

WHITE PAPER

The SunFire X4500: Addressing New Workloads with a New Blend of Server and Storage

Sponsored by: Sun Microsystems

Matthew Eastwood

Jean S. Bozman

December 2006

IDC OPINION

Data-intensive workloads need to be optimized I/O for improved performance and throughput because there are, increasingly, more of them heading straight for the datacenter. As next-generation Web 2.0 applications come online, a variety of technology elements are combining to transform the computing world. With all of the new and old data types coming down the line, new thinking around how to organize the compute, storage, and networking elements of the datacenter is needed to create new datacenter devices and server-based solutions. Ultimately, the "time to result" will improve as the "path" between processors and storage shortens — and as the return trip to storage devices takes less time.

Server and storage vendors have approached the marketplace with combined devices before, often positioning these devices as appliances or "smart" front-end devices to storage arrays. Examples of combined devices already exist, but the intended uses show a large variation in focus, depending on the heritage of the manufacturer. With its new SunFire X4500 product (previously code named Thumper), Sun has designed and produced a device that combines the attributes of server and storage products for the purpose of improving processing efficiency and time to results for a range of technical and commercial workloads. As a manufacturer of both server and storage systems, Sun is well positioned to provide a combined device, address the problem of shortening path length, and improve overall compute results for data-intensive workloads in the high-performance computing (HPC) and commercial spaces.

Servers that have access to large amounts of storage will be able to process the large data sets that are being generated by a new wave of data-intensive workloads — many of which are based on new data types, including images, IP-TV video, geographic information system (GIS) spatial and 3D data, and the very large databases that store all of the data near the system's processors. Without high levels of system bandwidth and ample storage space (ranging from hundreds of gigabytes to multiple terabytes), these workloads would slow down, reducing response time to end users by slowing data retrieval and overall system performance.

IDC would not be surprised if this approach to designing a computer for data-intensive workloads gains ground in the marketplace at large, given the prevalence of data-intensive workloads and the emergence of new data types that require large data sets, such as image-based workloads, video processing, digital content creation, and IP-TV broadcasts, in addition to traditional database-intensive workloads, such as business intelligence and data mining.

SITUATION OVERVIEW

Overview of Worldwide Server Market

The worldwide server market expanded in 2003 and 2004, driven most strongly by sales of volume servers (servers priced less than \$25,000). In 2005, this expansion continued, as the worldwide server market grew 5.9% year over year from \$49 billion in 2004 to nearly \$52 billion in 2005 — and it is expected to reach \$62.6 billion in the year 2010.

Within the worldwide market, various segments also showed strong growth, including the x86 server market with servers based on x86 architecture (including x86-32 and x86-64 processor designs as implemented by AMD, Intel, and other vendors). IDC market data shows that x86 servers account for more than 90% of all server units shipped worldwide and for more than 50% of the factory revenue worldwide.

The proliferation of x86 and x64 servers (supporting 32-bit and 64-bit computing) can be attributed to the widespread ecosystems supporting these platforms, including server vendors, independent hardware vendors (IHVs) supplying the system vendors, and independent software vendors (ISVs) providing software products such as applications, development toolkits, databases, and system management software. Therefore, these servers provide an opportunity for IT managers and the business organizations they support to consolidate workloads on x86 platforms, including technical (e.g., HPC) and commercial (e.g., ERP, database) workloads that run under multiple operating systems used within the enterprise, such as Unix, Microsoft Windows, and Linux.

Optimizing Rapid Storage for Networked Servers

The idea of a "storage server" could be said to have originated with storage vendors that wanted to combine storage area network (SAN) or network attached storage (NAS) hardware with additional hardware containing microprocessors. However, the microprocessors built into those systems are typically associated with running the system and not with performing the computational tasks associated with networked servers that are sending data to the SAN or NAS device.

Most servers encounter two types of delay in accessing data: the time it takes to access a data device, typically measured in milliseconds, and the time it takes to move that data across the network. These two categories of delay result in "latency" in accessing data that is needed for the next set of processing tasks to take place. Historically, systems developed intelligence that allowed them to "prefetch" data that was needed, either by frequency of use or by importance of the dataset to completing the next set of computing tasks.

In many computer systems, software algorithms have been designed to make smart selections of the data sets that need to be "staged" as they make their way to the processing units within a computing system. However, shortening the data-path itself has an effect, especially where the data sets to be transmitted via the network are large — as they would be in HPC or database workloads.

