCluster.

Figure 1. SPARC SuperCluster T4-4

Reliability, Availability, and Serviceability (RAS)

In casual conversations about system and services availability, the term RAS is often used to reflect availability. However, the other two components of RAS, reliability and serviceability, are equally important when considering the capabilities of a delivery platform or data center solution.

- **Reliability**
  There are several ways to deliver reliability. The first method is to reduce the number of components in a system. Compared to large, monolithic servers, SPARC SuperCluster has far fewer components. This reduction of components drives higher mean time between failures (MTBF). A second method for better reliability is to reduce the number of connectors for ASICS and boards. Larger servers tend to have more hot-swap or hot-plug components than rack form factor servers. This means far more connectors, which lowers reliability. The third method for better reliability is data integrity. Features such as error-correcting code (ECC), data parity, and instruction-level retry deliver this. It may be assumed that large servers typically used at the back-end layer of the data center's tiered compute model are selected because they are more reliable. Because of a large server's component count, though, it actually has lesser reliability than a middle-tier directed server.
- **Availability**
  The concept of availability is most often associated with the term RAS. As stated before, it has a unique meaning in and of itself. Availability is provided through three key methods. The first is redundancy of components. This does add cost, but is easily justified if the cost of an outage is greater than the cost of redundant components. A second method for greater availability focuses on data availability. This is delivered through memory mirroring, extended-ECC memory (a spare dynamic random access memory (DRAM) on every dual inline memory module (DIMM), and use of Live Migration of virtual machines between servers. Data availability is also provided by use of the ZFS filesystem and implementing mirroring across disk drives. A third feature that delivers greater availability isn't really a hardware feature, but it focuses at the data services level and that is clustering. There is a balance here that has to be made between redundant components in a server that still contains single points of failure, and a clustered solution that focuses on data service availability by constantly monitoring the health of the servers, storage, and networking, as well as the applications.
- **Serviceability**
  This is easier to measure. Components can be either hot-swappable or hot-pluggable to be counted as serviceable while the system is powered on and active. Hot-swap of components is the least disruptive RAS feature because it requires no assistance or preparation of the host operating system and server. Components can simply be replaced while the system continues to run, with no setup required. Another serviceability feature is hot-plug. It is similar to hot-swap in that it also allows the guest operating system to continue running while the component is inserted or removed. However, a hot-plug component does require the operating system and platform be prepared before the component is removed or inserted. Hot-plug components sometimes also require additional setup commands in order to properly use the device. Disk drives are an example of a hot-plug device. Hot plug and hot swap are a necessity of large servers with redundant components, but are used far less

when using a clustered solution. For the Oracle SPARC SuperCluster T4-4 configured for redundancy, individual components like compute nodes or storage nodes can be serviced while the system is running.

Delivering high levels of reliability and availability is achieved through a combination of several system elements:

- Processor
- Memory
- System I/O
- Service processor
- System ASICs
- System power and cooling
- Oracle Solaris operating system
- Oracle Solaris Cluster software

Each RAS element builds on top of the other to deliver the overall RAS capability within a complete system:

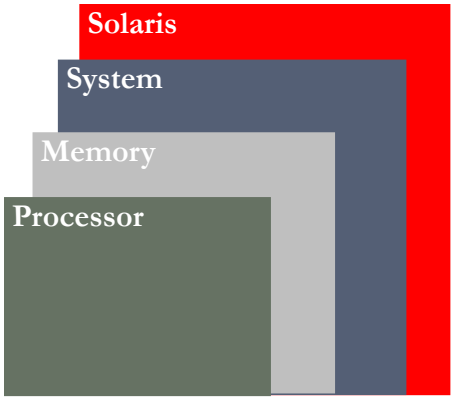

Figure 2. Elements that contribute to RAS

This layering of system elements provides a complete RAS picture for the SPARC T4-4 compute nodes. This paper will also discuss how reliability and availability are further enhanced by Oracle's Sun ZFS Storage Appliance, the Oracle Exadata Storage Servers, as well as the use of RAC for the Oracle Database and using Oracle Solaris Cluster software to maximize application services availability.

If the intent was to only deploy application services on a single server, then this stack might be sufficient. However, the Oracle SPARC SuperCluster system provides for a wide range of applications, as well as databases, which require even higher level of RAS support. Oracle Solaris Cluster is targeted to provide availability for the applications, while RAC is designed for database availability. Even shared file storage that is stored on the Sun ZFS Storage 7320 appliance is protected because the file system used is ZFS, and it provides high data integrity.

Trade-offs

It is important to remember that some features delivering RAS are a trade-off with other features. For example, integrating multiple features into a single chip, or surface-mounting all chips and ASICs to a single motherboard, delivers higher reliability, but with reduced availability (due to mean time to repair) and serviceability. Another example is hot-swap or hot-plug components, which require additional connectors for the components (among other things). While this feature improves availability, it reduces reliability. Selection of such features is always a business decision, since there must be a balance between system cost and data services availability.

While the previous examples were related to hardware, the software layers also impact RAS. First and foremost is Oracle Solaris. SPARC SuperCluster supports both Oracle Solaris 11 and Oracle Solaris 10. The Predictive Self Healing, Fault Management Architecture (FMA) and Service Management Facility (SMF) contribute greatly to the availability of hardware, as well as application data services. This is described in greater detail later on.

