Evaluator Group, Inc.

Comprehensive Information Storage Evaluation ™

THE EVALUATOR SERIES™

SUN STOREDGE 6920 SYSTEM ANALYSIS

APRIL 2005

7720 EAST BELLEVIEW AVENUE • SUITE 210 • GREENWOOD VILLAGE CO 80111

TEL: (303) 221-7867 • FAX: (303) 221-1615

WWW.EVALUATORGROUP.COM

Sun StorEdge 6920 System Analysis

Evaluator Group, Inc.
Comprehensive Information Storage Evaluation

Sun StorEdge 6920 System Analysis

INTRODUCTION

Sun announced the StorEdge 6920 storage system targeted at the mid-tier enterprise in September of 2004 and recently announced a significant set of additional features. The Sun StorEdge 6920 is unique in many aspects and provides opportunities for customers in areas beyond traditional storage systems. This paper will describe the StorEdge 6920, explain the features, and look into the customer value opportunity.

PRODUCT DESCRIPTION

The StorEdge 6920 is based upon the use of Sun's Data Services Platform (DSP) combined with Sun storage arrays as native storage or heterogeneous storage systems attached as external storage. The DSP is operating as a storage controller but uses other storage controllers in an aggregated manner. The key element of front-ending other storage controllers and performing the storage controller function is that it presents a single system image for the attached servers and for administrative purposes providing a consolidation of storage resources.

The Data Storage Platform uses a custom hardware and code implementation. Sitting in front of other storage systems has the potential for introducing extra latency in I/O operations. The DSP minimizes the extra latency in its design with purpose-built hardware to operate on individual requests with minimal overhead.

Figure 1 contains pictures of the front and back of the Data Services Platform.

Figure 1: Data Services Platform – Front and Back View

Architecture

The main components of the StorEdge 6920 are the Data Services Platform and the StorEdge arrays or external storage. The Data Services Platform is a rack mountable device that contains a passive mid-plane with six slots on both the front and back where cards that provide different functionality can be installed. The card types for the DSP are:

- Storage Resource Cards (SRC) – are the main processing elements of the DSP where commands from attached servers are executed and subsequent commands for the attached storage systems are generated. In addition, advanced functions such as snapshot, data replication and the data services to map host-presented LUNs to attached storage (virtualization) are executed in the SRC. The SRC contains five PowerPC processors, four of which perform the command processing from the servers with the other acting as a Control Processor.
- Storage I/O Cards (SIO) – provide the fibre channel connections to the attached storage systems or servers (either directly or through a fibre channel switch). On the SIO card are either eight 2Gb fibre channel interfaces or six 2Gb fibre channel interfaces and one gigabit Ethernet interface. The SIO cards are paired with the SRC cards in a 1:1 relationship.
- Switch Fabric Cards (SFC) – are the switched fabric implementation providing internal connectivity for the DSP. They are arranged in a pair to provide redundant crossbar switching with a 16 x 16 port connectivity for 1.28 Gb/s bandwidth on each internal port. One of the pair operates as a master (active) and the second is standby.
- Management Interface Cards (MIC) – are to communicate with the Storage Service Processor for control of the StorEdge 6920 and perform the administrative tasks on the DSP. The administrative tasks include booting the DSP, downloading code to the processors and collection and reporting of selected information. There are two MICs but only one is active and the other is a standby.

Evaluator Group, Inc.
Comprehensive Information Storage Evaluation

Sun StorEdge 6920 System Analysis

There are two slots on the front of the DSP for SFC cards and four for SRCs. On the back there are four slots for SIO cards and two for MICs. Figure 2 is a view of the DSP showing the structure. The DSP contains redundant power and cooling for the DSP chassis only.

