An Oracle Technical White Paper
February 2011

Oracle's Netra SPARC T3-1 Server Architecture

Introduction

Many telecommunications providers are moving toward a consolidated framework to converge telecom, Internet, and multimedia services using a single standards-based architecture. However, expanding voice and data services and escalating transaction volumes are putting a strain on existing IT infrastructures.

Even as the IT demand increases, economic realities are reshaping the requirements for the IT infrastructure. Real-estate constraints—along with rapidly rising energy costs—are significant factors that discourage the past practice of adding racks upon racks of servers. The cost and complexity of managing large numbers of systems creates formidable challenges, especially because low utilization levels can sometimes result, lowering the return on investment. Exacting service-level agreements (SLAs) are increasingly common because service availability is often closely tied to economic success.

In managing the IT infrastructure, organizations are looking to control capital and operational costs more precisely and are making deployment decisions that take into account overall system lifecycle and total cost of ownership (TCO). Oracle's Netra SPARC T3-1 server can help telecommunications companies address these challenges by providing extreme scalability, energy efficiency, and carrier-grade reliability in a compact 2RU rack-mountable server.

Oracle's Netra SPARC T3-1 Server

The Netra SPARC T3-1 server offers unprecedented processor core, memory, and I/O densities to enable dramatic expansion and configuration flexibility within a small, compact 2U form factor. CoolThreads chip multithreading (CMT) technology in the SPARC T3 processor helps to deliver high throughput for demanding telecommunications applications while minimizing power consumption.

The SPARC T3 processor in the Netra SPARC T3-1 server is Oracle's third-generation CMT processor design. It leverages the same CoolThreads technology proven in the Netra T5220 server based on the previous-generation UltraSPARC T2 processor. The SPARC T3 processor in the Netra SPARC T3-1 server—with up to 16 cores and 128 concurrent processor threads—takes throughput-per-watt to a new level, doubling the throughput of the earlier UltraSPARC T2 processor. It adds on-chip networking and improved floating-point capabilities to address the demands of intensive packet processing and multimedia processing workloads.

In addition to breakthrough performance and scalability, the Netra SPARC T3-1 server is certified to meet Network Equipment Building Specification (NEBS) Level 3 requirements for environments that demand continuous availability and simplified management. NEBS Level 3 certification (not merely compliance) indicates that the system has undergone a series of rigorous, independently conducted tests to prove that it can continue to operate under severe environmental conditions.

Numerous architectural features—including key redundant and hot-swappable components—enable the Netra SPARC T3-1 server to deliver outstanding levels of availability and reliability that contribute to continuous system operation. Redundant hot-swappable DC or AC power supplies and hot-pluggable disk drives, for example, help to enhance system uptime. A built-in service processor with Integrated Lights Out Manager (ILOM) capabilities enables remote monitoring and system management, providing the kind of reliability and serviceability needed by telecommunications companies or military organizations that operate in severe environments. The service processor is designed around an ASPEED Technology AST2200 microprocessor.

SPARC T3 Processor Architecture

The SPARC T3 processor extends Oracle's CMT initiative with an elegant and robust architecture that delivers real performance to applications. This section provides an overview of the SPARC T3 processor architecture. Additional details can be found on the web at:

- www.oracle.com/us/products/servers-storage/servers/sparc-enterprise/t-series/sparc-t3-171613.html
- www.oracle.com/technetwork/articles/systems-hardware-architecture/sparc-t3-server-architecture-176017.pdf.

Figure 1 provides a block-level diagram of the SPARC T3 processor.

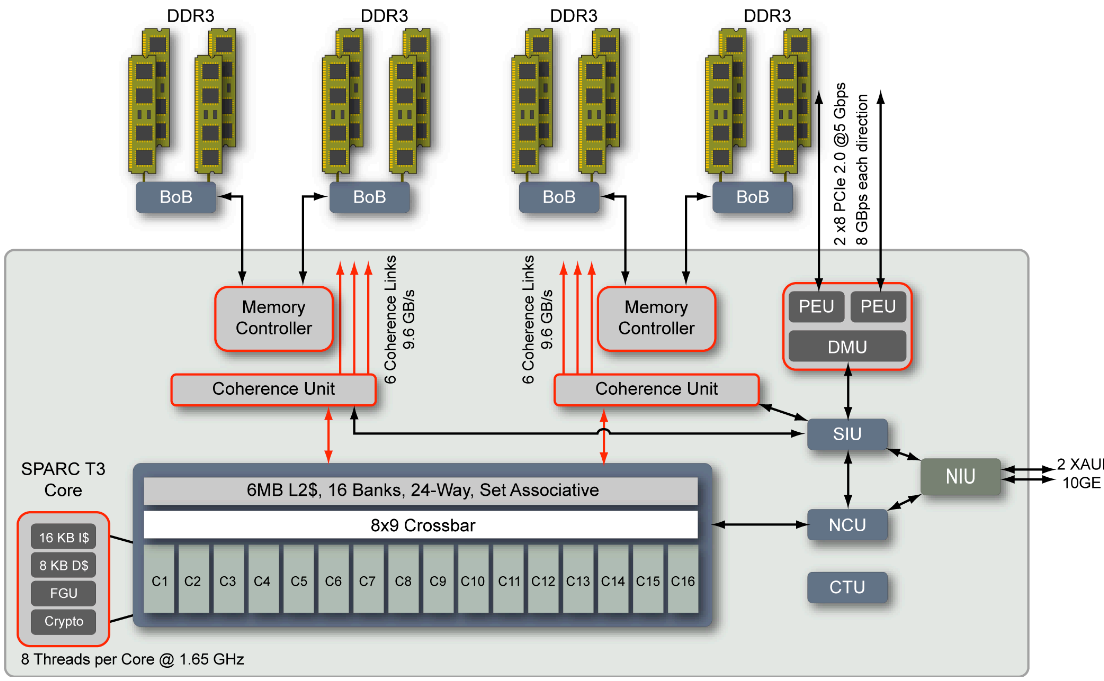

Figure 1. The SPARC T3 processor combines sixteen cores, memory management, cryptographic support, 10-GbE, and PCIe on a single chip.