As an engineered system, the design of SPARC SuperCluster achieves a balance between reliability and availability. Oracle RAC, along with cluster-related software, Oracle Solaris Cluster, and Oracle Data Guard, a feature of Oracle Database, can be used to extend data service availability across multiple nodes in SPARC SuperCluster.

SPARC SuperCluster T4-4 Components

The Oracle SPARC SuperCluster incorporates Oracle Exadata storage technology for enhancing the performance of the Oracle Database. SPARC SuperCluster also incorporates optional Oracle Exalogic Elastic Cloud Software technology to accelerate performance of Java middleware and applications, general-purpose applications, and the Oracle Database 11*g* Release 2. It is an engineered system designed to host the entire Oracle software solution stack. Third-party ISV software, as well as customer-developed software, is supported in general-purpose domains. In addition to the Oracle Exadata Storage Servers built in to SPARC SuperCluster, Oracle's Exadata Storage Expansion Racks can be used to add capacity and bandwidth to the system.

SPARC SuperCluster T4-4 is designed to fully leverage an internal InfiniBand fabric that connects all of the processing and storage components within SPARC SuperCluster T4-4 to form a single, large computing device. Each SPARC SuperCluster T4-4 is connected to the customer's data center networks via 10 GbE (client access) and 1 GbE (management) interfaces.

The SPARC SuperCluster system is a complete, pre-configured clustered solution utilizing the following components:

- **SPARC T4-4 servers** – Oracle's SPARC T4-4 server offers a large memory capacity and a highly integrated design that supports virtualization and consolidation of mission-critical applications. There are two SPARC T4-4 servers in half-rack SPARC SuperCluster configurations and four in the full rack. Each SPARC T4-4 server has the following:
  - **Four SPARC T4 processors** – Each processor comes with eight cores and eight threads per core. The SPARC T4 processor utilizes the latest advanced S3 core design. In conjunction with the

Oracle Solaris Fault Management Architecture feature, threads and cores can be proactively taken offline if too many correctable errors are detected. This provides for greater domain and application uptime.

  - **1 TB of memory** – Sixty-four of the latest 16 GB DDR3 memory DIMMs. Each DIMM is ECC protected and supports extended-ECC, a technology similar to IBM's ChipKill technology. Extended-ECC protects against single-bit errors by providing a redundant DRAM on every DIMM. If a page of memory is generating too many correctable errors, then the Fault Management Architecture will initiate a process to move the data to more reliable pages of memory and offline the suspect pages of memory. Lane failover is also provided for greater memory availability.
  - **Eight Disk drives** –Six 600 GB SAS2 disk drives and two 300 GB SSD disk drives are included.
  - **Oracle's Sun PCIe Dual Port QDR InfiniBand Host Channel Adapters** – Low latency 40 Gb/s InfiniBand HCA in a modular hot-pluggable PCI Express (PCIe) ExpressModule form factor. There are four InfiniBand (IB) cards in each SPARC T4-4 server. Each adapter is dual ported for redundancy.
  - **Oracle's Sun Dual 10 GbE SFP+ PCIe 2.0 Low Profile network interface cards** – These network interface cards (NICs) provide client access to each SPARC T4-4 server. They are in a modular hot-pluggable PCIe ExpressModule form factor. Four 10 GbE cards are included in each SPARC T4-4 server. Each adapter is dual ported for redundancy.

- **Sun ZFS Storage 7320 appliance** – Providing 60 TB of capacity, this appliance uses Flash-enabled Hybrid Storage Pools, a feature of Sun ZFS Storage Appliances, to improve application response times. It is configured with dual controllers for maximum availability. Each server also utilizes redundant power and fans, as well as ECC memory.
- **Oracle Exadata Storage Servers** – Provides three critical technologies to improve Oracle Database 11gR2 performance: Exadata Smart Scan, Exadata Smart Flash Cache, and Exadata Hybrid Columnar Compression. There are three Oracle Exadata Storage Servers in SPARC SuperCluster half-rack configurations and six in the full rack. Oracle Exadata Storage Servers can be set up in a mirrored, or even triple mirror, configuration for data redundancy. If the database is configured in an Oracle RAC configuration, then additional redundancy is achieved. For failover support, Oracle Clusterware should be used. Oracle Clusterware is portable cluster software that allows clustering of independent servers so that they cooperate as a single system. Oracle Clusterware is an independent server pool infrastructure, which is fully integrated with Oracle RAC, capable of protecting data access in a failover cluster.
- **Oracle Exalogic Elastic Cloud Software** – Provides extreme performance for Java applications, Oracle Applications, and all other enterprise applications, and reduces application implementation and ongoing costs versus traditional enterprise application platforms and private clouds assembled from separately sourced components. Application availability is maintained by using Oracle Solaris Cluster between the SPARC T4-4 nodes.

- **Sun Datacenter InfiniBand Switches** – Provide a high throughput, low latency, and scalable fabric suitable for fabric consolidation of inter-process communication, network and storage. InfiniBand delivers up to 63 percent more transactions per second for Oracle RAC over GbE networks. There are three InfiniBand switches in SPARC SuperCluster offering private connectivity within the system. Two of the switches provide redundant connectivity to the SPARC T4-4 nodes, the Sun ZFS Storage 7320, and the Oracle Exadata Storage Servers. See Figure 3 for a diagram of the InfiniBand fabric inside a half-rack SPARC SuperCluster. Each IB card in every SPARC T4-4 node, the controllers on the Sun ZFS Storage 7320, and the Oracle Exadata Storage Server are connected to the two IB leaf switches for redundant data access.

