An Oracle White Paper
April 2012

Taking Advantage of Multicore/Multiprocessor Technologies to Meet Today's Storage Needs

Sun ZFS Storage Appliances Based on Intel® Xeon® Processors

Introduction

Data growth is not only driving up storage costs, but also transforming performance and efficiency requirements for storage systems. IDC estimated that in 2011, the amount of information created and replicated would surpass 1.8 zettabytes (1.8 trillion gigabytes) — growing by a factor of nine in just five years.[1] This digital data is expected to grow 29X over the current decade, reaching 35,000 exabytes by 2020.[2]

IT managers are looking for new storage solutions that can address today's needs for managing large amounts of data. They want to do the following:

- Reduce the cost of storage administration by improving visibility and simplifying common administration tasks
- Deliver higher I/O throughput without the cost and complexity of a SAN solution
- Protect data integrity and prevent data loss or theft
- Reduce data center power and cooling costs

Oracle and Intel, as part of a broad strategic alliance, have been working together to ensure that Sun ZFS Storage Appliances are optimized to meet these challenges by unleashing the power and capabilities of current and future Intel® Xeon® processors. A key result of this collaboration has been the optimized use of the Oracle Solaris operating system as the

[1] "IDC Digital Universe Study, sponsored by EMC," June 2011. Available at http://idcdocserv.com/1142.
[2] IDC Worldwide Enterprise Storage Systems 2011-2015 Forecast, May 2011.

foundation upon which the Sun ZFS Storage Appliance has been built. Oracle Solaris has long been optimized for Intel Xeon processors, and Sun ZFS Storage Appliances leverage these significant optimizations as well as many other innovations in the world's leading enterprise operating system.

Oracle's Sun ZFS Storage Appliances are unique in the industry in that they provide much more processing power than those of other major storage vendors. With up to 40 cores per controller, the Sun ZFS Storage 7420 system has 10 times as many cores as the NetApp FAS3270 filer, enabling it to perform complex processing for compression, deduplication, real-time storage analytics, and other advanced data services without slowing down I/O throughput.

While offering more power and larger memory configurations than competing solutions, Sun ZFS Storage Appliances do not cost more to own. They simply bring more value. By combining the multithreaded Oracle Solaris operating system with the powerful multicore Intel Xeon processor family, Sun ZFS Storage Appliances can deliver scalable performance, innovative data services, advanced reliability, and high energy efficiency.

Scalable Performance

Today's enterprise IT, Web, and cloud infrastructures often support thousands of users and involve massive amounts of data. For these large-scale implementations, storage I/O throughput has become a critical factor in the ability to deliver high service levels for applications. Today's backup environments also require faster performance so that larger amounts of data can be protected within increasingly short backup and restore windows.

The scalable performance of Sun ZFS Storage Appliances enables organizations to meet application service levels while dramatically reducing storage costs. Higher performance at lower cost is achieved by using Oracle Solaris, a multithreaded operating system, to take advantage of up to four Intel Xeon processors, each of which can have up to 10 cores and 20 threads. This powerful platform enables the appliances to support large numbers of concurrent threads and deliver record-setting IOPS performance at a fraction of the cost of traditional storage solutions and SAN environments.

The Sun ZFS Storage Appliance leverages more than 20 years of development in Oracle Solaris to continuously improve its scalability and performance to match the consistent increases in processor count and memory capacity in Oracle servers over the years. This performance optimization, which has made Oracle Solaris run efficiently on servers with Intel Xeon processors, now helps ensure that the foundation for Sun ZFS Storage Appliances is bullet-proof and highly scalable.

Performance Features of the Intel Xeon Processor and Oracle Solaris

Sun ZFS Storage Appliances feature Intel Xeon processor family CPUs with four, six, eight, or 10 cores per socket. The Intel Xeon processor family is an expandable processor family for systems from two to eight (or more) sockets. The Intel Xeon processor family offers up to 10 cores and 20 threads per processor — up to 160 threads in an 8-socket system. These processors are massive data pump engines, moving more data per square millimeter of storage footprint than other leading storage silicon.