The SPARC T3 processor has coherence link interfaces to allow communication between up to four SPARC T3 processors in a system without requiring any external hub chip. There are six coherence links, each with 14 bits in each direction running at 9.6 Gbps. Each frame has 168 bits. Therefore, the maximum frame rate is 800M frames per second. The SPARC T3 processor has two coherence link controllers. Each includes two Coherence and Ordering Units (COU), three Link Framing Units (LFU) and a crossbar (CLX) between COUs and LFUs. Each COU interfaces to two L2 bank pairs. The coherence links run a cache coherence (snoopy) protocol over an FB-DIMM like physical interface. The memory link speed of the SPARC T3 processor was increased to 6.4 Gb/sec over the UltraSPARC T2 Plus processor's 4.8 Gb/sec, and 4.0 Gb/sec of the UltraSPARC T2 processor.

The SPARC T3 processor can have up to 16 cores and can support one-, two- and four-socket implementations. Each processor core provides an eight-stage integer pipeline and a 12-stage floating-point pipeline as shown in Figure 2. A new pipeline stage called “pick” has been added to choose two threads (out of the eight possible per core) to execute each cycle.

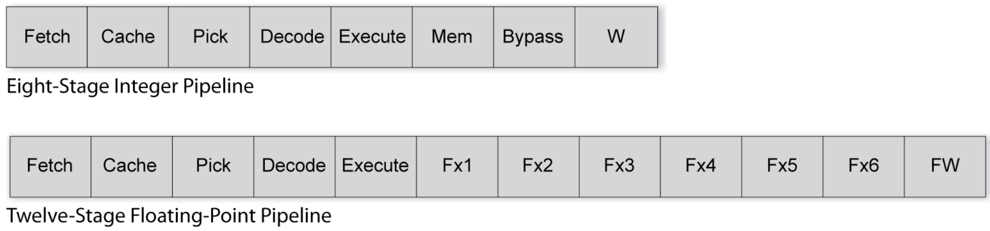

Figure 2. An 8-stage integer pipeline and a 9-stage floating-point pipeline are provided by the SPARC T3 processor core.

To illustrate how the dual pipelines function, Figure 2 depicts the integer pipeline with the load-store unit (LSU). The instruction cache is shared by all eight threads within the core. A least-recently-fetched algorithm is used to select the next thread to fetch. Each thread is written into a thread-specific instruction buffer (IB), and each of the eight threads is statically assigned to one of two thread groups within the core.

The pick stage chooses one thread each cycle within each thread group. Picking within each thread group is independent of the other, and a least-recently-picked algorithm is used to select the next thread to execute. The decode state resolves resource conflicts that are not handled during the pick stage. As shown in Figure 3, threads are interleaved between pipeline stages with very few restrictions. Any thread can be at the fetch or cache stage, before being split into either of the two thread groups. LSUs and FPUs are shared among all eight threads. Only one thread from either thread group can be scheduled on such a shared unit.

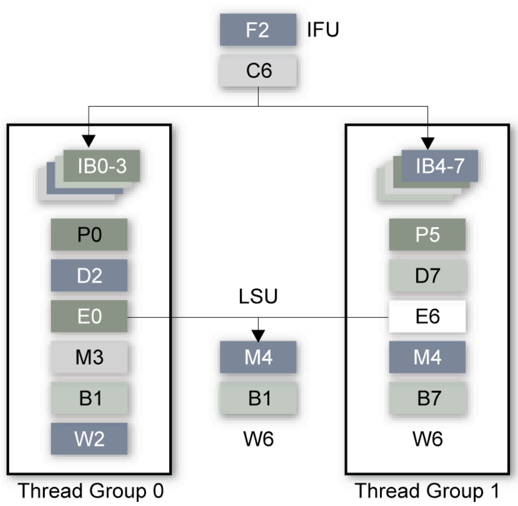

Figure 3. Threads are interleaved between pipeline stages with very few restrictions (integer pipeline shown; letters depict pipeline stages, and numbers depict different scheduled threads).

Integrated Networking

By providing integrated on-chip networking, the SPARC T3 processor is able to provide better networking performance for applications with intensive packet-processing requirements. All network data is supplied directly from and to main memory. Placing networking so close to memory reduces latency, provides higher memory bandwidth, and eliminates inherent inefficiencies of I/O protocol translation.

The SPARC T3 processor provides two 10-GbE ports with an integrated serializer/deserializer, providing line-rate packet classification at up to 30 million packets/sec (based on Layers 1 through 4 of the protocol stack). Multiple DMA engines (16 transmit and 16 receive DMA channels) match DMAs to individual threads, providing binding flexibility between ports and threads. Virtualization support includes provisions for eight partitions, and interrupts may be bound to different hardware threads.

Stream-Processing Unit (SPU)

The SPU on each SPARC T3 processor core runs in parallel with the core at the same frequency. The cipher/hash unit provides RC4, DES/3DES, AES-128/192/256, MD5, SHA-1, and SHA-256. The SPU is designed to achieve wire-speed encryption and decryption on both of the processor's 10-GbE ports.

Integral PCI-Express Gen2 Support

SPARC T3 processors provide dual on-chip PCIe Generation 2 interfaces. Each operates at 5 Gtps per x1 lane bi-directionally through a point-to-point dual-simplex chip interconnect. This means that each x1 lane consists of two uni-directional bit-wide connections, one for northbound and the other for southbound traffic. An integral IOMMU supports I/O virtualization and process device isolation by using the PCIe BUS/Device/Function (BDF) number. The total theoretical I/O bandwidth (for a x8 lane) is 4 GB/sec, with a maximum payload size of 256 bytes per PCIe Gen 2 interface. The actual realizable bandwidth is more likely to be approximately 2.8 GB/sec. An x8 SerDes interface is provided for integration with off-chip PCIe switches.

Power Management

Netra SPARC T3-1 server power management features are implemented utilizing several concurrent approaches. The most effective power management features are embedded in the SPARC T3 processor, which supports multiple cores and threads. There are also several other areas available at the chip, device, and system level where power savings can be achieved. These features include reduced instruction rates, parking of idle threads and cores, and ability to turn off clocks in both cores and memory to reduce power consumption. Substantial innovation is present in the areas of CPU configuration settings that offer extensive power management features:

- Thread parking and core disabling provides estimated power savings of around 1.75W per core.
- Power throttling allows extra stall cycles to be injected into the decode stage.
- CPU down clock provides estimated power savings per down clock step of about 9W-14W.

Netra SPARC T3-1 Server Architecture

The Netra SPARC T3-1 server leverages the SPARC T3-1 sever motherboard (see Figure 4) and uses the SPARC T3 processor, which is a multi-core multi-threaded implementation of the SPARC V9 architecture. The SPARC T3 processor includes 16 SPARC cores, each capable of running 8 threads concurrently, 8 floating-point units (FPU), 4 memory controllers, an L2 cache, and a central crossbar.