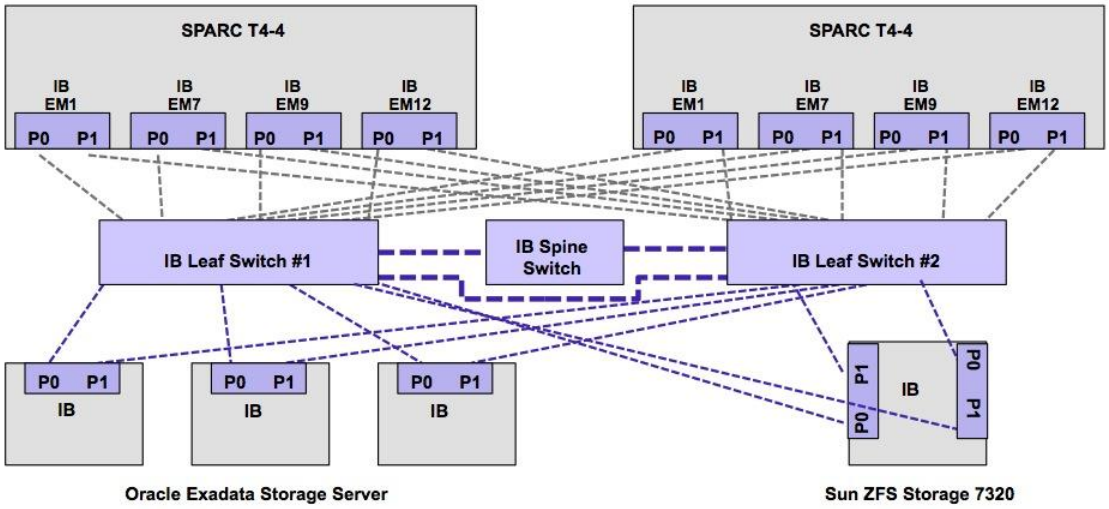

Figure 3. InfiniBand fabric in a half-rack SPARC SuperCluster

- **Ethernet Management Switch** – Provides network management connectivity to the management ports on all servers and switches used in SPARC SuperCluster. (A Cisco switch is provided although customers can substitute their own switch if desired.)
- **Integrated Virtualization -** Enhanced security, increased utilization, and improved reliability are delivered through Oracle Solaris Zones, a feature of Oracle Solaris, and through Oracle VM Server for SPARC (previously known as Logical Domains, or LDoms). Most important, Oracle Solaris Cluster can be used on any combination of logical domains and Oracle Solaris Zones.
- **Oracle Enterprise Manager Ops Center -** Oracle Enterprise Manager Ops Center delivers a converged hardware management solution for SPARC SuperCluster T4-4 that integrates management across the infrastructure stack to help IT managers deploy and manage SPARC SuperCluster T4-4 more efficiently.

SPARC SuperCluster configurations are fixed and contain a specific combination of servers, storage, network, and software elements. This architecture ensures the highest quality for integration and testing during production.

SPARC T4-4 Servers

The SPARC T4-4 server provides high throughput and computing density along with built-in virtualization and extreme scalability and is a highly efficient platform for deploying large scale, mission-critical applications.

Architected to reduce planned and unplanned downtime, the SPARC T4-4 server includes advanced reliability, availability, and serviceability capabilities to avoid outages and reduce recovery time. These design elements are vital for mission-critical systems such as SPARC SuperCluster. Design features that boost the reliability of SPARC T4-4 server include:

- **Advanced CPU integration** — The SPARC T4 is an eight-core processor, with each core featuring eight threads. Each processor provides 4 MB of L3 cache, critical in reducing the time required to read and process data. Each SPARC T4-4 server in SPARC SuperCluster comes with the maximum of four SPARC T4 processors.
- **Extended-ECC Memory** — The memory system has redundant components built on to each memory DIMM that allows a DIMM to continue to operate even with partial failure.
- **Fault-resilient power options and redundant components** — Systems feature redundant power supply and fan units. Redundant storage can be created using disk mirroring.
- **Hardware redundancy** — Redundant power, redundant fans, and redundant data paths.

The dual-port 10 GbE cards will be used by clients for datacenter access to SPARC SuperCluster compute nodes. In addition to the 10 GbE cards, also included are four Quad Data Rate (QDR) InfiniBand cards. These cards are used for internal communication and access to the Oracle Exadata Storage Servers and the Sun ZFS Storage 7320 appliance. The cards are dual-ported for redundant communications.

Applications that run on the SPARC T4-4 server run in one of two types of logical domains:

1. **Database Domain**: A domain dedicated to running Oracle Database 11*g* Release 2, using Oracle Exadata Storage Servers for database storage. This domain must run Oracle Solaris 11.
2. **Application Domain**: A domain dedicated to running applications on either Oracle Solaris 11 or Oracle Solaris 10. Application Domains running Oracle Solaris 10 also support the use of legacy Oracle Solaris Zones, for applications that require either a Solaris 8 or 9 environment. Application Domains running Oracle Solaris 11 also support the use of Oracle Solaris 10 branded zones. The optional Oracle Exalogic Elastic Cloud software run only in an Application Domain on Oracle Solaris 11.