BASIC Data Services Platform

SP - Storage Processor
CP - Control Processor
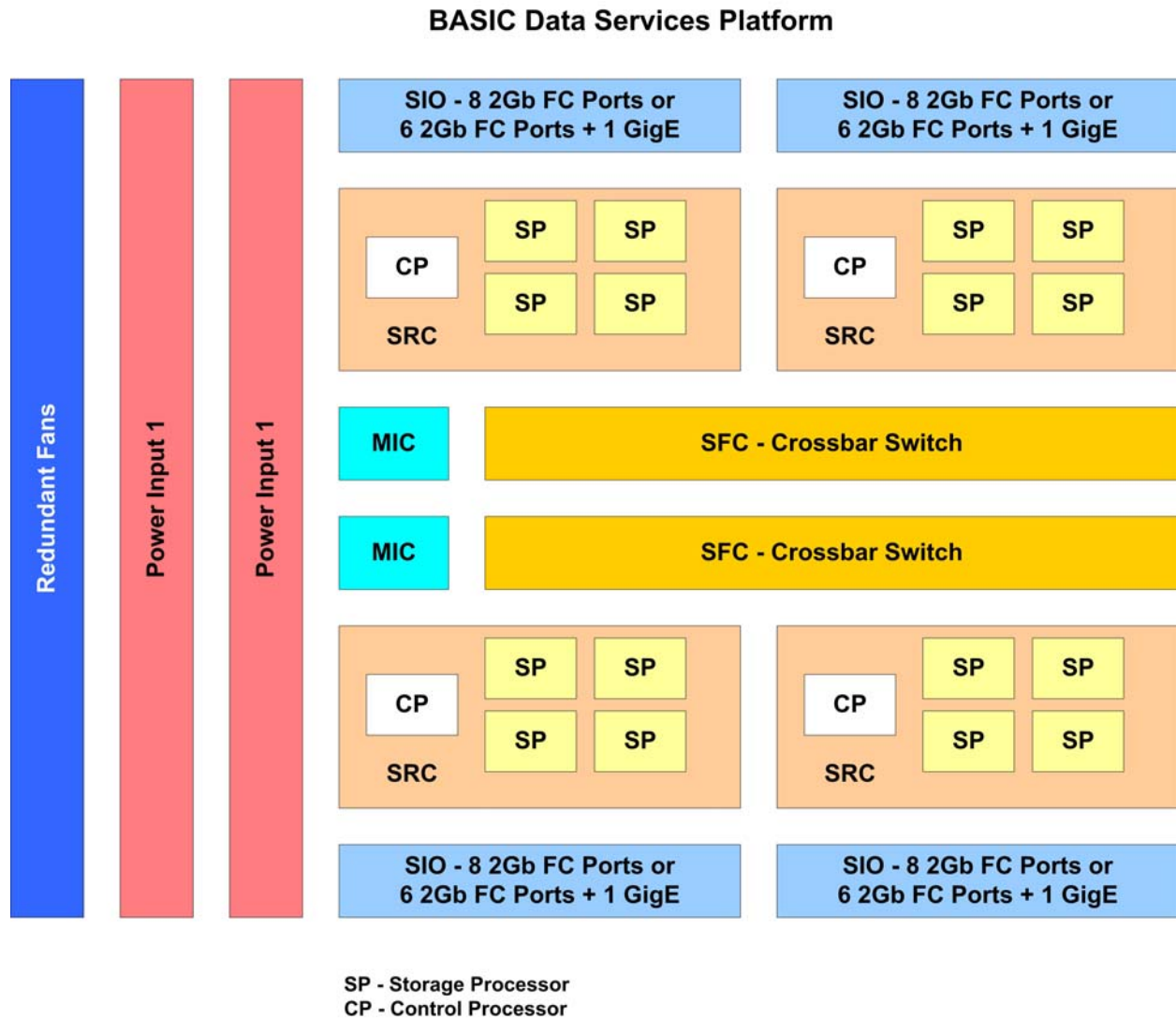
Figure 2: Data Services Platform Elements

The DSP is installed in a cabinet with StorEdge arrays, a Storage Service Processor, an accessory tray, and an Ethernet hub. Additional cabinets with StorEdge arrays and