What Would a "Storage Server" Look Like?

In contrast, a storage server would have the advantage of an extremely short path to data storage, which has the effect of optimizing the entire system for higher performance. It would combine the attributes of both server and storage devices, and its efficiency in doing so would determine how quickly data sets could be prepared for processing and returned to the storage device until needed again. The short round-trip between the server and storage components of the system would, itself, speed overall throughput, as it reduces the time it takes to transfer large data sets for processing.

Computing fabrics interweave the enterprise, and they are based on different types and speeds of networking links. That is why customers would expect that the storage server would be able to support multiple types of I/O because it would fit into a network "fabric" that supports many devices, each with its own set of I/O specifications. Included in this mix of I/O types would be PCI-X, PCI-Express, 1 Gigabit Ethernet (1GbE), 10GbE, InfiniBand, and others.

Vendors that can mix the attributes of both server and storage devices would be well-positioned to reduce the total transit time for data within a network that includes both servers and storage and to improve efficiencies in data-processing compared to standalone server and storage devices. Some have looked to high-speed I/O fabrics to reduce the transit time for data to span the distance between computers and the storage devices they use as data repositories.

The computer industry has seen the basic components of computing — compute, memory, I/O, and storage — combined and recombined in a variety of ways over the last two decades. Basic principles of computer science demonstrate that if the distance between these elements can be reduced, then the performance of the overall system will improve and the system can be optimized to support demanding workloads.

Over the years, a variety of approaches have been taken, including:

- Application-specific appliances
- Server-based filers (i.e., file-based storage based on NFS protocols)
- Storage arrays connected to high-speed interconnects
- Optimized server systems that balance the compute, I/O, and storage components, and increase system bandwidth to optimize throughput of the overall system

Summing up, a storage server would combine the best attributes of both server and storage hardware, leveraging common components where possible and leveraging multiple types of high-speed I/O interconnects to link that device with the networks that exist at the edge of the datacenter. It would benefit by reducing the distance between computing and storage elements of the system, which would have the practical effect of speeding performance and optimizing the system's ability to support high-throughput workloads where large data sets are being processed efficiently and quickly.

New Workloads with New Data Types

A new generation of data-intensive workloads is emerging in the marketplace. These workloads begin with large datasets: HPC needs persistent data store; video-based applications, such as video archiving/search and video surveillance applications for security purposes, digital content creation (DCC), which generates thousands of images to manage and store; and applications accessing large databases.

Given the 64-bit computing taking place within the server, the server is capable of addressing multiple terabytes of data directly—a dramatic change from 32-bit systems that have a 4GB limit on the amount of data that can be addressed at any given time. Now, that limit has been surpassed, and multiple terabytes can be addressed directly and access/retrieval time can be limited by the inability to reach the data store quickly.

Workloads that Benefit from Optimized Server/Storage Links

The types of workloads that would benefit most, and first, from this innovation include the following:

- ☒ **HPC workloads.** HPC workloads are typically compute-bound but must access large data sets to set data processing in motion and to store completed results, updating a large persistent store database that retains post-processing data for later access by end users.
- ☒ **Database-associated workloads.** Data warehouse, data mart, and business intelligence (BI) workloads all benefit when access to the underlying database and ability to update that database are improved. Faster searches of large volumes of data speed the time-to-solution by locating the patterns in the data more quickly and getting the results into the hands of end users.
- ☒ **Processing of new data types.** IT is increasingly confronted with the need to process large data sets in new categories of computing. These modes of computing leverage large data sets, such as those for animation/rendering, video archiving (inclusive of video surveillance), new forms of IP-enabled broadcast (e.g., IP-TV), or military applications processing geographic information via GIS data types.

DESIGNING THE SUNFIRE X4500

As already established in computing, distance equals delay. The more distance between two computing elements, the longer it takes to transfer data between two computing elements or between a computing element and a data-storage element. The SunFire X4500 was designed to reduce the distance between the compute and storage elements, with the aim of optimizing system throughput for processing large data sets. HPC and database workloads typically leverage 64-bit computing and generate large data sets, which need a persistent store mechanism.