Platform management is done via the service processor, or Oracle Integrated Lights Out Manager (Oracle ILOM 3.0). Oracle Integrated Lights Out Manager provides a command-line interface (CLI), a Web-based graphical user interface (GUI), and Intelligent Platform Management Interface (IPMI)

functionality to aid out-of-band monitoring and administration. The management software, Oracle Enterprise Manager Ops Center, communicates with the Oracle ILOM to manage and monitor the SPARC T4-4 servers. All system telemetry and health diagnostics are recorded by the Oracle ILOM and forwarded to Oracle Enterprise Manager Ops Center for further analysis and action. If a service event is deemed necessary, Oracle Enterprise Manager Ops Center works with Oracle Auto Service Request (ASR) to notify service that action needs to be taken.

Oracle Exadata Storage Servers

Oracle Automatic Storage Management is used as the file system and volume manager for Oracle Exadata. Oracle Automatic Storage Management virtualizes the storage resources and provides the advanced volume management and file system capabilities of Oracle Exadata. Striping database files evenly across the available Oracle Exadata cells and disks results in uniform I/O load across all the storage hardware. The ability of Oracle Automatic Storage Management to perform non-intrusive resource allocation, and reallocation, is a key enabler of the shared grid storage capabilities of Oracle Exadata environments. The disk mirroring provided by Oracle Automatic Storage Management, combined with hot-swappable Oracle Exadata disks, ensure the database can tolerate the failure of individual disk drives. Data is mirrored across cells to ensure that the failure of a cell will not result in loss of data, or inhibit data accessibility. This massively parallel architecture delivers unbounded scalability and high availability.

Oracle's Hardware Assisted Resilient Data (HARD) Initiative is a comprehensive program designed to prevent data corruptions before they happen. Data corruptions are very rare, but when they happen, they can have a catastrophic effect on a Oracle Database, and therefore a business. Oracle Exadata Storage Servers have enhanced HARD functionality embedded in it to provide even higher levels of protection and end-to-end data validation for data. Oracle Exadata Storage Servers perform extensive validation of the data stored in them including checksums, block locations, magic numbers, head and tail checks, alignment errors, etc. Implementing these data validation algorithms within Oracle Exadata will prevent corrupted data from being written to permanent storage. Furthermore, these checks and protections are provided without the manual steps required when using HARD with conventional storage.

Oracle Data Guard

Oracle Data Guard is the software feature of Oracle Database that creates, maintains, and monitors one or more standby Oracle databases to protect your database from failures, disasters, errors, and corruptions. Oracle Data Guard works unmodified with Oracle Exadata and can be used for both production and standby databases. By using Oracle Active Data Guard with Oracle Exadata storage, queries and reports can be offloaded from the production database to an extremely fast standby database and ensure that critical work on the production database is not impacted while still providing disaster protection.

Oracle Recovery Manager (Oracle RMAN)

SPARC SuperCluster T4-4 works with Oracle Recovery Manager (Oracle RMAN) to allow efficient Oracle Database backup and recovery. All existing Oracle RMAN scripts work unchanged in the SPARC SuperCluster T4-4 environment. Oracle RMAN is designed to work intimately with the server, providing block-level corruption detection during backup and restore. Oracle RMAN optimizes performance and space consumption during backup with file multiplexing and backup set compression. The integrated Sun ZFS Storage 7320 appliance could be used by Oracle RMAN to backup the Oracle Database, or an external Sun ZFS Storage Appliance could be connected to the SPARC SuperCluster T4-4.

Oracle Solaris

SPARC SuperCluster supports both Oracle Solaris 10 and Oracle Solaris 11. Oracle Solaris 11 is used exclusively for the Database Domains, and can also be used in an Application Domain. Those applications and third-party applications that are not yet ready for Oracle Solaris 11 can use an Oracle Solaris 10 Application domain.

Oracle Solaris includes the following features that strengthen the reliability and availability of SPARC Supercluster:

- **Oracle Solaris Fault Management Architecture** — Oracle Solaris Fault Management Architecture feature reduces complexity by automatically diagnosing faults in the system and initiating self-healing actions to help prevent service interruptions. This software helps increase availability by configuring problem components out of a system before a failure occurs—and in the event of a failure, this feature initiates automatic recovery and application restart using Oracle Solaris Service Management Facility feature. The Fault Management Architecture diagnosis engine produces a fault diagnosis once discernible patterns are observed from a stream of incoming errors. Following diagnosis, the Fault Management Architecture provides fault information to agents that know how to respond to specific faults. While similar technology exits with other hardware and software vendors, each of their solutions is either software, or hardware-based fault detection. Oracle Solaris provides a fault detection and management environment that allows full integration of hardware fault messages to be passed on to the operating system, and the operating system services can adjust hardware resources as needed so outages are greatly reduced.
- **Oracle Solaris Service Management Facility** — With Service Management Facility, system administrators can use simple command line utilities to easily identify, observe, and manage the services provided by the system and the system itself. Service Management Facility describes the conditions under which failed services may be automatically restarted. These services can then be automatically restarted if an administrator accidentally terminates them, if they are aborted as the result of a software programming error, or if they are interrupted by an underlying hardware problem. Other operating systems on the market today use either a monolithic start up script, or a series of smaller scripts that are executed sequentially. They cannot provide a dependency between scripts, or a restart of services when problems are corrected.