The Intel Xeon processor family works together with Oracle Solaris to deliver the performance features described below and enable the Sun ZFS Storage Appliance to achieve record-setting performance.

Intel Hyper-Threading Technology

Intel Hyper-Threading Technology provides two hardware threads per core, yielding 20 hardware threads in a 10-core processor. Oracle Solaris takes advantage of Intel Hyper-Threading Technology so that threads can be executed in parallel within each processor core. This helps ensure optimal scheduling of software threads to minimize resource contention and maximize performance. In addition, years of performance optimization work have gone into the Oracle Solaris threading model so that the Oracle Solaris multithreaded kernel and threaded user applications can both take full advantage of Intel multicore processors. Since Sun ZFS Storage Appliances include many powerful data services that can be executed in parallel with other activities, the ability to effectively use the

available threads on the Intel Xeon processor enables Sun ZFS Storage Appliances to increase overall performance throughput.

Intel QuickPath Technology (Intel QPI)

Intel QuickPath Technology is a platform architecture that provides high-speed (up to 25.6 GB/s), point-to-point connections between processors as well as between processors and the I/O hub. Each processor has its own dedicated memory that it accesses directly through an Integrated Memory Controller. If a processor needs to access the dedicated memory of another processor, it can do so through a high-speed Intel QuickPath Interconnect that links all the processors. Intel QuickPath Interconnect architecture also includes capabilities such as an optimized scheduler and memory placement optimization capability that work together with Oracle Solaris to deliver proven performance benefits. The result is faster movement of data between processors and memory so that the Sun ZFS Storage Appliance can move data to and from DRAM or SSD cache quickly. This, in turn, improves read and write performance for I/O requests.

Oracle Solaris Memory Placement Optimization (MPO)

Multiprocessor systems generally demonstrate some memory locality effects, which means that when a processor requests access to data in memory, that operation will occur with somewhat lower latency if the memory bank is physically close to the requesting processor. Oracle Solaris determines the NUMA configuration of the underlying hardware platform at boot time and then uses this information to allocate memory and schedule software threads such that memory is as close as possible to the processors that access it.

Intel Turbo Boost Technology

Intel Turbo Boost Technology, together with Intel Intelligent Power Technology, delivers performance on demand, letting processors operate above the rated frequency to speed specific workloads and reduce power consumption during low utilization periods. In those situations where Oracle Solaris determines that maximum processing power is required, the Intel Xeon processor increases the frequency in the active core(s) to the maximum extent possible given other conditions such as load, power consumption, and temperature.

Hybrid Storage Pools

Large memory and cache sizes can be used in Sun ZFS Storage Appliances due to the powerful capabilities of Oracle Solaris and Intel Xeon processors as well as a technology called Hybrid Storage Pools. Hybrid Storage Pools enable Sun ZFS Storage Appliances to automatically place data in memory, flash, or hard disk drives to optimize performance and cost. They continuously monitor data usage patterns to determine whether and how to use the different storage media on the appliance. For example, large synchronous writes, such as video streaming, do not benefit from caching, so there is no attempt to place this type of data to write cache. Similarly, the read cache is populated based on an intelligent algorithm that takes into account not only the most recently used data but also anticipated

read requests and estimated data to be held in DRAM. Hybrid Storage Pools enable many data requests to be serviced by fast memory and cache rather than disk, thus speeding reads and writes.

Outstanding Performance Benchmarks

A recently published SPECsfs2008 benchmark offers a good example of the industry-leading performance the Sun ZFS Storage Appliance is able to achieve by taking advantage of the extreme multithreaded capabilities of Oracle Solaris and Intel Xeon processors. SPECsfs2008 is the latest version of the Standard Performance Evaluation Corporation benchmark suite measuring file server throughput and response time.