Sun StorEdge 6920 System Analysis

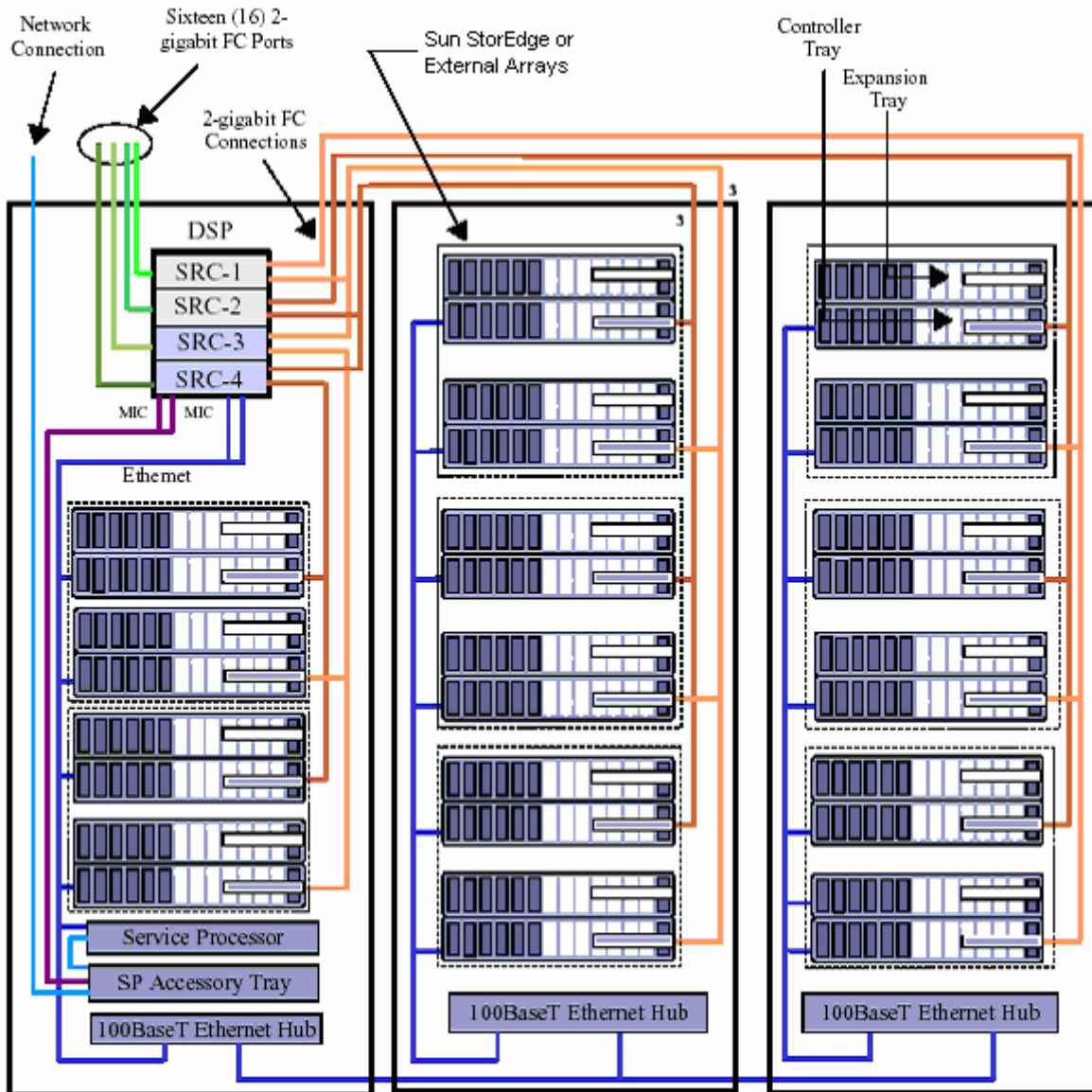Ethernet hubs or heterogeneous External Storage systems may be attached as represented in Figure 3.

Figure 3: StorEdge 6920 System Organization   Source: Sun

The other elements of a 6920 are:

Sun StorEdge 6920 System Analysis

- Storage Service Processor – is a small server that is installed in the same rack as the DSP and is used to manage the DSP by providing the configuration function and serving as the monitoring and reporting point. The configuration data is stored on the Storage Service Processor hard disk and on a removable flash disk. The Storage Service Processor is not redundant and does not have redundant power inputs.
- Storage Service Processor Accessory Tray – contains additional equipment used in the remote support of the 6920. The additional equipment includes a modem, a firewall/router, integrated power supply module, and a network terminal concentrator.
- Ethernet Hub – provides the connection aggregation for all the components in the 6920 system for maintenance purposes. The Ethernet hub is not used in any data transfer operations.
- Power Sequencers – provide for redundant power source inputs which are used by the DSP and the storage arrays. Other elements only have a single power input.
- StorEdge Array – are the same arrays available as the Sun StorEdge 6120 and Sun StorEdge 6320. The arrays consist of a dual controller implementation with each controller having two fibre channel interfaces. Multiple storage arrays may be installed as listed in the configuration description that follows.
- External Storage – heterogeneous storage arrays may be attached to the Data Services platform in the base cabinet.