- **Oracle Solaris ZFS** — Oracle Solaris ZFS provides unparalleled data integrity, capacity, performance, and manageability for storage. ZFS provides high resiliency features, such as metadata logging to guarantee data integrity and speed recovery in the event of system failure. What differentiates Oracle Solaris ZFS from other competitive file system offerings is strong data integrity and high resiliency. Oracle Solaris ZFS dramatically simplifies file system administration to help increase protection against administrative error. Oracle Solaris ZFS uses techniques such as copy-on-write and end-to-end checksumming to keep data consistent and eliminate silent data corruption. Because the file system is always consistent, time-consuming recovery procedures like fsck are not required if the system is shut down in an unclean manner. In addition, data is read and checked constantly to help ensure correctness, and any errors detected in a mirrored pool are automatically repaired to protect against costly and time-consuming data loss and previously undetectable silent data corruption. Corrections are made possible by a RAID-Z implementation that uses parity, striping, and atomic operations to aid the reconstruction of corrupted data.

  Oracle Solaris ZFS protects all data by 256-bit checksums, resulting in 99.99999999999999999 percent error detection and correction. Oracle Solaris ZFS constantly reads and checks data to help ensure it is correct, and if it detects an error in a storage pool with redundancy (protected with mirroring, ZFS RAIDZ, or ZFS RAIDZ2), Oracle Solaris ZFS automatically repairs the corrupt data. It optimizes file system reliability by maintaining data redundancy on commodity hardware through the delivery of basic mirroring, compression, and integrated volume management.

- **Oracle Solaris DTrace** — DTrace, a feature of Oracle Solaris, is a dynamic tracing framework for troubleshooting systemic problems in real time on production systems. DTrace is designed to quickly identify the root cause of system performance problems. DTrace safely and dynamically instruments the running operating system (OS) kernel and running applications without rebooting the kernel and recompiling—or even restarting—applications. This design greatly improves service uptime.

The Predictive Self Healing and fault management combination of Fault Management Architecture and Service Management Facility can offline processor threads or cores in faults, retire suspect pages of memory, log errors or faults from I/O, or deal with any other issue detected by the server.

Sun ZFS Storage 7320 Appliance

The combination of multithreaded Oracle Solaris operating system with the powerful multicore Intel Xeon processor family, Sun ZFS Storage Appliances deliver scalable performance, innovative data services, advanced reliability, and high-energy efficiency. Today's vast repositories of data represent valuable intellectual property that must be protected to ensure the livelihood of the enterprise. While backup and archival capabilities are designed to prevent catastrophic failures, they do not necessarily protect against silent data corruption and they offer no protection against data theft. The Sun ZFS Storage Appliance brings together the data center reliability and security features of both Oracle Solaris and Intel Xeon processors.

To keep the file system internally consistent, Oracle Solaris ZFS combines a copy-on-write approach

(data is written to a new block on the media before the pointers to the data are changed and the write is committed) with end-to-end checksumming (explained below). Because the file system is always consistent, time-consuming recovery procedures are not required if the system is shut down in an unclean manner, thus saving time for administrators and improving service levels for users.

Oracle makes extensive use of the Intel Xeon Machine Check Architecture Recovery (MCA Recovery). MCA Recovery enables Sun ZFS Storage Appliances to detect and correct errors in memory and cache that were previously "uncorrectable" through ECC memory or other means. MCA Recovery accomplishes this by first detecting and containing errors before the data is consumed by an application. Then it works with Oracle Solaris Predictive Self Healing to determine the best course of action to keep the system and applications running.

Oracle Solaris Cluster

For applications run in an Application Domain, Oracle Solaris Cluster helps organizations deliver highly available application services. To limit outages due to those single points of failure, mission-critical services need to be run in clustered physical servers that efficiently and smoothly take over the services from failing nodes, with minimal interruption to data services. While SPARC SuperCluster is designed with full redundancy at the hardware level, Oracle Solaris Cluster provides the best HA solution for Oracle SPARC servers running Oracle Solaris and applications. Oracle Solaris Cluster is focused on failover between zones and Application Domains within SPARC SuperCluster. Tightly coupled with Oracle Solaris, Oracle Solaris Cluster detects failures without delay (zero-second delay), provides much faster failure notification, application failover, and reconfiguration time. Oracle Solaris Cluster offers high availability for today's complex solution stacks, with failover protection from the application layer through to the storage layer. Figure 4 shows an example of the time it takes Oracle Solaris Cluster to detect a failure and recover on the redundant node. Oracle Solaris Cluster is compared to another vendor, which offers a similar cluster software solution.