Oracle announced in February 2012 that the Sun ZFS Storage 7320 appliance delivered 134,140 SPECsfs2008_nfs ops/sec with an overall response time (ORT) of 1.51 milliseconds on the SPECsfs2008_nfs benchmark for network file system throughput. As shown in Figure 1, this result beat the NetApp FAS3270 filer by 33% with a configuration that is less than one-fifth the list price.[3]

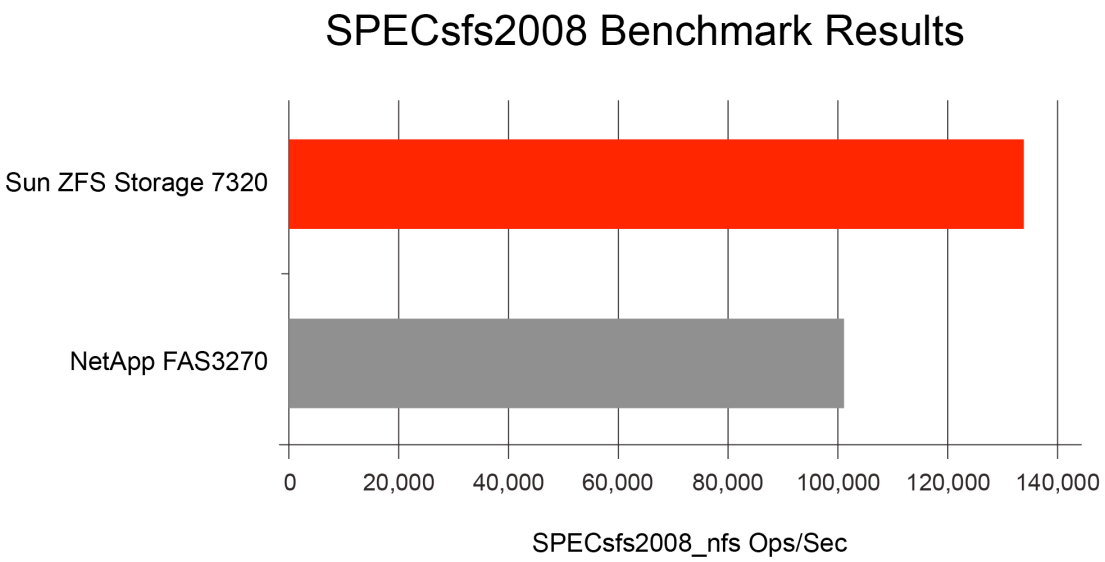

Figure 1. The Sun ZFS Storage 7320 beat the NetApp FAS3270 by 33% in the SPECsfs2008 benchmark.

Oracle's list of storage benchmarks showing proven performance continues to grow. For more information visit *http://www.oracle.com/us/solutions/performance-scalability/sun-storage-gateway-160373.html*.

[3] Results as of February 22, 2012. For more information see *www.spec.org*. Sun ZFS Storage 7320 Appliance: 134,140 SPECsfs2008_nfs ops/sec, 1.51 msec ORT. NetApp FAS3270: 101,183 SPECsfs2008_nfs ops/sec, 1.66 msec ORT.

Innovative Data Services Made Possible by a Scalable Platform

With scalability up to 40 cores per controller and 80 cores in a high-availability cluster, Sun ZFS Storage Appliances have ample processing power to perform complex calculations for data services such as compression, deduplication, and real-time storage analytics — without slowing down I/O throughput. The powerful Intel Xeon processor family, coupled with the multithreaded Oracle Solaris operating system, enables Sun ZFS Storage Appliances to deliver innovative services that would not be practical with a less scalable platform. In fact, some vendors require customers to purchase specialized add-on systems to perform functions such as deduplication or compression without overloading the primary storage controller.

The sections that follow describe how key data services in Sun ZFS Storage Appliances take advantage of mulitprocessing power to deliver more value.