The embedded software in the Data Services Platform runs on the Storage Processors, the Control Processor, and the Management Interface Card. The Storage Processors handle the I/O requests from a server by mapping the selected LUN to the appropriate volume in the correct storage partition (storage container) and setting up the data transfer between the attached storage system and the server. The Control Processor handles the provisioning and storage management functions. The Management Interface Card processing controls the boot of the DSP and interacts with the device management software from the Storage Service Processor.
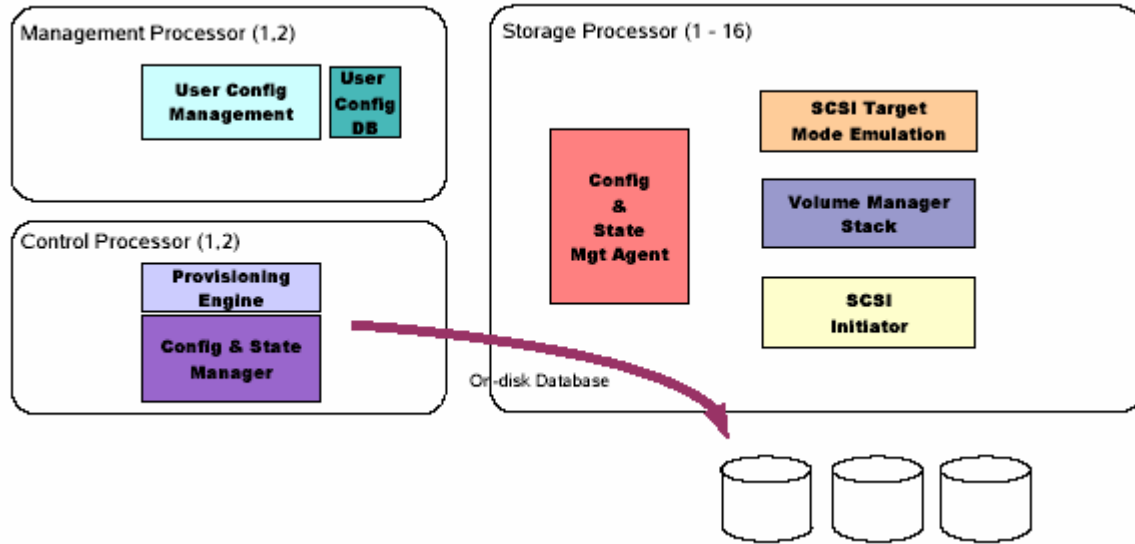
Sun StorEdge 6920 System Analysis

Figure 4: Embedded Software Organization   Source: Sun

Configuration

The configuration of the StorEdge 6920 is determined by the number of cards installed in the Data Services Processor and the number of StorEdge arrays installed. As an example if the StorEdge 6020 storage array is used, there will be a base cabinet with the DSP and 6020 arrays and up to two expansion cabinets with 6020 arrays as referenced in Table 1 using 146GB disks. The 6020 arrays are organized as two controllers with up to six drive trays. The drive trays can have from 7 to 14 drives. The minimum capacity for a drive tray is 504GB and the maximum is 2TB. Up to 32 drive trays may be installed in the 6920.   The drives supported are 36GB 15K rpm, 73GB 10K and 15K rpm and 146GB 10K rpm.

StorEdge 6020 Array	Maximum Drives	Maximum Raw Capacity
Base Cabinet	112	16.4TB
Base + expansion	280	41TB
Base + two expansions	448	65.4TB

Table 1: Cabinet Capacity with 6020 Arrays and 146GB Drives

The DSP ships with two SRCs and two SIO cards. Two additional SRCs and SIO cards may be added for a total of four each. Sun has made the recommendation that the fibre channel ports be split evenly between the server connections and the attached storage

Sun StorEdge 6920 System Analysis

arrays. The number of server connections and attached storage connections will determine the number of SRCs and SIO cards required to be installed.

Features

The StorEdge 6920 contains additional features that are enabled in software. The features may be offered via a separate licensed piece of software or included with the base offering. The additional licensing charge can vary depending on the method employed.

External Storage

External Storage is the attachment of heterogeneous storage systems to the 6920 via the fibre channel interfaces. The external storage may be from Sun or from other vendors. Currently the qualified external storage systems for attachment include:

Vendor	Storage System
Sun	T3B
Sun	6020
Sun	6130
Sun	3500 Series
Sun	9900 Series
EMC	CLARiiON CX Series
HP	EVA Series

Table 2: Support External Storage Systems

Additional heterogeneous storage systems will be added to the qualification list over time. To utilize non-Sun storage systems as external storage, a heterogeneous version of the StorEdge Pool Manager is required which is an additional license fee. If the non-Sun storage is attached as external storage for migration only, the Heterogeneous Pool Manager is not required.

