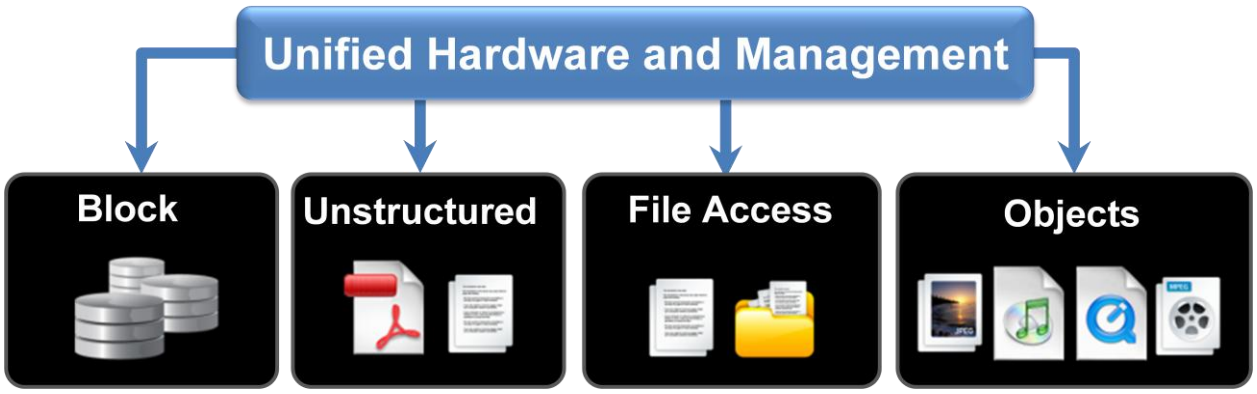
WHITE PAPER

Hitachi Unified Storage: One Platform for Block, File and Object

An Analyst Product Review



By Josh Krischer, Josh Krischer and Associates GmbH

April 2012

Table of Contents

Executive Summary

Hitachi Data Systems, a full subsidiary of Hitachi, Ltd., Japan, is now shipping its new mid-range storage platform branded Hitachi Unified Storage (HUS) 100 series, which is a follow-on model for Hitachi Adaptable Modular Storage (AMS) series subsystems. In addition to adding file services, the new series offers significantly improved performance, enhanced functionality, new advanced features and much better environmental figures. The controller is not commodity hardware, as opposed to several competitive products, but specially designed, full active-active hardware.

Storage Requirements and Challenges in a Weak Economy

Despite the weak world economy, the market for storage subsystems is not showing any indications of slowing down. It shrunk a little in 2009 but fully recovered in 2010 and 2011. The major impact on storage market was the wide adoption of server virtualization, which accelerated storage networking deployment decisions by many small and medium enterprises (SMEs), and the increased growth of unstructured data in organization.

Availability and Business Continuity remain on top of the list of requirements. The nonstop global economy, fierce competition and new levels of service requirements raise the requirements for business continuity. The last decade the midrange storage users requested advanced functionality and flexibility, which allow them to increase utilization, decrease storage management efforts and deploy disaster recovery schemes.

Scalability is a must. A storage subsystem should be able to scale seamlessly in capacity, connectivity or performance, without lowering service levels. The capacity should support multitier storage media: solid state drives (SSDs), performance hard disk drives (HDDs) in different capacities (Fibre Channel or SAS), and capacity [nearline Serial-Attached SCSI (SAS) or Serial Advanced Technology Attachment (SATA)] disks. In simple words, scalability should support tiered storage “in a box.”

Performance has two aspects: the first is the throughput measured in number of I/O per second (IOS), and the second is the response time measured in milliseconds. Performance should meet the Service Level Agreement (SLA) requirements, regardless of the used capacity and the workload. Erratic performance levels irritate users more than a slightly slower but constant response time.

Advanced Functionality is required to cope with storage management. The average organization’s storage capacity grows by approximately 50 percent per year while the size of the storage management staff generally remains the same. Data centers which supported terabytes at the end of the previous decade today support petabytes. The only way to cope with this capacity explosion is sophisticated functionality, advanced automation and user-friendly management tools and interfaces.

Storage Efficiency in usage and energy consumption is equally important. Storage efficiency can be achieved by using thin provisioning, tiered storage, automated data placement, deduplication, compression, small-form-factor HDDs, SSDs and virtualization. Efficient storage subsystems allow for better storage utilization, which translates to lower capacity and lower capital and operational expenditure (CAPEX and OPEX), reduced floor space requirements, lower energy consumption and effective usage of people.