The SPARC T3 processor has the following major differences from the Ultra SPARC T2 processor:

- 2 FBDIMM2 memory controllers
- 2 PCI Express Gen2 interfaces
- Included Network Interface Unit (NIU) with 2 10Gbps XAUI interfaces
- Enhanced power management features

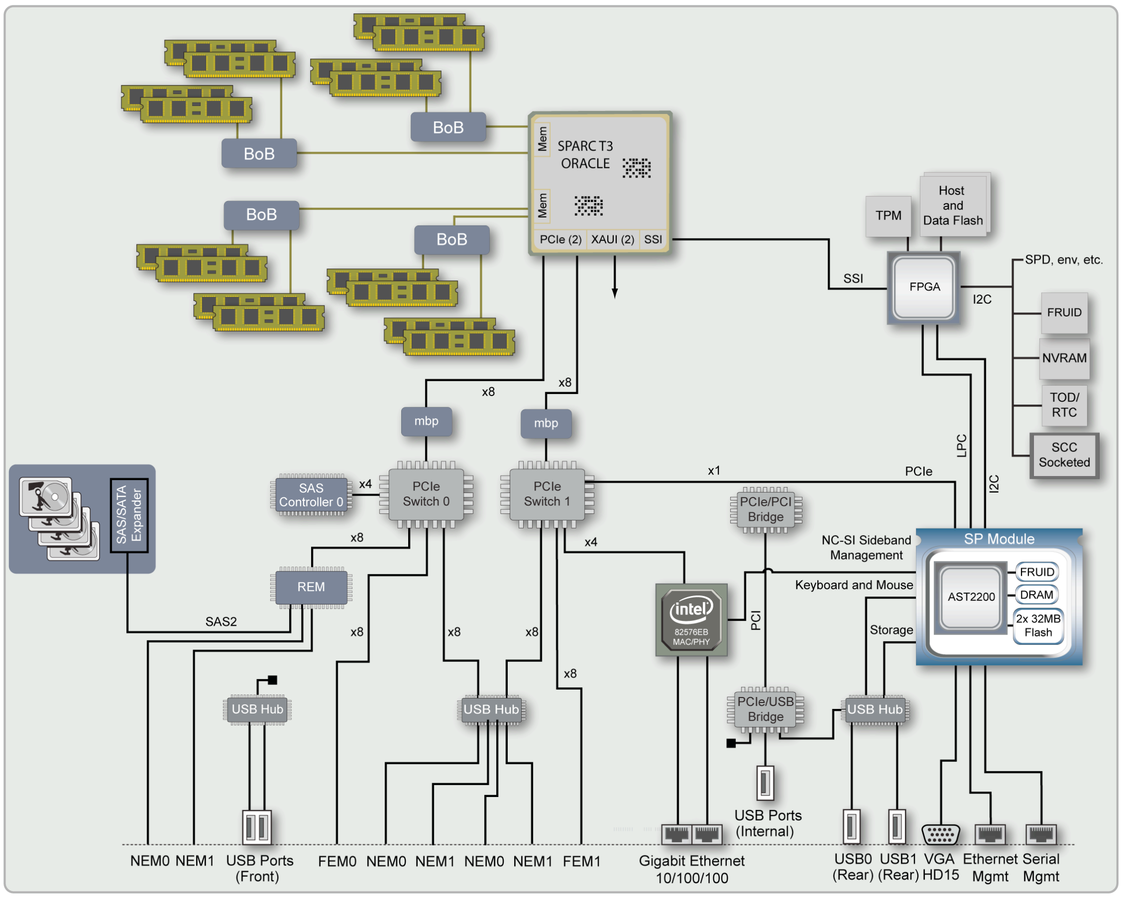

Figure 4. The SPARC T3 processor is a central component in the Netra SPARC T3-1 system motherboard design.

The motherboard interconnect has been greatly simplified over prior-generation servers. Power is distributed to the motherboard through a pair of metal bus bars connected to a power distribution board (PDB). A single flex-circuit connector routes all critical power control and DVD drive signaling to the PDB. One or two mini-SAS cables connect the motherboard to the disk drive backplane, providing data access to the internal SAS hard drives.

Memory Subsystem

In Netra SPARC T3-1 servers, the SPARC T3-1 processor's on-chip memory controllers uses buffer-on-board to communicate with DDR3-DIMM memory through high-speed serial links. Equipped with two memory controller units (MCU) supports a pair of FBDIMM2 links. Each FBDIMM2 link connects to a buffer-on-board (BoB) ASIC on the motherboard. Each BoB has two DDR3 channels, and each channel can contain up to two DDR3 DIMMs. A total of sixteen memory socket locations provide sufficient board space for two rows of DDR3-DIMMs per channel. DDR3-DIMM modules must be populated in like pairs—with 4-GB, or 16-GB modules—enabling a maximum capacity of 256 GB using 16-GB modules.

When populating DIMMS, the following sequence(s) must be followed to ensure a valid DIMM configuration is achieved. A minimum of 4 DIMMs (1 per BoB) must be installed. The first 4 DIMMS should be installed in the blue DIMM slots. These are channel 1, DIMM0 on each BoB. If additional DIMMs are to be added to the node a minimum of 4 additional DIMMs should be installed in the white DIMM slots. These are channel 0, DIMM 0 slots. If still more memory is required a minimum of 8 DIMMS should be added to black DIMM slots.

I/O Subsystem

There are two PCI Express switches on the motherboard, operating independently of each other. All ports on both switches can run at either 2.5 Gbps or 5.0 Gbps, and support link widths of 1, 2, 4, or 8 lanes. Both switches connect to both SPARC T3, and provide an interface to the internal and expansion I/O such as those listed below:

- **Disk controller.** Disk control is managed by a single LSI Logic SAS2008 controller chip that interfaces to a four-lane (x4) PCIe port.
- **Dual GBE.** Two x4 PCIe ports connect to two Intel Ophir dual GbE chips, providing four 10/100/1,000 Mb/sec Ethernet interfaces on the rear panel.
- **SB and DVD.** One x4 PCIe port connects to a PLX PEX8111 PCI bridge device. A second bridge chip converts the 32-bit 33-MHz PCI bus into multiple USB 2.0 ports. The system's USB interconnect is driven from those ports. In addition, a bridge chip interfaces one of the USB ports to IDE format, which is used to drive the optional DVD-RW drive.