The external storage has two possible ways to be used. The first is as **RAW Storage** which is the same as if it were native storage from the standpoint of allocation and usage of the capacity. Adding external storage as raw storage will mean that any existing data on the storage system will not be preserved but the space management will be handled by the DSP as if it were native storage.

The other possible use of external storage is in what is typically called "passthru" mode where the existing volumes on the external storage are maintained. Sun calls these

Sun StorEdge 6920 System Analysis

Virtual Legacy Volumes (VLV). In this case, the VLV's can be accessed through the DSP and additional features such as remote replication and data mirroring can be applied to those volumes but the data from when they are attached is preserved. By implementing VLV's, a customer can use the DSP for centralized remote replication or for migration from external storage at a convenient time. The overall value of attaching external storage to a 6920 includes protecting an existing investment, centralizing storage capacity management through consolidation, and scaling of additional capacity. The device level management (configuration and service) of the third party external storage will still need to be done with the device management tools offered for those systems.

Sun StorEdge Pool Manager software is the main functional element of the DSP that provides the virtualization of the connected storage systems to the servers attached to the DSP. The Pool Manager software allows administrators to create storage pools across the attached storage systems and then allocate LUNs from the storage pools. The LUNs can range in capacity from 16MB to 2TB and can be dynamically expanded in capacity up to the 2TB limit.

The StorEdge Pool Manager software has pre-defined "storage profiles" which creates storage pools and allocates volumes that are in a form for optimal operation with specific applications. This feature is targeted at reducing additional tuning requirements, easing the provisioning workload, and optimizing storage operations. There are seventeen pre-defined profiles and the administrator can create additional profiles.

For up to 2TB of raw capacity, the Pool Manager software is included with the DSP configuration. Beyond 2TB, a separate license is required but is for an unlimited capacity.

Sun StorEdge Data Snapshot Software is a point-in-time copy implementation to create volume copies. The implementation for data snapshot is a copy-on-write technology where another identity is established for the snapshot copy and only on write operations is the original data copied to a snapshot reserve space location (sometimes called a side file) before the write data is written to the original volume. The snapshot reserve space must be allocated in the same storage pool ahead of time by the administrator. The amount of space to reserve is dependent on the amount of changed data to the snapshot copy and the time the copy may exist. The maximum is 100% per copy but guidelines can be established based on usage cases. In many prior implementations, 30-40% has proven adequate. The snapshot reserve space may be dynamically increased as needed up to fifteen times to avoid an out of space condition for the snapshot copy.

Sun StorEdge 6920 System Analysis

Up to eight snapshot copies may be made per volume. The snapshot copy may be mounted for read/write usage. Data snapshot is a separately licensed product and the charge is in tiers based on the storage capacity that is being copied (and not for the entire capacity of the 6920). With the ability to attach external storage that may be of a different class (in performance and cost), the snapshot copy can be placed on a different class of storage. This enables the ability to move data between classes of storage within the control of a 6920.

The Point-In-Time copy can be "rolled back" to the original volume as initiated by an administrative task. After the rollback, the original PIT copy is not deleted. The rollback process is a background operation in the 6920 and requires the volume be offline to ensure coherency of the data.

Sun StorEdge Data Replicator Software provides remote replication for volumes in the 6920. The replication is from one 6920 to another 6920 over either a fibre channel or Ethernet link of the SIO card. Multiple 6920's may replicate volumes to a target 6920 (called a many to 1 replication) with up to 256 simultaneous replications. The Ethernet connection on the SIO card has TCP/IP acceleration with a TCP/IP Offload Engine to improve performance of replication over the Ethernet interface. Both synchronous and asynchronous replication methods are supported. Write ordering dependencies are handled by the implementation of *consistency groups* where the administrator can group volumes that may have dependent writes. With consistency groups the write order at the remote site is ensured to be maintained.

For synchronous replication, the writes are sent over the remote connection to the remote 6920 and after completion, the primary 6920 will signal write completion to the originating server. The asynchronous replication is a continuous copy where writes are accepted by the primary 6920 and queued on disk for the transmission to the remote 6920 without causing the application to wait on the remote write completion. When write data is present in the queue it is sent over the remote link.

If a failure occurs such that the primary volume (in the local 6920) is unavailable, the secondary (remote) volume becomes the primary volume for access which is called *role reversal*. During this time, changes are tracked with a bit map (called a scoreboard) for the volume and after the problem is resolved and the replication pair is re-established, the modifications are applied to the original primary.