Hitachi Unified Storage 100 Family: Control Unit Architecture

Block Access Modules

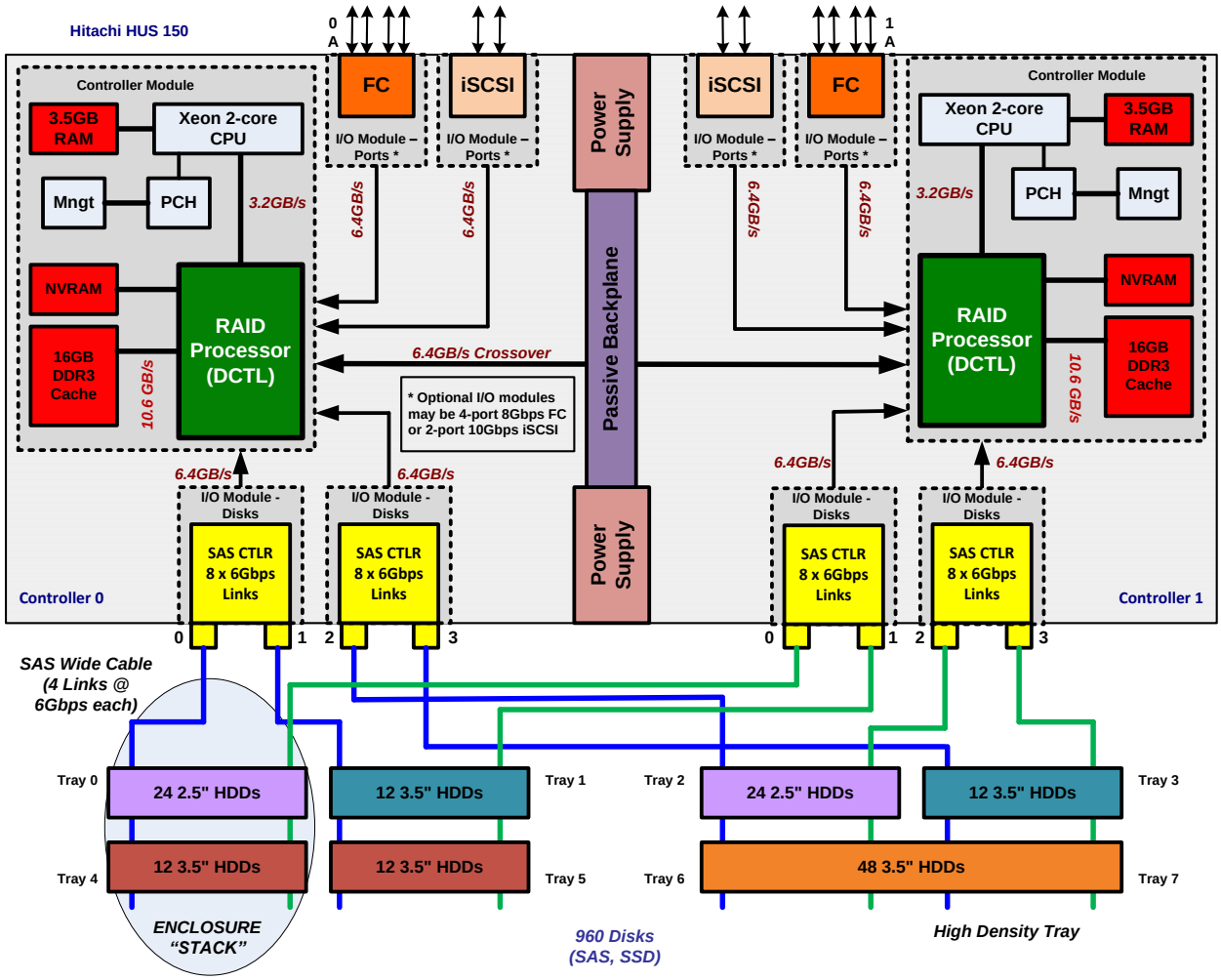
The HUS 100 series unified midrange storage subsystem, unlike some other midrange storage subsystems is a purpose-built storage array. It has 3 different "block" module types (controllers) and 2 different "NAS head" module types for file access. The models differ in scalability, performance and connectivity. As in the high-end Hitachi Virtual Storage Platform, some of the components are specially designed and produced by different divisions of Hitachi Ltd. The "tailored" design is one of the factors in achieving very high performance, large scalability and unmatched reliability. The "heart" of the HUS family is the DCTL ASIC (Data Control Application Specific Integrated Circuit), which, among other functions, is managing the RAID controller and cache memory. All the HUS 100 subsystems use dual controller, symmetrical active-active configuration and mirrored cache. See the top-of-the-line HUS 150 subsystem (3U) block diagram in Figure 2. There is one DCTL chip in each side of the double controller. Additional computing power is contributed by dual core Intel Xeon processors, Tachyon processors and SAS controllers. The Xeon processors manage all I/O operations within a controller for the set of LUNs on their management list.