Technical Specifications

The target markets for the SunFire X4500 are those that use data-intensive workloads, such as large databases (data warehousing and BI) and data-intensive HPC applications (video streaming, video-based surveillance, DCC, and medical imaging storage/retrieval). To support those workload types, each SunFire x4500 has the following technical features:

- ☒ A form factor that combines a two-socket dual-core AMD Opteron computing component with 8 DDR-1 DIMM slots (providing up to 16GB of memory), and up to 48 hot-swappable 3.5in. SATA disks (providing up to 24TB of storage capacity) in a 4U rack-optimized server footprint. This yields up to 6TB per 1U rack factor.
- ☒ A chassis with 4GbE I/O ports and two PCI-X slots, and supports ILOM integrated lights-out management via a dedicated on-board service processor.
- ☒ Solaris 10 U2 operating system, which has the ZFS file system and a built-in volume manager as integrated features of the operating system,
- ☒ The ZFS file system inside Solaris 10 is an extensible file system that can support high-speed transactions and parallel processing, as well as clustering of multiple server nodes. Using the ZFS file system and integrated volume manager, customers can support scale-out processing for HPC or commercial purposes. This approach to managing file systems includes a 64-bit checksum feature that is designed to identify and to avoid "silent" data corruption and data errors that would affect computing results. Because ZFS is a 128-bit file system, it can support an extremely large amount of data — up to multiple terabytes due to the ability to address "2 to the 128th power" in place of "2 to the 64th power" as a theoretical limit bounded by the actual amount of storage accessible by the system.
- ☒ Support for Linux distributions and Java-based software, along with remote lights-out management software.

A Cross-Over Device Combining Server and Storage Components

The SunFire X4500 is a cross-over device that combines server and storage capabilities in a single system. Sun already sells products in both the server and storage categories, but one of the innovations in this product design is that by bringing the server and storage components closer together, the performance and throughput in the device can be optimized and speeded up.

Now, IT managers have an opportunity to leverage new technologies built on standard server and storage hardware components in order to gain performance and throughput improvements that will benefit the business process and the infrastructure components that support that business process. Because the x4500 leverages industry-standard components and interfaces, costs for the total solution can be delivered at lower price points than single-vendor solutions that have been optimized for business intelligence, data warehousing, and other data-intensive workloads.

In Computer Science, Smaller Is Better

This is a consequence of computer science principles that improve performance when latency is decreased — and when computing devices' access to I/O and storage is increased. According to the online encyclopedia, the Wikipedia: "The performance of a computer can be measured using other metrics, depending upon its application domain. A system may be CPU-bound (as in numerical calculation), I/O bound (as in a Web-serving application), or memory bound (as in video editing)."

Therefore, some customers may be unsure just how to regard this new storage server hybrid design, which combines server and storage components in a single, rack-optimized server chassis. However, one way to think of this product is that it has taken the "network" out of the CPU-to-disk connections. Thus, I/O-intensive workloads can access storage nearly as quickly as data requests exit the computing components of the system.

Another key advantage of combining the server and storage capabilities in one integrated system is the small form factor of this system, which fits into the 4U rack-optimized server category. Another is the speed of data transfer, which can be done at "inside the system" speeds, rather than slowing down through a transfer at a slower network speed, which would introduce a dimension of latency into the data-delivery process. Data is transferred to memory at 2GB/second and from the system's disks to the network at 1GB/second.

Compared to the traditional combination of servers and several storage arrays linked by switches that connect the system elements together, the SunFire x4500 consumes approximately one-third of the space for traditional combined solutions that link multiple discrete server and storage systems. In addition, it provides an integrated solution at lower price points (estimated at \$2/GB) and with lower price/cooling requirements than many competing combined server and storage solutions. Its compact rack-optimized form factor also saves valuable real estate within the datacenter because it requires less space than many of the traditional data warehouse systems.

IDC ANALYSIS

Combining the server and storage elements of a computing solution is a variation on the trend toward linking these elements via high-speed interconnects. Indeed, IDC sees high-speed interconnects deployed in high-performance computing and in database-oriented commercial computing to address the latency issues that surface in all workloads that leverage large data sets and large memory to reduce database-scanning times and to speed up database-oriented throughput. When 64-bit computing was introduced to the IT marketplace in the 1990s, both HPC and database workloads benefited through the introduction of 64-bit addressing, which allowed high-speed processing of large data sets. This type of addressing broke through the traditional 4GB limit of 32-bit data processing, by allowing multiple gigabytes of data to be addressed any given time. However, to take full advantage of this 64-bit computing capability, server vendors needed to provide large memory features near the processing complex as well as fast I/O paths to large data stores located in, or near, the processing computer.