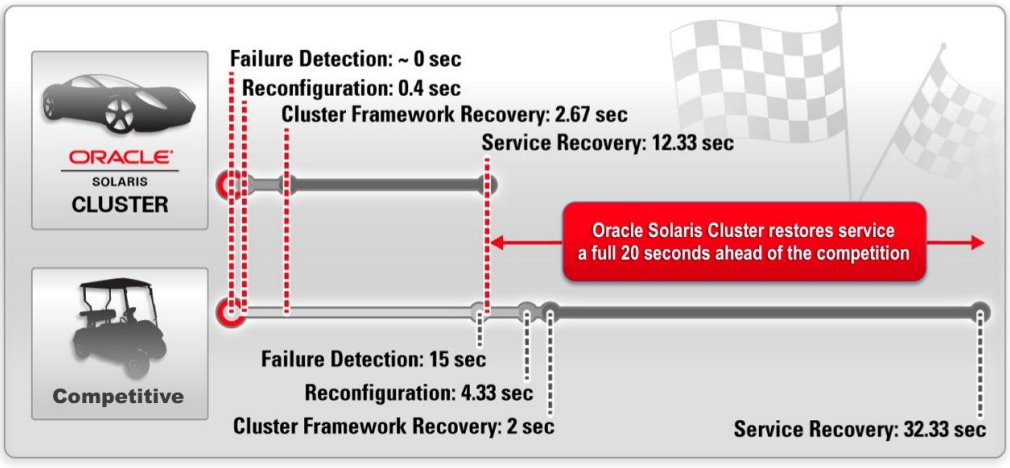

Figure 4. Oracle Solaris Cluster recovery times

Use of Oracle Solaris Cluster achieves much faster resumption of IT services. Features in Oracle Solaris Cluster that speed this process on SPARC SuperCluster include:

- Integrates tightly with the Predictive Self Healing framework and supports applications controlled by Service Management Facility in Oracle Solaris Zones and logical domains.
- Makes extensive use of Oracle's storage management and volume management capabilities.
- Supports Oracle Solaris ZFS as a failover file system and as a boot file system, allowing the use of ZFS storage as the single file system type used.
- Leverages ZFS features such as pooled storage, built-in redundancy, and data integrity.
- Integrates with Oracle Enterprise Manager Ops Center.

Here is a typical cluster configuration for a full-rack SPARC SuperCluster:

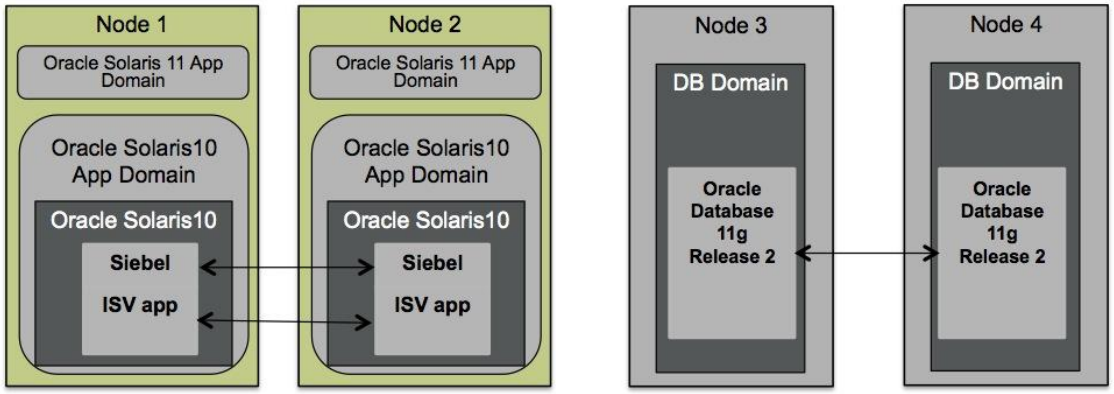

Figure 5. Application and Oracle Database Clustering

While beyond the scope of this paper, customers should consider the requirements for replication and disaster recovery between SPARC SuperCluster racks. The racks can either be co-located in the same data center, or to remote data centers.

Integrated System Monitoring

SPARC SuperCluster T4-4 provides comprehensive monitoring and notifications to enable administrators to proactively detect and respond to problems with hardware and software components. With direct connectivity into the hardware components of SPARC SuperCluster T4-4, Oracle Enterprise Manager Ops Center can alert administrators to hardware-related faults and log service requests automatically through integration with Oracle Automatic Service Requests (ASR) for immediate review by Oracle Customer Support. Problems that would have required a combination of database, system, and storage administrators to detect them in traditional systems can now be diagnosed in minutes because of integrated systems monitoring for the entire SPARC SuperCluster T4-4.

Oracle Configuration Manager

Oracle Configuration Manager is used to personalize the support experience by collecting configuration information and uploading it to the management repository. When customer configuration data is uploaded on a regular basis, customer support representatives can analyze this data and provide better service to the customers.

Some of the benefits of using Oracle Configuration Manager are as follows:

- Reduces time for resolution of support issues
- Provides proactive problem avoidance
- Improves access to best practices and the Oracle knowledge base
- Improves understanding of customer's business needs and provides consistent responses and services

Figure 6 shows the possible architecture configurations of Oracle Configuration Manager, depending on customer requirements and restriction on connectivity to the Internet.