To minimize cabling and increase reliability, a variety of smaller boards and riser cards are deployed. These infrastructure boards serve various functions:

- PDBs distribute system power from the dual power supplies to the motherboard and to the disk backplane (via a connector board).
- Connector boards eliminate the need for many discrete cables, providing a direct card plug-in interconnect to distribute control and most data signals to the disk backplane, the fan boards, and the PDB.

- Fan boards provide connections for power and control for both the primary and secondary fans in the front of the chassis. No cables are required because every dual fan module plugs directly into one of these PCBs, which in turn plugs into the connector board.
- PCIe Riser Cards plug directly into the motherboard, allowing PCIe cards to be installed. Two XAUI riser cards provide slots that access to the on-chip 10-GbE interfaces on the SPARC T3 processor, or alternatively provide access to PCIe interfaces. Each PCI/XAUI slot can accept either an Oracle proprietary optical/copper XAUI card or a low-profile PCIe card with up to an x8 form factor edge connector (x4 electrically). All PCI cards are installed in a horizontal orientation.
- A four-disk backplane mounts to the disk cages in the chassis, delivering disk data through a single, discrete mini-SAS cable from the motherboard.

Netra SPARC T3-1 Server Overview and Subsystems

The expandable Netra SPARC T3-1 server is optimized to deliver high performance and throughput. With considerable expansion capabilities and many high-reliability features, the Netra SPARC T3-1 server is an ideal platform for demanding mission-critical and consolidated workloads.

Enclosure

The Netra SPARC T3-1 server features a compact yet expandable 2U rackmount chassis that provides the flexibility to scale processing and I/O needs without wasting valuable datacenter space.

TABLE 1. DIMENSIONS AND WEIGHT

SERVER/DIMENSION	UNITED STATES	INTERNATIONAL
Height	3.44 inches (2U)	87.37 millimeters
Width	16.75 inches	425.45 millimeters
Depth	20 inches	502 millimeters
Weight (without PCI cards or rackmounts)	34.78 pounds	15.78 kilograms

Major Components

The Netra SPARC T3-1 server includes the following major components:

- A SPARC T3-1 processor running at 1.65 GHz with sixteen cores
- Up to 256 GB of memory in 16 DDR3-DIMM slots (populated in pairs of 8 or 16-GB DDR3-DIMMs)
- Four onboard 10/100/1,000-Mb/sec Ethernet ports
- Five PCIe slots: two full-height, half-length x8 PCIe slot 64bit at 133MHz, three low-profile PCIe slots, or one low-profile, and two low-profile PCIe or XAUI combination slots
- Two USB 2.0 ports on the rear panel

- Four internal SAS drives with DVD-RW is installed
- Onboard service processor with ILOM
- Two hot-swappable high-efficiency AC or DC 1100-W power supplies
- Five hot-swappable fan modules

Front and Rear Perspectives

Figures 5 and 6 illustrate the front and back panels of the Netra SPARC T3-1 server.

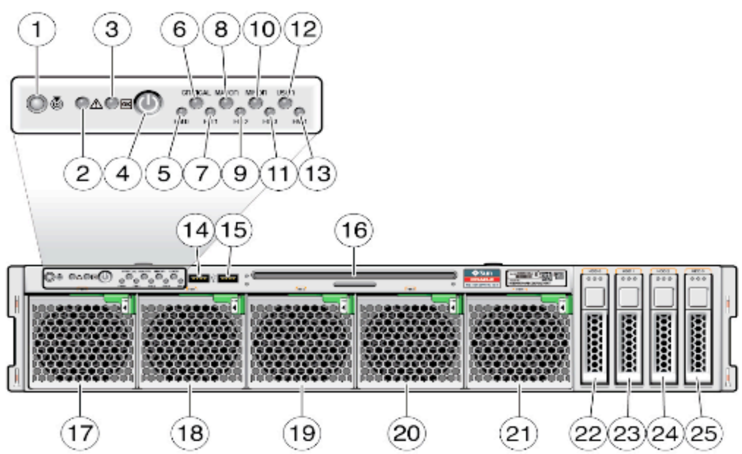

Figure Legend

1- System Locator
2- System fault indicator
3- System status indicator
4- Power button
5 - System Fan0 indicator
6- Critical Alarm indicator
7- System Fan1 indicator
8- Major Alarm indicator
9- System Fan2 indicators
10- Minor Alarm indicator
11- System Fan3 indicator
12- User Alarm indicator
13- System Fan4 indicator
14- Front USB0 port
15- Front USB1 port
16- DVD media push button
17- System Fan 0
18- System Fan 1
19- System Fan2
20- System Fan3
21- System Fan4
22- HDD0
23- HDD1
24- HDD2
25- HDD3

Figure 5. Front perspective of the Netra SPARC T3-1 server and its major components

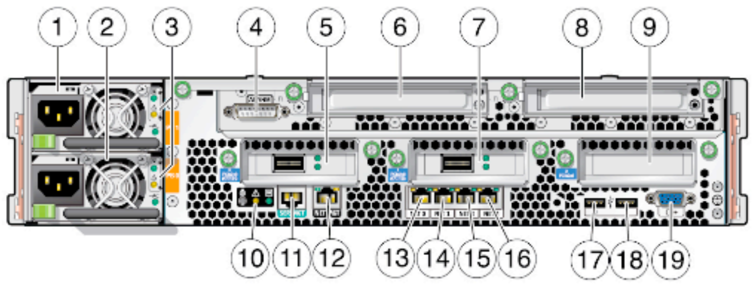

Figure Legend

1- Power Supply (PS1)
2- Power Supply (PS0)
3- PSU status indicators
4- Telco Alarm I/O Port
5- PCIe or XAUI slot0
6- PCIe slot3
7- PCIe or XAUI slot1
8- PCIe slot4
9- PCIe slot2
10- System status indicators
11- Serial Port
12- 10/100 Ethernet port
13-16 10/100/1000 Ethernet ports (0-3)
17-18 Rear USB port 0, 1
19- Video Port

Figure 6. Rear perspective of the Netra SPARC T3-1 server and its major components

External features of the Netra SPARC T3-1 server include:

- Front and rear system status indicator lights provide “Locator” (white), “Service Required” (amber), and “Activity Status” (green) for the system.
- A hinged access door opens to provide access to the front panel, where hot-swappable SAS disk drives and the optional DVD-RW device are accessible.
- Two USB ports are provided on the rear panel.
- Two hot-pluggable/hot-swappable AC or DC power supply units (PSUs) insert from the rear (rear indicator lights convey the status of each PSU).
- Four autosensing 10/100/1,000Base-T Ethernet ports are available on the rear panel.
- Two 10-GbE ports (that connect directly to the SPARC T3 processor) are available via optional XAUI networking cards.
- Five PCIe card slots are provided, two of which are combination PCIe/XAUI slots for optional XAUI networking cards.
- Two RJ-45 management ports (serial and network) are located on the rear panel for use with the ILOM system controller. The serial port is the default connection to the ILOM controller, and optionally the network port supports a 10/100Base-T connection to the controller.
- One telco alarm output port (DB-15) on the rear panel is connected to the service processor.