The establish phase of remote replication requires an entire copy of the volume which can be time and bandwidth consuming. The 6920 has an optional *Fast Start* feature where a copy of data (typically on tape) can be made and applied at the remote site while changes are being tracked subsequent to the copy creation at the primary site. Once the initial copy is loaded at the remote site, the establish phase will apply the changes and achieve synchronization.

Sun StorEdge 6920 System Analysis

The Data Replicator Software is an optional feature with a license charge based on the amount of data replicated at both sites. A license will have to be purchased for both the primary and remote 6920's.

Sun StorEdge Data Mirror Software is a local mirroring option where a volume may be dynamically mirrored to another volume. Support exists for 2, 3 or 4-way mirroring. Starting the mirror operation will copy data from one volume to the other(s) in the mirror until a complete mirror exists. Any writes to the primary volume will be mirrored. After completion of the establish, a mirror may be split and the volumes written independently. A change bit map (scoreboard) is kept after the split so that a rejoin may be done and the user can chose whether to *resynchronize* from the primary or *reverse resynchronize* from the secondary copy.

The data mirror function can be exploited in the case of attaching external storage as a means of migrating data. The optional Data Mirror feature is a license charge based on the size of the original volume that is being mirrored.

Sun StorEdge Enterprise Storage Manager software is the storage network manager software offered by Sun. From it, SAN management can be done with a centralized view of the storage infrastructure that discovers the storage network and does topology visualization. Storage management tools to manage devices can be launched from the ESM. The ESM utilizes the SMI-S standard for information content about the storage elements connected to the SAN. The launching of device management tools includes the StorEdge 6920, the StorEdge 9900, the Qlogic switch manager, and the Brocade switch manager. A diagnostic expert is also included to monitor the health of the SAN. This software will be integrated with the AppIQ storage resource manager software that Sun has established as its direction for Storage Resource Management.

Sun StorEdge Traffic Manager software provides for load-balancing and alternate path failover capabilities on servers that are attached to the StorEdge 6920. The software provides performance improvement with the dynamic load-balancing across paths and availability by detecting path failures and redirecting I/O over alternate paths. This software is for individual servers attached to the 6920 and is individually licensed.

Performance

The performance of the StorEdge 6920 is a factor of two fundamental elements: the base performance of the attached arrays and the overhead introduced by the Data Services Platform in its role as the consolidation storage controller. The StorEdge 6020 array has had its performance characterized in the form of multiple controllers offered as the StorEdge 6320 which is shown in figure 5. The 6920 also has been

Sun StorEdge 6920 System Analysis

characterized for performance with multiple controllers of 6020 arrays. Even though there are different numbers of controllers (22 versus 20), there are comparisons that can be made to infer the overhead introduced by the DSP.

The DSP does not do any caching and because of its switch-type architecture adds a minimal amount of overhead. In addition, the DSP will control data placement across the arrays with striping which can improve performance by allowing more disks to be transferring data. The performance charts in figure 5 bear this out and show exactly the same shape which indicates that the overhead introduced by the DSP is consistent and minimal. It also shows that with different storage arrays attached to the DSP, the performance would have an improvement over the arrays without the DSP.

Evaluator Group, Inc.