ZFS File Compression

The Sun ZFS Storage Appliance reduces the amount of space required to store user data through several methods, including file-level compression built into the Oracle Solaris ZFS file system. This in turn increases the effective storage capacity available to applications. Having compression built into the file system (as opposed to deploying an appliance between the client application and storage) not only simplifies the management of complex storage architectures but also helps minimize the impact on application performance. In addition, there are some cases in which compression can improve system performance due to the fact that compression results in fewer bytes of data and therefore, less I/O traffic to and from disks.

Compression typically provides good results for unstructured data in a file sharing environment, often reducing the amount of physical storage needed to house data by 50% or more. Sun ZFS Storage Appliances provide four levels of compression so administrators can choose the compression algorithm best suited for their particular application environment. Sun ZFS Storage Appliance compression algorithms are implemented in software so they leverage the power of Intel multicore processors to achieve fast results.

Efficient Inline Deduplication

Deduplication identifies and stores only unique blocks of data and can potentially save significant amounts of storage capacity by eliminating redundant data from within the same configured pool.

Many storage systems use a method of deduplication called post-processing in which the complete set of data is initially written to disk and then a subsequent pass is made, usually during off hours, to process the data and remove duplicate copies. This approach avoids putting a heavy load on the storage system or application during the initial writes. However, it incurs a penalty in the sense that it requires enough extra storage capacity to store the duplicate data until the post-processing step is complete.

The approach used by Sun ZFS Storage Appliances avoids this over-subscription of storage space because deduplication is performed on the fly while the data is being written. This approach requires more processing power in the storage appliance, which plays to the strengths of Intel Xeon processors with multiple cores and built-in support for cryptographic hash functions.

Sun ZFS Storage Appliances perform deduplication at the block level rather than at the file level. Every new block of data being stored is checksummed using the 256-bit secure hash function to uniquely identify the data block. With this 256-bit checksum, the odds of finding a duplicate hash entry where the data inside the blocks is not also identical is only one in 2^{256}. In other words it is 50 orders of magnitude less likely than an undetected, uncorrected memory error in error-correcting code (ECC) memory on the most reliable hardware platforms.

Because the 256-bit hash entry effectively serves as a unique identifier, the Sun ZFS Storage Appliance deduplication process does not have to do a byte-by-byte comparison of data blocks.[4] This fact, coupled with the ability to leverage Intel Xeon processors for parallel processing and hardware accelerated hash lookup, enables Sun ZFS Storage Appliances to keep up with I/O throughput rates and perform the deduplication inline.

As illustrated in Figure 2, inline deduplication saves storage capacity during the temporary period when multiple copies of data would be stored in a post-processing approach. It also has the advantage of avoiding an extra processing step.

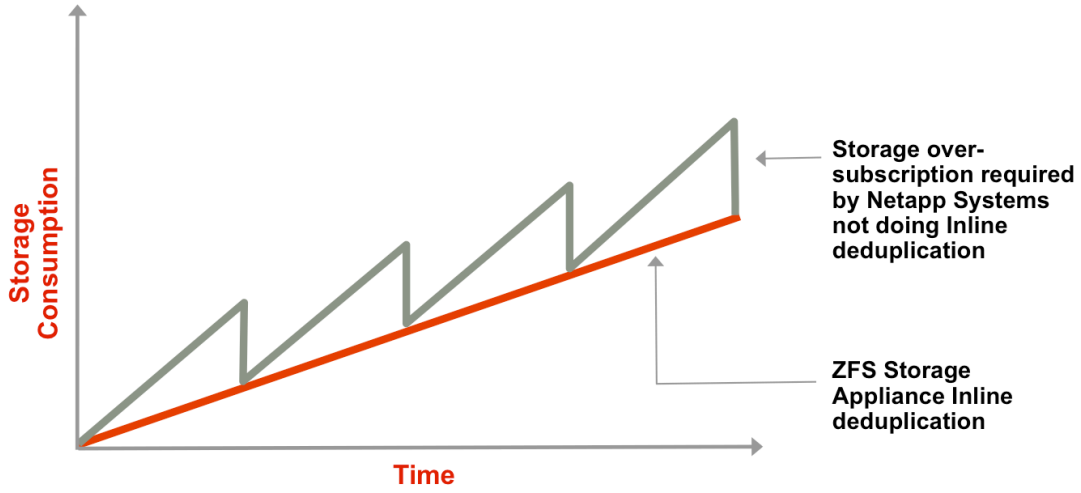

Figure 2. Inline deduplication saves disk space by avoiding storage over-subscription between deduplication runs.