PCI Expansion

When the Netra SPARC T3-1 server is viewed from the rear, the PCI slots are in numeric order from bottom left-to-right numbered from zero to two, and top left-to-right numbered from three to five. The Netra SPARC T3-1 server has five PCIe card slots, two of which can support XAUI cards connected to the SPARC T3 processor 10-GbE interfaces (Slots 0 and 1). Slot numbering and type are summarized in Table 2.

TABLE 2. SLOT NUMBERING AND TYPE FOR NETRA SPARC T3-1 SERVER

SLOT NUMBER	TYPE	MECHANICAL	ELECTRICAL
0	PCIe or XAUI	X8	X8
1	PCIe or XAUI	X8	X8
2	PCIe Mezzanine	X16 (FH/HL)	X8
3	PCIe	133 MHz at 64-bit,	
4	PCIe Mezzanine	X16 (FH/HL)	X8

System Network Interfaces

The Netra SPARC T3-1 server features four 10/100/1,000-Mb/sec Ethernet ports (RJ-45) located on the rear panel. Multiple onboard GbE ports promote flexibility as well as network configurations that support interface failover. The ports are numbered in sequence from left to right. Each port autonegotiates its link connection, and LEDs above the port indicate the speed of the established link (green signifies that the link is established at 1,000 Mb/sec). All four ports support 10/100/1,000 Mb/sec; full-and half-duplex operation; IEEE 802.3ab autonegotiation for speed, duplex, and flow control; and PXE boot for network booting.

Internal Storage

The Netra SPARC T3-1 server supports four internal SAS drives, and one internal DVD-RW device. The SAS2 hard drives are hot-swappable, 300-GB-capacity, 10,000-RPM, 3-Gb/sec, 2.5-inch hard drives. The drives are also 100 percent duty-cycle, small-form-factor drives that are NEBS certified.

The onboard LSI SAS2008 disk controller enables hardware RAID, providing data redundancy and increased performance without additional cost. The controller provides either two-disk RAID 1 volumes (integrated mirroring) or two-, three-, or four-disk RAID 0 volumes (integrated striping).

The DVD-RW device available with the four-drive configuration provides the ability to read and write to a removable DVD media access device, enabling users to store data without external storage or hard drive requirements. In addition, the virtual storage feature of the ILOM service processor allows host access to remote CD-ROM ISO images, just as if the CD device were a locally attached USB CD-ROM device, eliminating the cost, complexity, and need for a separate optical disk drive for each individual server.

Power and Thermal

Typical heat dissipation and power consumption metrics for a Netra SPARC T3-1 server equipped with a 1.65-GHz SPARC T3 processor and 16 GB RAM are as follows:

- AC power: 100 to 240 VAC, 50 to 60 Hz
  - Maximum input current: 6.0 A at 100 to 120 VAC, 3.0 A at 200 to 240 AC
  - Maximum operating input power: 600 watts
  - Maximum heat dissipation: 1,365 BTU/hour
- DC power: 48 VDC or 68 VDC (nominal), 40 VDC to 75 VDC (range)
  - Maximum input current: 12.5 A at 48 VDC, 10 A at 60 VDC

Power Supplies

Engineered for high availability as well as low energy consumption, the Netra SPARC T3-1 server is equipped with two highly efficient, redundant, hot-swappable AC or DC power supplies with separate power cords. The power supplies are rated at 1100 watts each and can auto-detect between 120/240V and 50/60Hz.

One power supply is sufficient to run a fully populated server. For maximum protection against power supply failures, however, dual power supplies should be installed at all times. In normal operation, the power supplies share the power demands of the system equally.

To further reduce power requirements and to meet NEBS certification, the Netra SPARC T3-1 server can be run on DC power. Using DC power reduces overall operating costs by lowering energy use, reducing heat, and increasing reliability.

Cooling

Netra SPARC T3-1 servers feature an innovative chassis design engineered to lessen the need for cooling. The power and cooling efficiency of these systems exceeds that of many competitive carrier-grade systems configured with similar processing, memory, and storage capacities. The effective front-to-back airflow design helps to lower component temperatures, reducing the number of fans needed to cool the system. The chassis design features hex-shaped, honeycombed air inlet holes that enhance airflow and provide EMI shielding.

The chassis is divided into two distinct airflow chambers as shown in Figure 7. One chamber draws airflow for the motherboard (CPUs, RAM) and PCI trays. The other chamber provides airflow for the hard drives, DVD-RW device (if present), and power supplies. The power supply/storage chamber is cooled by individual fans on the back of each power supply. The motherboard/PCI chamber is cooled by a row of fans mounted in front of the server behind the bezel.

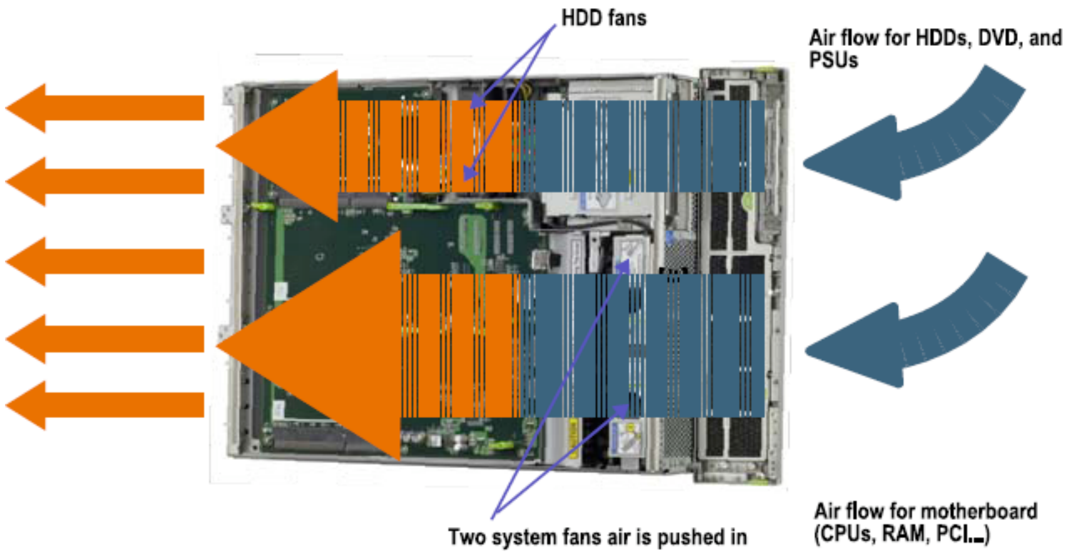

Figure 7. The chassis is divided into two distinct airflow chambers that are both cooled front-to-back.