SPC Posted Benchmark Results 9/23/04
(Multiple Systems - mirrored cache results - Sun comparison)
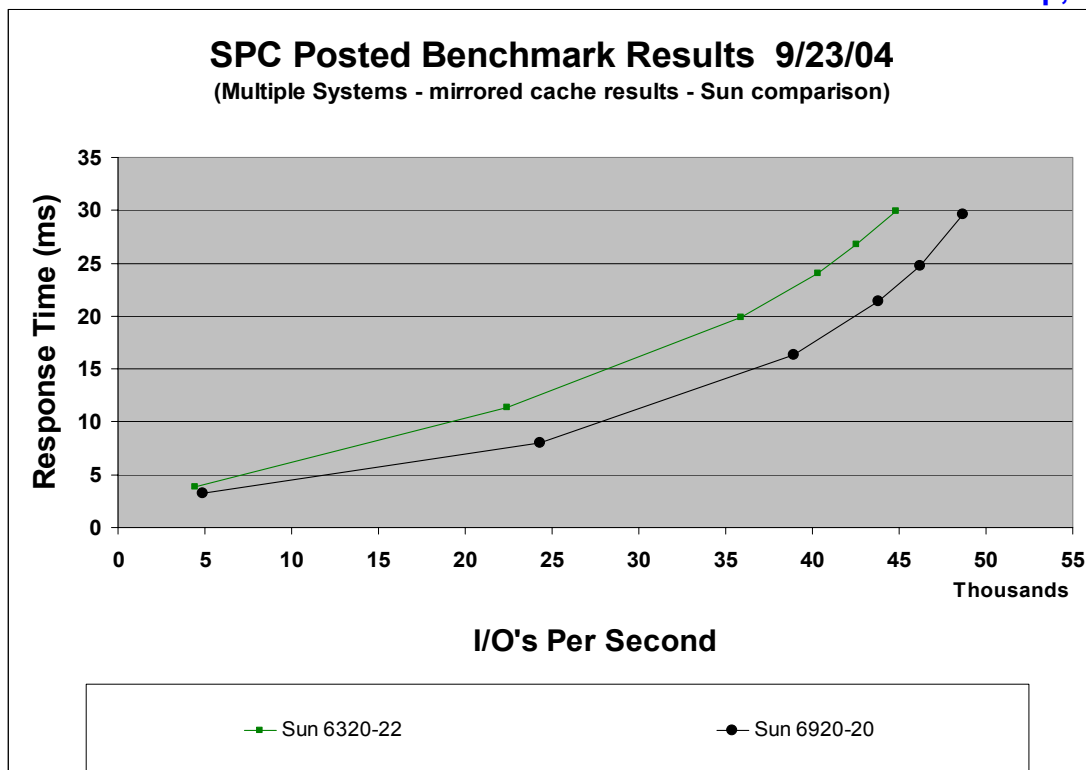
© 2004 Evaluator Group, Inc.

Figure 5: SPC-1 Performance Results

There are other factors that have bearing on the 6920 performance. One is the use of StorEdge Traffic Manager software or other load balancing software on the attached servers to balance the workload across paths to the SRCs on the Data Services

Sun StorEdge 6920 System Analysis

Platform to maximize use of DSP resources. Another is the configuration of the storage pool. There are variables that can be set that influence the effectiveness of the application based on I/O characteristics. These variables include:

- Segment size
- RAID level
- Array type
- Stripe virtualization versus volume concatenation
- Stripe size

Normally in disk systems the amount of cache can have a significant impact on performance. The DSP does not cache any data and relies on the caching in the storage arrays. The 6020 has 1GB of cache per controller (2GB per pair) for read ahead caching and write data caching.

One of the real values in the StorEdge 6920 is its ability to scale out meaning that additional storage controllers (storage arrays) can be added to linearly increase performance with capacity.

CUSTOMER VALUE

There are many high value opportunities for the customer with the StorEdge 6920. Which opportunity is of greatest value depends on the individual circumstances but it is useful to point out the most valuable features of the StorEdge 6920.

An obvious high-value opportunity is the ability to consolidate storage into a single environment by virtualizing the attached storage arrays. Consolidation brings with it many advantages:

- Centralized administration – with fewer storage systems to administer, there is less work to be done and administrators can greatly increase the amount of storage they can handle.
- Increase resource utilization – effectively abstracting the capacity allows better of the resources than if individual storage systems are used.
- Centralized replication – both for control and for single licensing of remote replication and data mirroring.

Another major opportunity related to consolidation is the ability to configure storage partitions called **storage containers**. There can be up to 64 storage containers defined where resources are allocated to the container and administration can be handled on a per container basis. It is easier to do chargeback of resource usage to business

Evaluator Group, Inc.
Comprehensive Information Storage Evaluation

Sun StorEdge 6920 System Analysis

processes or applications when containers are established. Probably the greatest value of storage containers is in consolidation in that an application or business process that had its own storage can still get the value of consolidation while maintaining perceived independence which is typically characterized as meeting specific Service Level Agreements (SLAs).

The ability to scale out where additional controllers can be added to the configuration without having to add more administrative work is a relatively new capability for storage systems and can be of immense value to customers. This is due to the 6920's ability to present a single system image for administration and attachment of servers regardless of the number of storage systems (storage arrays in this case) attached behind the DSP. Scale out or linear scaling of performance and capacity can be exploited by customers having to deal with the demand for additional capacity by providing significant value in maintaining administrative costs while capacity is increased. The single image capability is the main enabler for providing centralized administration over multiple storage systems in a consolidation solution. It also provides for high leverage in using the storage management software and operational procedures developed when additional storage resources are required.