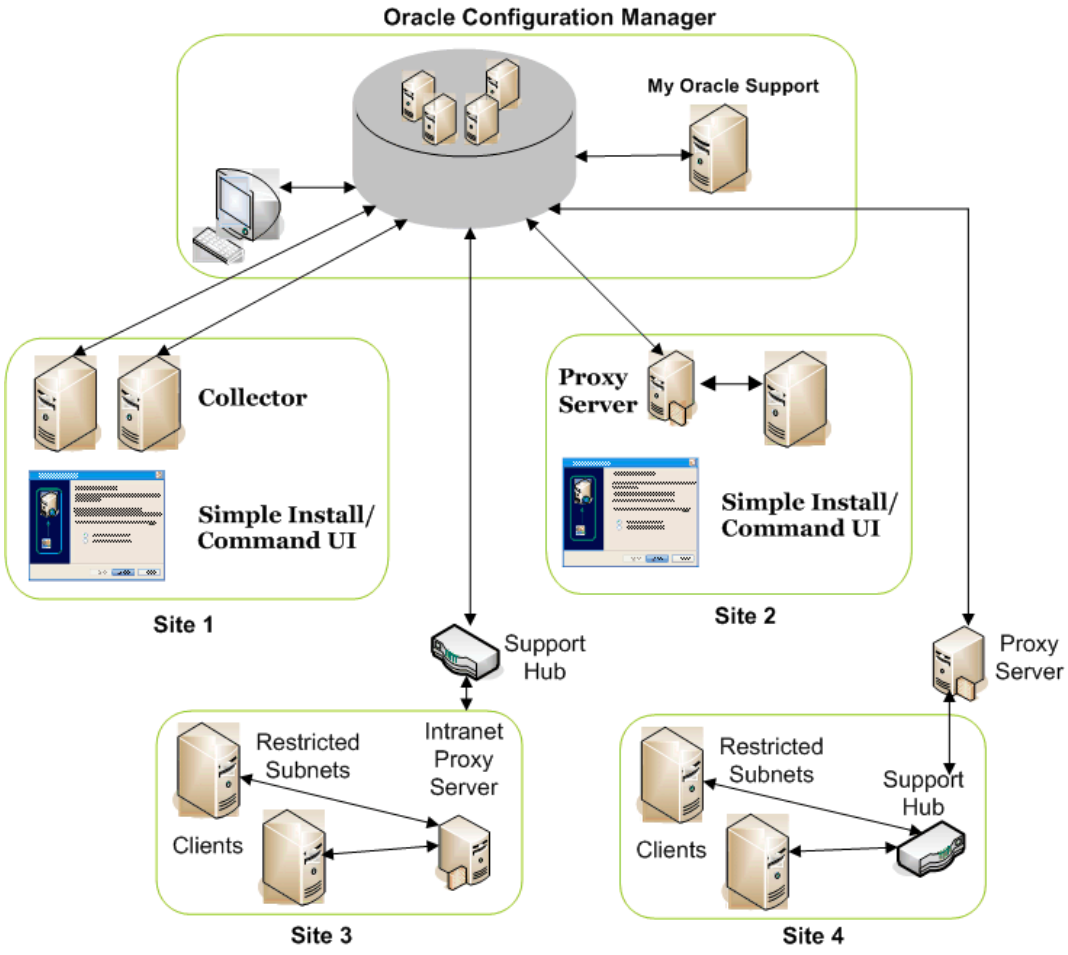

Figure 6. Oracle Configuration Manager Architecture

- **Site 1:** Systems that are directly connected to the Internet.
- **Site 2:** Systems that are connected to the Internet through a proxy server.
- **Site 3:** Systems that do *not* have direct access to the Internet but do have access to an Intranet proxy server, which in turn has Internet connection through the Support Hub.
- **Site 4:** Systems that do *not* have direct access to the Internet but do have access to the Support Hub, which in turn is connected to the Internet through a proxy server.

Oracle Enterprise Manager Ops Center 12*c*

Oracle Enterprise Manager Ops Center 12*c* helps IT staff understand and manage every architectural layer—from bare metal to operating systems and applications. It provides a centralized interface for physical and virtual machine lifecycle management, from power-on to decommissioning. In addition, it offers IT administrators a unique insight into the user experience, business transactions, and business services, helping administrators quickly detect changes in system health and troubleshoot issues across the entire environment.