Variable-speed fans run under the control of the onboard service processor, which monitors processor temperatures and system ambient air temperature. Based on these readings, the fans operate at the lowest speeds possible to provide sufficient cooling—conserving power usage, prolonging fan life, and reducing acoustical noise.

Fan failures are indicated by an amber “Service Required” LED on the front panel. Fan modules are designed for redundancy—a backup fan enables system continuity in the event of a fan failure. The fans are also hot-swappable so that a module with a failed fan can be removed and a new fan module inserted without shutting down the system.

Rackmounting

The Netra SPARC T3-1 server ships with a hard rackmount 19-inch 4-post kit. Optional kits are available for 19-inch two-post, 23-inch two-post, and 600 x 600 millimeter hard rackmount, as well as 19-inch 4-post slide rackmount. For example, the following options are available from Oracle to facilitate rack mounting:

- **Rack-Mounting Slide Rail Kit.** This is a four-point mounted slide rail kit (mounting points are located at the rack front and rear).
- **Tool-less Rack Kit.** As the name implies, this rackmounting kit snaps into most third-party racks without requiring any tools.
- **Cable Management Arm.** The Cable Management Arm supports and protects cables as the server slides into and out of the rack.

RAS Features

Customer data and applications comprise critical business assets. Enterprise computing technologies strive to furnish a high degree of data protection (reliability), to provide virtually continuous application access (availability), and to incorporate procedures and components that help to resolve problems with minimal business impact (serviceability). Commonly referred to as RAS, these capabilities are a standard part of Oracle's mission-critical computing solutions.

The Netra SPARC T3-1 servers are engineered for hardware failure prevention, near-continuous operation, fast recovery, and easy serviceability. RAS features for these systems include

- **Reduced component count.** Integration of key functionality into the SPARC T3 processor means fewer components and reduced incidence of component failures to increase overall availability.
- **Hot-swappable redundant components.** Mirrored disks, redundant fan modules, and redundant power supply units can be quickly and easily changed out, increasing system uptime.
- **Accessible components for improved serviceability.** Front-accessible, hot-swappable disk drives can be replaced quickly. The optional DVD-RW drive can also be removed without opening the top cover of the chassis. Fan modules and power supply units can be replaced without completely removing a system from the rack.
- **Indicator LEDs on the back of the chassis.** Easily visible LEDs allow problems to be identified and isolated easily. Diagnostic LEDs are also included on the motherboard.
- **Integrated out-of-band management.** Standard on the Netra T5220 servers at no additional charge, ILOM provides powerful tools for remote system management, simplifying administrative tasks, reducing on-site personnel needed, and lowering the cost of operations.
- **NEBS Level 3 certification.** Telecordia NEBS Level 3 certification demonstrates that the system meets stringent requirements for extremely reliable operation, even in harsh environmental conditions.

Enterprise Scalability and Availability with Oracle Solaris

The SPARC T3 processor technology is mature and well tested, and Netra SPARC T3-1 servers share binary compatibility with earlier SPARC systems—they come preloaded with the solid and secure foundation of Oracle Solaris 10. Moreover, Oracle Solaris 10 includes a variety of sophisticated tools that let organizations easily consolidate and manage workloads while taking advantage of processor technology innovation and availability features that are part of Oracle Solaris.

The subsections that follow provide additional information about some of the most relevant characteristics of Oracle Solaris for telecommunications providers. Additional information about Oracle Solaris capabilities for carrier-grade applications can be found in the white paper, "Oracle Solaris: The Carrier-Grade Operating System," which is available at http://www.oracle.com/technetwork/server-storage/solaris11/documentation/solaris-thecarriergradeos-wp-308726.pdf.

Scalability and Support for CoolThreads Technology

Oracle Solaris 10 is specifically designed to deliver advanced capabilities of SPARC T3 processor–based systems. In fact, Oracle Solaris 10 provides key functionality for virtualization, optimal utilization, high availability, unparalleled security, and extreme performance for both vertically and horizontally scaled environments.

One significant advantage of a system based on the SPARC T3 processor is that it appears as a familiar symmetric multiprocessing (SMP) system to Oracle Solaris and applications. In addition, Oracle Solaris 10 incorporates features that help to improve application performance on CMT architectures:

- **CMT awareness.** Oracle Solaris 10 is aware of the SPARC T3 processor hierarchy so that the scheduler can effectively balance the load across all the available pipelines. Even though it exposes the SPARC T3 processor as 128 logical processors, Oracle Solaris understands the correlation between cores and threads they support.
- **Fine-grained manageability.** Oracle Solaris 10 has the ability to enable or disable individual processors. In the case of the SPARC T3 processor, this ability extends to enabling or disabling individual cores and threads (logical processors). In addition, standard Oracle Solaris features such as processor sets provide the ability to define a group of logical processors and schedule processes or threads on them.
- **Support for virtualized networking and I/O, and accelerated cryptography.** Oracle Solaris contains technology to support and virtualize components and subsystems on the SPARC T3 processor, including support for the on-chip 10-GbE ports and the on-chip PCIe interface. Accelerated cryptography is supported through the Oracle Solaris cryptographic framework.

Fault Management and Predictive Self-Healing

Oracle Solaris 10 introduced a new architecture for building and deploying systems and services capable of fault management and predictive self-healing. Oracle Solaris Predictive Self Healing is an innovative capability in Oracle Solaris 10 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.

Oracle Solaris Predictive Self-Healing and Oracle Solaris fault management provide the following specific capabilities on Netra SPARC T3-1 servers:

- CPU Offlining takes a core offline that has been deemed faulty. Offlined CPUs are stored in the resource cache and stay offline on reboot unless the processor has been replaced, in which case the CPU is cleared from the resource cache.
- Memory Page Retirement retires pages of memory that have been marked as faulty. Pages are stored in the resource cache and stay retired on reboot unless the offending FB-DIMM has been replaced, in which case affected pages are cleared from the resource cache.