[4] For organizations that want to take the extra step to verify that blocks are identical if their checksums match, Sun ZFS Storage Appliances have a "verify" option that causes the system to perform a full byte-by-byte comparison before deduplicating blocks with matching checksums.

Browser User Interface (BUI) and Real-Time Dashboard

Sun ZFS Storage Appliances come with an elegantly simple-to-use graphical browser user interface (BUI) that gives administrators access to storage management with the rapid familiarity of a point-and-click browser user interface. A real-time dashboard (Figure 3) acts as an administrative home page from which all administration tasks can be initiated. The dashboard provides continuous monitoring of key performance metrics as well as a real-time feed of relevant alerts. Administrators can use the dashboard for at-a-glance usage statistics or as an entry point for running more-advanced Oracle Solaris DTrace Analytics that show detailed usage statistics.

Figure 3. The Sun ZFS Storage Appliance real-time dashboard helps simplify management.

Better Management Visibility with DTrace Analytics

Because most storage management interfaces do not provide detailed visibility into storage subsystems and I/O traffic patterns, enterprises trying to optimize storage performance must use a trial and error process that often leads to subpar results. DTrace Analytics provides the industry's first comprehensive and intuitive analytics environment for storage systems. It uses built-in instrumentation in Oracle Solaris to provide real-time visibility into I/O usage patterns and performance issues. The powerful Intel Xeon processor enables DTrace to provide this real-time visibility without slowing storage I/O, enabling DTrace Analytics to be used while systems continue running in production environments. DTrace Analytics helps reduce the cost of troubleshooting performance issues and can even help organizations avoid additional capital outlays for upgrades by enabling administrators to identify and fix storage bottlenecks.

Figure 4 shows an example of the kind of data that can be visualized with DTrace Analytics. The top part of the screen shows NFS operations per second broken down by client. This kind of detail makes

it easy to identify peak usage periods and also provides insight into which clients are putting the heaviest load on the storage system. The bottom part of the screen shows network I/O coming in and going out of the storage system, giving administrators a chance to see how the storage system is responding in terms of total throughput.