Fast forwarding to 2006, a hybrid design is one that accomplishes this goal by packaging the compute and storage elements in near proximity and supporting an array of fast I/O paths between compute elements and data elements. From a storage perspective, the support of fast I/O "pipes" handles requests from multicore processors that are running multithreaded applications or database workloads.

Usage Scenarios

Usage scenarios for this type of hybrid server/storage device could include the following, which are part of a range of scenarios enabled by the server/storage combination within the Sun Fire x4500 system:

- ☒ **Video archiving.** This data-intensive workload can slow delivery times to end users unless it has access to a high-speed network link and a dedicated server. By directing the workload to a SunFire X4500 system, the data storage area is proximate to the compute-serving devices — and both are close to the multiple high-speed I/O ports, accessing 1GbE, 10GbE, or InfiniBand networks. Archiving of video data, including video on demand and video surveillance data for security purposes, is expected to rise in coming years.
- ☒ **Business intelligence (BI).** This workload was historically handled by supercomputers and highly scalable servers due to the need for 64-bit addressing to speed up scans of large databases. The presence of a SunFire X4500 would allow customers to off-load BI functions to a dedicated server that can gather up data from multiple servers on the network — and then process it quickly — as the BI software searches for patterns within the data.
- ☒ **BI and data warehousing (DW).** An approach that works particularly well for this type of workload is a shared-nothing architecture, which allows for massively parallel processing (MPP) of queries and other database operations. In this scenario, clusters of SunFire X4500 servers can be deployed with a shared-nothing architecture to bring large amounts of disk I/O, multicore Opteron processors, and 64-bit addressing to bear on the problem of full-table scans, which are typical of BI and DW workloads; many general-purpose database

systems are not optimized for the large-scale data demand for these kinds of workloads. The Data Warehouse Appliance powered by Sun and Greenplum is one such solution, leveraging the X4500 server and Greenplum Database in turnkey appliance package. The relative performance, horizontal scale-out capability, and economics of this Sun/Greenplum solution could prove disruptive in the high-end BI/DW market, which has been the domain of single-vendor, proprietary database machines.

- ☒ **Video streaming and digital media streaming.** One of the reasons that video streaming was discouraged in some early Web 1.0 installations was the high aggregate bandwidth needed to support this workload in a standard business environment. Now, a dedicated storage server can deliver I/O-intensive video and digital media data streams without slowing the performance of more conventional rack-optimized servers. The arrival of Web 2.0 workloads will make faster throughput a key component of successful video-streaming implementations, in which seamless delivery of video content is one component among many in the end-user "deliverable" on the client desktop.

- ☒ **Server/storage resource for compliance requirements.** The growth in regulatory oversight legislation, particularly in the United States and in Europe, is causing many IT organizations to inventory and evaluate the "state" of their existing archiving technology. This is prompting many companies to expand their investments in data replication, mirroring, snapshot, and other storage techniques to protect data, even in the event of a man-made or natural disaster, and to protect business data and data records from tampering by employees, over time. Archiving, involving near-line and secondary storage, as well as permanent taped backup of all business data, is accelerating. Technologies that support compliance initiatives will be tapped for temporary and permanent storage of operational business data per government and board-of-director requirements.

- ☒ **HPC and technical grid platform.** HPC is increasingly an element of business operations, as automotive and aerospace companies, among others, leverage HPC-style computing in support of engineering and design. Similarly, financial institutions, oil/gas companies and pharmaceutical companies use HPC computing as an important keystone for their ongoing R&D operations. The use of HPC computing in technical "grids" that allocate compute resources, as needed, to support parallelized computing tasks is becoming well-established. The server and storage components of the SunFire X4500 can participate in a "grid-in-a-box" implementation—or multiple units can be used as part of a larger HPC grid that includes a job-scheduler for processing on multiple compute nodes. As for all grids, the presence of persistent data-store resources on the network is an important prerequisite for being able to retrieve results from previous calculations for further processing — and later long-term storage.