- I/O Retirement logs errors and faults.
- fmlog logs faults detected by the system.

High Availability with Oracle Solaris Cluster

Mission-critical network infrastructure applications in telecommunications often require clustered physical servers to avoid single points of failure and the potential for service interruptions or data loss. Oracle Solaris Cluster is designed to provide the highest availability for applications running on Oracle Solaris. It also allows telecommunication providers to combine high availability clustering with virtualization to offer increased service levels at lower cost.

Oracle Solaris Cluster enables:

- Sub-second failure detection and lower service recovery time through Solaris kernel-level integration
- Optimized resource use along with high availability and fault isolation by supporting both Oracle VM for SPARC and Oracle Solaris Containers
- Consolidation of multiple applications within the same cluster of physical servers using virtual nodes or virtual clusters
- Disaster Recovery through multi-site, multi-cluster configurations

For more information on Oracle Solaris Cluster refer to: http://www.oracle.com/technetwork/server-storage/solaris-cluster/.

Important New Features in Oracle Solaris 11 Express 2010.11

Oracle Solaris 11 Express provides access to advanced Oracle Solaris features that have been in development for more than five years. It allows customers to deploy the latest Oracle Solaris features today while preparing their environments to take advantage of Oracle Solaris 11.

Two new Oracle Solaris 11 Express features are noteworthy for telecommunications providers:

- **Network-based package management.** The Oracle Solaris 11 Express Image Packaging System (IPS) is a new network-based package management system designed to greatly decrease planned system downtime and provide for completely safe system updates and upgrades. IPS offers a framework for complete software lifecycle management such as installation, upgrade, and removal of software packages. Administrators can install software from network-based package repositories with full automatic dependency checking for any additional libraries that might be required during a software package install.
- **Network virtualization**. A new network stack architecture, also known as the Crossbow project, has been introduced in Oracle Solaris 11 Express. The architecture offers improvements in performance, observability, security, and ease of use. Amongst other features, it enables administrators to create Virtual NICs (VNICs) and virtual switches, resulting in a virtual network infrastructure that is completely independent from the underlying hardware. This gives IT organizations the ability to deploy a "datacenter in a box" solution that brings increased levels of flexibility and TCO savings, much like a larger datacenter consolidation project. Network

virtualization also gives telecommunications carriers a method to control network bandwidth and network traffic flow for different services deployed on the system. Thus services that are very sensitive to network latency can get their share of network bandwidth before others, and the physical network can be spared from being flooded with unwanted traffic levels.

End-to-End Virtualization Technology

Virtualization technology is increasingly popular as organizations strive to consolidate disparate workloads onto fewer but more-powerful systems to improve utilization. Netra SPARC T3-1 servers are specifically designed for virtualization, providing very fine-grained division of multiple resources—from processing to virtualized networking and I/O. Most importantly, Oracle's virtualization technology is provided as a part of the system.

A Multithreaded Hypervisor

Like the UltraSPARC T2 processor, the SPARC T3 processor provides a multithreaded hypervisor—a small firmware layer that provides a stable virtual machine architecture that is tightly integrated with the processor. Multithreading is crucial because the hypervisor interacts directly with the underlying chip-multithreaded SPARC T3 processor. This architecture is able to context-switch between multiple threads in a single core, a task that can require additional software and considerable overhead in competing architectures that do not interface with the hardware directly.

The strength of Oracle's approach is that all of the layers of the architecture are fully multithreaded, from the processor up through applications that use the fully threaded Java application model.

Oracle VM Server for SPARC

Supported in all Oracle servers using CMT technology, Oracle VM Server for SPARC provides full virtual machines that can run an independent OS instance—each containing virtualized CPU, memory, storage, console, and cryptographic devices. Within the Oracle VM Server for SPARC architecture, each instance of Oracle Solaris 10 is written to the hypervisor, which provides a stable, idealized, and virtualizable representation of the underlying server hardware to the OS in each virtual machine.

By taking advantage of virtual machines, organizations gain the flexibility to deploy multiple OS instances simultaneously on a single platform. In addition, administrators can leverage virtual device capabilities to transport an entire software stack hosted on a virtual machine from one physical machine to another. Virtual machines can also host Oracle Solaris Containers to capture the isolation, flexibility, and manageability features of both technologies.

To support virtualized networking, Oracle VM Server for SPARC implements a virtual Layer 2 switch, to which guest virtual machines (guest OSs) can be connected. Each guest virtual machine can be connected to multiple vSwitches, and multiple guest virtual machines can also be connected to the same vSwitch. vSwitches can either be associated with a real physical network port, or they may exist without an associated port, in which case the vSwitch provides only communications between virtual machines in the same server, thus saving valuable network resources.

Oracle Solaris Containers

Providing virtualization at the OS level, Oracle Solaris Containers consist of a group of technologies that work together to efficiently manage system resources, virtualize the environment, and provide a complete, isolated, and secure runtime environment for applications. Oracle Solaris Containers can be used to isolate applications and processes from the rest of the system. This isolation helps enhance security and reliability because processes in one zone are prevented from interfering with processes running in another zone.

Resource management tools provided with Oracle Solaris help allocate resources such as CPUs to specific applications. CPUs in a multiprocessor system (or threads in the SPARC T3 processor) can be logically partitioned into processor sets and bound to a resource pool, which in turn can be assigned to an Oracle Solaris Container. Resource pools provide the capability to separate workloads so that consumption of CPU resources does not overlap. They also provide a persistent configuration mechanism for processor sets and scheduling class assignment. In addition, the dynamic features of resource pools enable administrators to adjust system resources in response to changing workload demands.

System Management Technology

As the number of systems grows in any organization, managing the infrastructure throughout its lifecycle becomes increasingly complex. Effective system management requires both integrated hardware that can sense and modify the behavior of key system elements, as well as advanced tools that can automate key administrative tasks.

Netra SPARC T3-1 servers can be managed through an integrated lights out management service as well as through Oracle Enterprise Manager Ops Center, and enterprise-level management solution.

Oracle Integrated Lights Out Manager

The Oracle Integrated Lights Out Manager service processor, which is provided across many of Oracle's x64 servers, acts as a system controller, simplifying the remote management and administration of Netra SPARC T3-1 servers. The service processor is similar in implementation to that used in other Oracle modular and rackmount x64 servers. As a result, these servers integrate easily with existing management infrastructures, including on-site enterprise management frameworks or element managers.