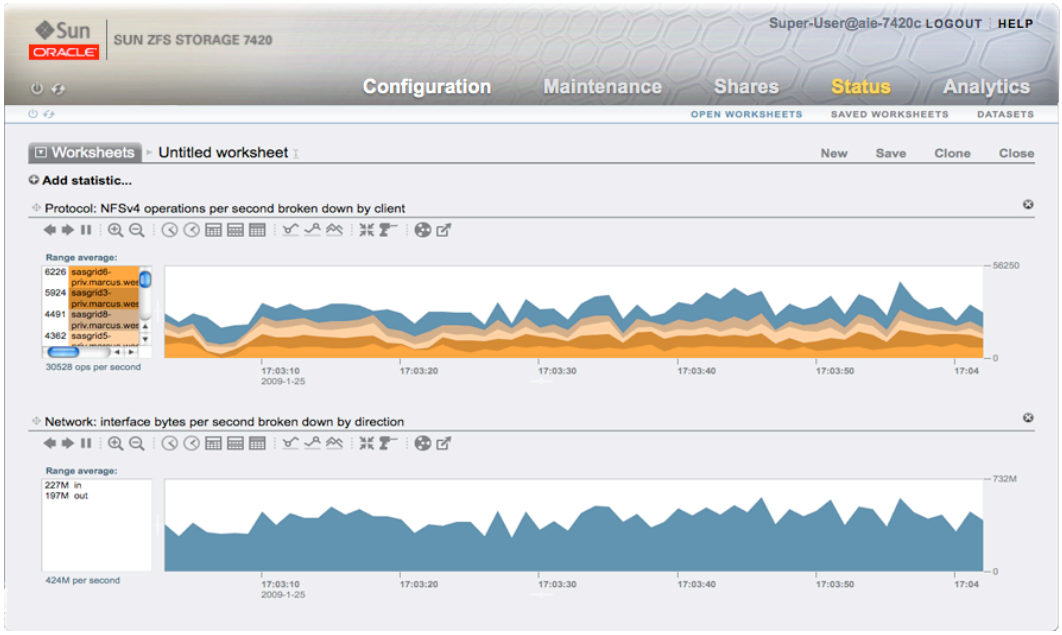

Figure 4. DTrace Analytics screenshot showing real-time storage performance statistics.

Not only does DTrace Analytics give administrators the tools they need to respond to issues to maintain SLAs, it also enables them to proactively manage the environment by providing answers for questions such as:

- How do write-optimized and read-optimized SSDs help specific storage workloads?
- Are CPU, memory, or networking causing bottlenecks?
- What is the read/write/metadata mix of a particular workload?
- How do configuration and application changes affect the storage system?

Space-Efficient Snapshots and Clones

Sun ZFS Storage Appliances enable the use of an unlimited number of snapshots and clones due to the copy-on-write design of the Oracle Solaris ZFS file system. Snapshots are read-only point-in-time copies of data and clones are writable copies. The copy-on-write approach means that no physical storage space is allocated to the snapshot or clone until the data is changed. The powerful multicore multiprocessor platform enables Sun ZFS Storage Appliances to perform snapshot and cloning operations as background tasks that do not materially affect other storage system processes.

Advanced Reliability and Security

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. The subsections below provide an overview of how Oracle Solaris and Intel Xeon processors work together to deliver industry-leading reliability and security.

Protecting Data Integrity and Availability

Protecting data integrity involves the prevention of data errors first, then the detection of data errors that might occur, and finally the containment of any corrupt data to keep it from poisoning other data in the system. The best scenario, of course, is to prevent data loss and securely protect sensitive data in the primary disk storage system.

Oracle Solaris ZFS keeps on-disk data self-consistent and eliminates silent data corruption. 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.

The Oracle Solaris ZFS file system works in conjunction with reliability features on Intel Xeon processors to offer several advanced data protection capabilities that can help increase productivity by improving data availability and data integrity. Many of these capabilities in Oracle Solaris would not be practical without the use of powerful multicore processors. The fact that Sun ZFS Storage Appliances can deliver industry-leading I/O performance while simultaneously executing sophisticated data integrity checks and delivering advanced data services is a tribute to the value of having many Intel Xeon processor threads available for parallel execution.

Intel Xeon Machine Check Architecture (MCA) Recovery

MCA 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 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 Predictive Self Healing

Oracle Solaris Predictive Self Healing automatically diagnoses, isolates, and helps resolve hardware and application faults. When a faulty hardware component is discovered, the self-healing capability in the Oracle Solaris Fault Management Architecture automatically responds by taking the faulty component

offline. This capability is always running in the background of Sun ZFS Storage Appliances without affecting I/O performance. It enables faulty components to be identified immediately and then replaced later during planned maintenance cycles. Sun ZFS Storage Appliances are able to recover and keep running in situations where other storage systems would not. Hardware faults thus become less urgent, allowing IT teams to extend the time between maintenance cycles.

Oracle Solaris ZFS End-to-End Checksumming

Oracle Solaris is the only known operating system designed to provide end-to-end checksumming for all data. Oracle Solaris ZFS constantly reads and checks data to help ensure that it is correct. If it detects an error in a mirrored pool, the technology can automatically repair the corrupt data.