Challenges/Opportunities

As with any new product, the Sun X4500 will be evaluated by customers, hardware vendors, and industry analysts to see whether it speeds up data processing in the datacenters where it is deployed. It will also be compared and contrasted with storage appliances and with servers that include near-by storage that optimize 64-bit

processing of large databases for faster scanning of the databases and faster access to the requested data.

A limiting factor in 64-bit deployments is the ability to store large data sets close to the processor complex, which operates at "system speed" rather than at the slower speed of the attached disk system or systems. Network connectivity between the server and the nearby storage is often another source of latency that delays processing and, thus, data results. Therefore, customers are likely to be curious about an alternative device that optimizes throughput by reducing the total path length between compute and storage elements of a system. Moving the data store closer to the server requesting the data and reducing the number of outside-the-box jumps to access that data should have a strong impact on improving overall throughput.

Another challenge will come from storage vendors, which will question why Sun is innovating in this cross-over product area at a time when storage appliances are shipping more widely within the enterprise marketplace. They may also start to seriously consider taking a similar approach to bringing server and storage capabilities closer together and may compete closely with Sun going forward.

There will also be competitive challenges in the marketplace from well-established data warehouse providers (e.g., NCR/Teradata, HP Non-Stop systems hosting data warehouses, and IBM DB2-based data warehouses) that already provide BI and data warehousing solutions. Sun plans to compete with these vendors on the basis of price/performance — and lower cost — in a disruptive innovation play in the market. These kinds of challenges typically involve long sales cycles, but they could pay off for new entrants bringing reduced hardware costs and ongoing operational benefits.

Sun can address these challenges by generating demand for the new devices by proving customer acceptance and adoption for data-intensive workloads and for new workload types and by showing that this is meeting an emerging customer demand in the wider server marketplace, especially in the HPC and digital media segments.

Opportunities, including new and unforeseen opportunities, may emerge that will broaden the market potential for the X4500, as customers try early-ship units, test them, and pilot them for next-generation deployments. In IT sites where system consolidation is leading to data consolidation, there will likely be additional demand for larger block-mode storage devices. In HPC environments, large clusters or grids of compute devices will generate extremely large data sets that will need large persistent data stores to protect data for future use. And, as video-based applications become more widely used, there will be a need to archive and to rapidly retrieve the video data that has been collected or generated.

Future Outlook

IDC expects the worldwide server market to expand from \$54.9 billion in 2005 to \$62.8 billion in 2010, with a CAGR of 2.7% over a five-year period ending in 2010. This increase in IT spend bodes well for increased investment in server systems and storage devices. Drivers for this growth include the following:

- The process of IT transformation, which is allowing customers to replace aging server and storage devices in the installed base

- ☒ The emergence of Web 2.0 workloads that span multiple tiers of the computing infrastructure and combine data resources across the network
- ☒ The use of more and varied data types as part of next-generation workloads, combining video, data, speech, images, spatial data for geographic information systems (GIS), along with traditional data types stored in relational databases
- ☒ Enforcement of regulatory compliance, requiring extensive data archiving of data and email over a period of many years, driving expansion of network storage
- ☒ Emphasis on Dynamic IT, which is the term that IDC uses to describe the process of aligning IT priorities with business priorities. Dynamic IT initiatives have a dual goal: improving IT flexibility and improving business agility. Virtualization, provisioning, and automation through the use of data-use policies within a system management framework are elements of achieving these goals.

CONCLUSION

The digital world has become more varied over time, and the types of data that must be stored as part of regulatory and compliance standards encompass more formats than ever before. Video images, medical images, document images, audio files, and large data sets generated by HPC workloads and database workloads all will need to be readily accessible by other devices on the network.

Sun has taken the initiative to combine the server and storage components of networked computing closer together. Its SunFire X4500 system has been optimized to support quick archiving for a range of data types — old and new — and rapid access to that data once it has been stored. The approach that Sun has taken to bring high-speed data access and high-speed data storage to high-intensity computing workloads will likely have a speed-up effect on applications that are built around next-generation Web 2.0 computing tasks.

Copyright Notice

External Publication of IDC Information and Data — Any IDC information that is to be used in advertising, press releases, or promotional materials requires prior written approval from the appropriate IDC Vice President or Country Manager. A draft of the proposed document should accompany any such request. IDC reserves the right to deny approval of external usage for any reason.

Copyright 2006 IDC. Reproduction without written permission is completely forbidden.