The service processor offers the following capabilities:

- Provides IPMI 2.0–compliant management functions to the server module's firmware, OS, and applications, and to IPMI-based management tools that access the service processor via the ILOM Ethernet management interface. It provides visibility of environmental sensors, both on the server module and elsewhere in the chassis.
- Manages inventory and environmental controls for the server module, including CPUs, FB-DIMMs, and power supplies, and supplies HTTPS, CLI, or SNMP access to this data.
- Supplies remote textual console interfaces.

- Provides a means to download upgrades to all system firmware.

The Oracle Integrated Lights Out Manager service processor allows the administrator to remotely manage the server independently of the OS running on the platform and without interfering with system processing. Oracle Integrated Lights Out Manager can send e-mail alerts of hardware failures and warnings, as well as other system-related events. The Oracle Integrated Lights Out Manager circuitry runs independently from the server, using server standby power. As a result, Oracle Integrated Lights Out Manager firmware and software continue to function when the server OS goes offline and even when the server is powered off.

Oracle Integrated Lights Out Manager monitors the following Netra SPARC T3-1 server conditions:

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

Oracle Enterprise Manager Ops Center

Oracle Enterprise Manager Ops Center is infrastructure lifecycle management software for deploying, monitoring, patching, and managing large and small installations of Oracle systems. Oracle Enterprise Manager Ops Center takes a step-by-step approach to unraveling the challenges of getting systems operational quickly:

- **Discover.** As systems are added to the management network, administrators can use Oracle Enterprise Manager Ops Center to discover bare metal systems based on a given subnet address or IP range.
- **Group.** Given the number of systems to manage and the constant repurposing of systems, it is critical for IT organizations to find ways to group resources together. Oracle Enterprise Manager Ops Center enables users to logically group systems together and perform actions across a group of systems as easily as performing actions on a single system. Systems can be grouped by function (Web servers versus grid computing), administrative responsibility, or other categorization based on organizational needs.
- **Provision.** Oracle Enterprise Manager Ops Center remotely installs OSs (Oracle Solaris, RedHat, or SuSE Linux) onto selected systems. Administrators can use this functionality to provision OSs onto bare metal systems or reprovision existing systems. As the infrastructure lifecycle continues, Oracle Enterprise Manager Ops Center can update firmware and provision software packages and patches to selected systems.

- **Monitor.** When systems are up and running, administrators can use Oracle Enterprise Manager Ops Center to monitor system health, helping to ensure that everything is running at the optimal levels. The software provides detailed hardware monitoring for attributes such as fans, temperature, disk, and voltage usage, including bare metal systems. Oracle Enterprise Manager Ops Center also monitors OS attributes such as swap space, CPU, memory, and file systems. Administrators can define specific threshold levels and set preferred notification methods, including e-mail, pager, or Simple Network Management Protocol (SNMP) traps for each monitored component, as business needs demand.
- **Manage.** Businesses require that infrastructure lifecycle management extend beyond just deploying and monitoring systems. Oracle Enterprise Manager Ops Center includes lights-out management capabilities, such as powering systems on and off, and remote serial console access to help IT organizations manage their IT infrastructures from remote locations. Leveraging Oracle Enterprise Manager Ops Center software's Role-Based Access Control (RBAC) feature, organizations can grant permissions to specific users to perform specific management tasks.
- **Hybrid user interface.** Oracle Enterprise Manager Ops Center offers users a hybrid user interface (UI) accessible from the Web, which integrates both the GUI and CLI into one console. With this hybrid UI, operations performed in the GUI are simultaneously reflected in the CLI, and vice versa.

Chip Multithreading Tools Enable Fast Performance and Rapid Time-to-market

No matter how compelling new hardware or OS platforms may be, organizations must be ensured that the costs and risks of adoption are in line with the rewards. In particular, organizations want to be able to continue to leverage the considerable advantages of popular commercial and open source software. Developers don't want to have to switch compilers and basic development tools. Administrators can scarcely afford a more-complex support matrix or more time spent getting applications to run effectively in a new environment. Previous generations of CMT servers have relied upon individual tools that provide recommendations as to whether an existing application would execute with maximum efficiency on a CMT server. Developers could employ a variety of individual tools to optimize applications for execution within a CMT environment. However, Oracle, in its ongoing efforts to produce a tightly integrated software platform, has integrated much of what were separate tools into Oracle Solaris or Oracle Solaris Studio.

Oracle Solaris Studio delivers a comprehensive development platform for building scalable, secure and reliable applications for CMT servers. For more information about Oracle Solaris Studio and Oracle developer tools visit the following links:

- http://www.oracle.com/technetwork/server-storage/solarisstudio/overview/index.html
- http://www.oracle.com/us/products/tools/

Accelerating Development and Reducing Costs with Oracle Netra Data Plane Software Suite

Oracle’s Sun Netra Data Plane Software Suite can help telecommunications carriers accelerate development and reduce R&D and sustaining costs by exploiting the CMT architecture of the SPARC T3 processor for line rate packet processing performance. Comprised of an open unified development environment and a light-weight runtime environment, Netra Data Plane Software Suite enables better economics through rapid development. It can also leverage virtualization capabilities in the hypervisor using Oracle VM Server for SPARC (previously called Sun Logical Domains, or LDoms) for radical consolidation of Control and Data Plane functions into a single unified architecture.

Conclusion

Oracle's Netra SPARC T3-1 servers combine breakthrough performance and energy efficiency with unprecedented processor core, memory, and I/O densities to help telecommunications companies scale their IT infrastructures in a cost-effective manner. The SPARC T3-1 processor used in Netra SPARC T3-1 servers delivers approximately twice the throughput and efficiency of the previous generation UltraSPARC T2 processor. With 128 threads per processor, on-chip memory management, two 10 GbE interfaces, PCIe, and on-chip cryptographic acceleration, the SPARC T3 processor fundamentally redefines the capabilities of a modern processor. By incorporating cache coherency for multiprocessor support, SPARC T3 processors allow these capabilities to be multiplied incrementally.

Netra SPARC T3-1 servers leverage these strengths to provide powerful and highly scalable server platforms while delivering even higher levels of performance in a compact rack-mount chassis. The result is a highly compact IT infrastructure that can truly scale to meet today’s telecommunications challenges and budget constraints.

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

Oracle's Netra SPARC T3-1 Server Architecture
February 2011, Version 1.1

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