As shown in Figure 5, conventional block-based checksumming (left side of figure) is based on checksums that are stored with the data block, allowing any self-consistent data block to pass the checksum. This approach can only verify media data integrity and can't even detect stray writes. By contrast, Oracle Solaris ZFS validates the entire data block tree (right side of Figure 5), thus validating the entire I/O path. Since Oracle Solaris ZFS validates much more than the media (bit rot), it can catch issues such as phantom writes, driver bugs, and accidental overwrites.

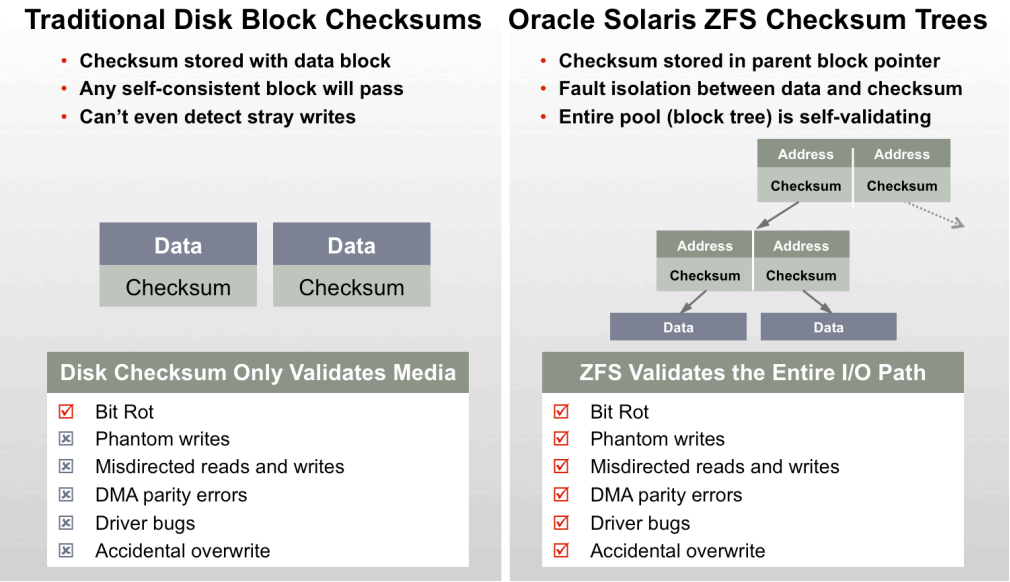

Figure 5. Oracle Solaris ZFS checksums are used to validate the entire data block tree.

This relentless vigilance on behalf of availability protects against costly and time-consuming data loss — even previously undetectable silent data corruption. Corrections are also facilitated by a RAID-Z implementation that uses parity, striping, and atomic operations to aid in the reconstruction of corrupted data. Because of the powerful processing capabilities of Intel Xeon processors, these checksumming operations are performed as background tasks, enabling Sun ZFS Storage Appliances to deliver exceptional I/O performance with this important data protection feature turned on.

Triple-Parity RAID

Sun ZFS Storage Appliances protect data from as many as three drive failures with triple-parity RAID. With disk drives getting larger and many organizations implementing wider striping across more drives for increased performance, a third-parity bit helps reduce the risk of data loss.

By configuring wide stripes to maximize capacity and giving each stripe three disks for parity, the appliance can deliver both high capacity and high availability. This configuration, however, requires more calculations than double-parity RAID, which is one reason why Sun ZFS Storage Appliances are currently the only storage solution that offers triple-parity RAID. Other storage arrays simply do not have the CPU performance power to manage a third parity bit without significantly dragging down I/O throughput.

Triple Mirroring

Triple mirroring maintains two extra copies of a storage volume, providing additional data protection as well as an easy way to increase random read performance for application workloads such as the Oracle database.

Engineered for Energy Efficiency

Energy efficiency is becoming increasingly important as data growth creates sprawling disk farms. Traditional disk-only solutions with 15,000 RPM Fibre Channel hard disk drives are often used to meet today's application performance needs, but these solutions take up extra floor space and require much more power and cooling than Sun ZFS Storage Appliances. Sun ZFS Storage Appliances leverage Hybrid Storage Pools and Intel power-saving features to offer reductions in power and cooling costs.