The storage profiles feature of the StorEdge 6920 provides easier configuration and provisioning for customers that should result in administrative savings. Many times the actual details for optimal configuration for a particular application get overlooked or are misunderstood. Having the storage profiles for particular applications is a positive benefit in administration by simplifying the tasks required.

One overlooked benefit but one that may be the greatest economic gain in operational expenses is establishing a long and consistent life for the storage management procedures that may be developed. What's behind this advantage is that storage systems traditionally have a life-span roughly equivalent to the warranty on them which is usually about three years and in a few cases four years. The reason for the warranty period being such and the real motivation to replace storage systems is the change in technology of the disk drives in the storage system and their associated warranty. After the warranty period is over, the vendor necessarily has to charge for maintenance that reflects the expected service characteristics as the systems age and failures begin to occur. For the customer there will be newer, faster disk drives that take up less space so there are other justifications to change technology. With the Data Services Platform, the storage management software and operational procedures can have a much longer life than if storage systems are replaced, typically requiring different management. The storage systems (storage arrays) behind the DSP can be added and replaced without affecting the operational environment. The operational expense savings include the administrative time to add and administer additional new or changed management software, training on new products, and new operational procedures.

Evaluator Group, Inc.
Comprehensive Information Storage Evaluation

Sun StorEdge 6920 System Analysis

RELIABILITY, AVAILABILITY, SERVICEABILITY

Every storage system has an expected set of Reliability, Availability, and Serviceability (RAS) characteristics and the 6920 does as well. What is different about the 6920 is that the RAS characteristics are accomplished by either the Data Services Platform or the attached storage systems. For the purposes of this document, the unique or special RAS features will be highlighted. In addition, there are two offerings that provide enhances RAS.

Storage Automated Diagnostic Environment (StorADE) is software that performs monitoring, fault isolation, and diagnostics for the StorEdge 6920 and executes on the Storage Service Processor. The StorADE software provides a consolidated view of all the components of the 6920 to system monitors and reporting software. The functionality includes a long list of diagnostics, data collection, and reported information. Information can be explicitly requested or will be reported through SNMP traps to system monitoring software.

Sun StorEdge Remote Response Service is a remote monitoring offering to allow Sun personnel at a central site call the Sun Solution Center to monitor the 6920 and perform diagnosis and service operations.

Other notable RAS characteristics:

- The DSP contains redundant components:
    - Alternate paths through the SRC-SIO card pairs
    - Power and cooling
    - Management Interface Cards (one active and one standby)
    - Switch Fabric Cards (one active and one standby)
- The StorEdge arrays also have the redundancy expected of an enterprise storage system including dual controllers with failover, redundant components, RAID protection of disks, and hot sparing.

A Command Line Interface capability is provided for remote configuration through the Storage Service Processor. The remote configuration CLI allows communication through **https** for configuration functions. Also available is a Java-based Thin Scripting Client for installation on select servers to communicate with the Storage Service Processor.

Security

In addition to the RAS characteristics, the security of the StorEdge 6920 is highlighted by a few special capabilities:

Sun StorEdge 6920 System Analysis

- LUN masking prevents access to LUN from hosts not are not authorized.
- Administrative, role-based ID's with individual passwords control authorization of access.
- SSH is used for communication between the Storage Service Processor and the modem and the SSRR server.
- A firewall/router controls access to the Ethernet connection for service on the 6920.
- All management traffic is encrypted.
- Storage containers provide for isolation.

EVALUATOR GROUP OPINION

With the release of the StorEdge 6920 storage system Sun has delivered a product with significant capabilities and great opportunity for customers. The recent feature additions give the StorEdge 6920 functionality in a mid-tier system that is usually only seen in the high-end, cache-centric storage systems that include mainframe connectivity. The scale-out capability has to be seen as a logical evolution of storage systems. The 6920 should have the performance to meet most customer requirements in the mid-tier. The big value is that customers can scale performance and capacity with the 6920 and not compromise on any of the high-value features.

It seems obvious that there will be additional enhancements to improve the offering. One major set of features has been added since the initial announcement. We can expect more capabilities that will improve the customer value to be derived from deployment of the 6920.

We think that the StorEdge 6920 represents a significant technical direction for Sun storage. Sun has the basis for additional products that provide new levels of scalability and consolidation that can be leveraged by customers. With enhanced product offerings, Sun can be very competitive in the storage market.