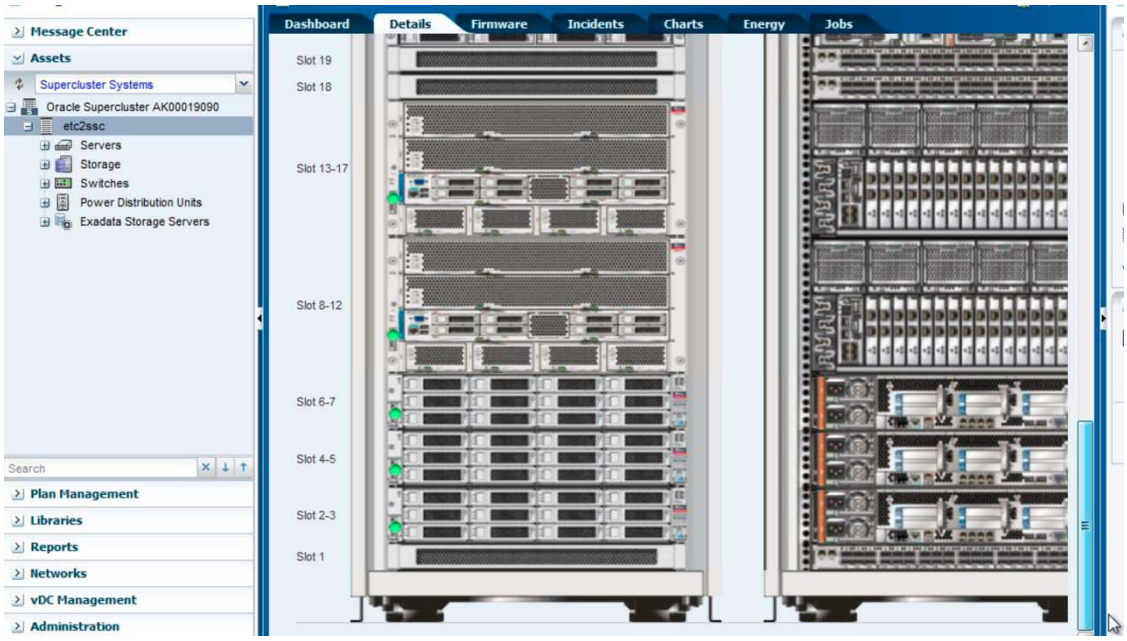

Figure 7. Oracle Enterprise Manager Ops Center 12*c* User Interface

Oracle Enterprise Manager Ops Center 12*c* is designed to discover and manage the SPARC SuperCluster T4-4 as a whole system, and not just a collection of components in a rack. All hardware and software events are consolidated into Oracle Enterprise Manager Ops Center 12*c* for a single system view of events. These events are then sent to My Oracle Support for further analysis and if needed, proactive action is taken. Oracle Enterprise Manager Ops Center 12*c* delivers holistic fault management for events coming from the SPARC T4-4 servers, Infiniband switches, Oracle Exadata Storage Servers, the Sun ZFS Storage 7320 appliance, and even Oracle Solaris Cluster events are gathered.

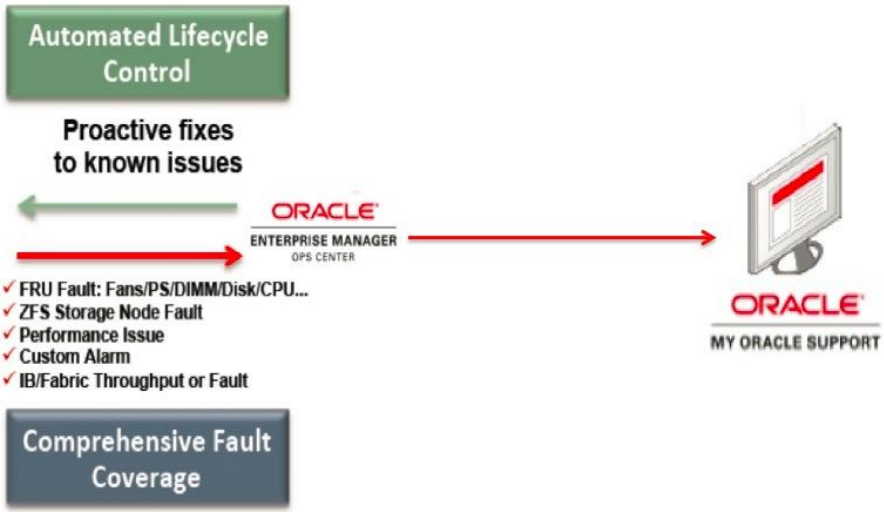

Figure 8. Oracle Enterprise Manager Ops Center 12*c* Phone Home Capability

Automated Service Requests (ASR)

Once a problem is detected in a business-critical system, every minute until the problem is resolved is costly. Oracle Enterprise Manager Ops Center 12*c* provides deep connections to My Oracle Support systems and processes to enable automatic problem detection, analysis, automated service requests, and access to the Oracle knowledge base and community for optimal problem resolution.

ASR is a feature of the Oracle hardware warranty and Oracle Premier Support for Systems. ASR resolves problems faster by automatically opening service requests when specific hardware faults occur. Oracle's ASR is integrated with My Oracle Support. Customers must use My Oracle Support to activate ASR Assets. Oracle Platinum Services provides remote fault monitoring with faster response times and patch deployment services to qualified Oracle Premier Support customers—at no additional cost.

Oracle Maximum Availability Architecture

Oracle Maximum Availability Architecture is a framework and set of best practices for maximizing systems availability and meeting the most aggressive service level agreements (SLAs) for system availability, quality of service, and data protection.

In addition to the best practices of Oracle Maximum Availability Architecture, the Oracle Optimized Solutions program offers best practices for setting up the Oracle SPARC SuperCluster T4-4 in High Availability (HA) and Disaster Recovery (DR) configurations. Visit the Oracle Optimized Solutions home page for more details, at http://www.oracle.com/us/solutions/oos/overview/index.html.

Conclusion

SPARC SuperCluster T4-4 is designed to help IT organizations consolidate multiple workloads in an environment that has been optimized for performance and RAS. SPARC SuperCluster offers high levels of reliability and availability in both hardware and multiple layers of software. It strikes a perfect balance between the cost of hardware redundancy and the flexibility of software for data service availability.

SPARC SuperCluster: Taking Reliability and Availability to the Next Level
December 2012
Version 2.0
Author: Gary Combs

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

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

Hardware and Software, Engineered to Work Together