Hybrid Storage Pools enable Sun ZFS Storage Appliances to deliver high I/O performance with dramatically lower power consumption. For optimal power savings, Sun ZFS Storage Appliances can be configured with capacity-optimized SAS-2 hard disk drives, with performance-optimized SAS-2 hard disk drives, and with multiple SSD caches that help meet high I/O throughput requirements. In most enterprise workloads, the combination of SAS-2 drives and SSD cache enables the Sun ZFS Storage Appliance to deliver higher levels of performance with fewer disk drives and better overall energy consumption.

In addition, engineers from Intel and Oracle have worked together for several years to optimize Oracle Solaris for power management and efficiency when operating on Intel Xeon processors. Optimizations in Oracle Solaris for Intel Xeon processors enable Sun ZFS Storage Appliances to utilize the powerful Intel processors when needed and let them be idle in energy-saving mode when the appliance does not need the extra compute power. The Oracle Solaris kernel dispatcher — the part of the kernel that decides where threads should run — is integrated with the power management subsystem of the Intel Xeon processor family to enable threads to be scheduled according to the power state of the processor.

Intel Intelligent Power Technology

When the processing power of a Sun ZFS Storage Appliance is not being fully utilized, Intel Intelligent Power Technology automatically minimizes power consumption with the following features:

- Integrated power gates allow individual idling cores to be reduced to near-zero power independent of other operating cores. This feature reduces server idle power consumption by up to 50 percent versus the previous generation of two-socket server processors.
- Automated low-power states automatically put processor and memory into the lowest available power states that will meet the requirements of the current workload. Processors are enhanced with more and lower CPU power states, and the memory and I/O controllers have new power management features.

Conclusion

Today's IT managers are challenged to reduce storage costs in the face of continued increases in storage capacity requirements and urgent needs for I/O throughput that can keep up with the requirements of enterprise applications. Sun ZFS Storage Appliances take advantage of the latest features in the multicore Intel Xeon processor family and the enterprise-class Oracle Solaris operating system to meet today's challenging storage requirements. The appliances provide much more processing power and larger memory configurations than those offered by other major storage vendors. At the same time, Oracle's Sun ZFS Storage Appliances do not cost more to own. They simply bring more value.

These capabilities enable Sun ZFS Storage Appliances to deliver the following advantages:

- *Scalable performance* — Sun ZFS Storage Appliances take advantage of Intel Hyper-Threading Technology, Intel Turbo Boost Technology, and Intel QuickPath Technology along with Oracle Solaris optimizations and Hybrid Storage Pools to deliver record-setting storage performance.
- *Innovative data services* — The multicore multiprocessor foundation of Sun ZFS Storage Appliances and the years of optimization work behind Oracle Solaris on Intel processors give the appliance the processing power it needs to deliver innovative data services that were not practical in earlier generations of storage systems.
- *Advanced reliability and security* — Oracle Solaris provides a proven architecture for mission-critical reliability and security and takes advantage of Intel Xeon processor features to protect data as well as detect and recover from errors.
- *High energy efficiency* — New features in Intel Xeon processors and related optimizations in Oracle Solaris enable the Sun ZFS Storage Appliance to reduce operating costs by scaling energy usage to match workload requirements.

For More Information

For more information about Oracle's Sun ZFS Storage Appliances, visit the Web sites in Table 1 or call +1.800.ORACLE1 to speak to an Oracle representative.

TABLE 1. WEB SITES FOR ADDITIONAL INFORMATION

PRODUCT OR SOLUTION AREA	WEB SITE URL
Sun ZFS Storage Appliances	*http://www.oracle.com/us/products/servers-storage/storage/unified-storage*
Oracle Solaris	*http://www.oracle.com/solaris*
Intel Xeon Processor family	*http://www.intel.com/xeon*

Taking Advantage of Multicore/Multiprocessor Technologies to Meet Today's Storage Needs
April 2012

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