"All HUS 100 subsystems use dual controllers with Active-active configuration and mirrored cache."

The Tachyon processors manage the front-end host connections and the SAS controllers the back-end HDDs. The control unit is based on a PCI Express (PCIe) 2.0 bus within each controller. These are 8-lane connections, which can provide up to 3.2Gbps send and 3.2Gbps receive (simultaneously) per bus. The storage enclosures are connected via 6Gbps redundant SAS links, 16 links on each side.

HUS 130 and the HUS 110 subsystems use 2U enclosures. The midline HUS 130 subsystem supports 16 SAS connections and fewer disks (see scalability section). The entry-level HUS 110 supports 8 SAS links on the back end and fewer host channels in comparison to the bigger models. It also uses slower processor and smaller cache.

Figure 1. HUS 150 High-end Storage Subsystem Block Diagram/Boards



Source: Hitachi Data Systems
Figure 2 : HUS series high-end subsystem block diagram/Boards, source: Hitachi Data Systems
DCTL: Data Control Processors
SAS CTLR Back-end SAS Director - (disk interface)

File Access Modules (NAS Function)

The HUS series file controllers are installed either as single nodes, or in pairs with a Cluster Management Switch (SMU). The two types of the “NAS head” controllers; HUS 110 and HUS 130 subsystems are based on Hitachi NAS Platform (HNAS) 3080 (dual modules cluster capable) while the HUS 150 subsystem is based on HNAS 3090, with dual node and 4 node clustering. Both models are powered by BlueArc NAS technology. Global file naming is an option. The BlueArc NAS family is an industry-leading NAS series, which uses specially designed hardware acceleration to control file access operations promising very high performance. For example, HNAS 3090 supports much higher scalability with better

throughput that the comparable products from EMC and NetApp (VNX5500 and FAS3270, respectively).

BlueArc was founded in 1998 in the UK. Hitachi Data Systems has sold BlueArc NAS through an OEM deal since 2006 and acquired the corporation in September 2011. Most manufacturers have chosen to build their solutions using "commodity" hardware components, such as Intel Xeon-based processors and chipsets. BlueArc takes a different approach to high-performance scale-out NAS by using dedicated Field-Programmable Gate Arrays (FPGAs) within its architecture to deliver very high scalability and capacity. Unlike commodity CPUs, FPGAs are customized for specific tasks and can be configured to run different processes concurrently. The use of dedicated hardware chips differentiates the BlueArc technology from other vendors' technology in the market today.

In addition to leading throughput and scalability, the platform supports industry-unique functionalities such as native ***Intelligent File Tiering -*** automated policy-based data migration. Intelligent File Tiering allows automated storing files at the appropriate storage tier, thus ensuring optimal cost efficiencies. Combined with Hitachi Content Platform and Hitachi Data Discovery Suite federated search software, the policy manager is able to incorporate content-aware rules to move select files or folders, which are inactive, to an external tier of storage. Thus, the primary dataset is reduced and backup window requirements are decreased.

The Hitachi-BlueArc NAS Platform file-system, SiliconFS, which directs and manages file access performance and scalability, is really an object-based file system, a collection of object structures referring to data on disks, and a set of rules which govern the organizational layout and management of objects on the system. This capability which is unique to Hitachi NAS enables very fast searches, automated intelligent file tiering and faster replication over WANs.

Scalability

Maximum Capacity

The 3 HUS 100 series subsystems support large capacity and different HDD and SSD types which allow "tailoring" configurations to any end-user requirements. Both the HUS 110 and 130 controller enclosures are 2U in height and may be populated with 12 x 3½"Large Form Factor (LFF) HDDs or 24 x 2½" Small Form Factor (SFF) HDDs. There are two different storage expansion enclosures:

1. 2U General Purpose Tray for 24 x 2½" SAS HDDs or SSDs, or 12 x 3½" SAS 7200 RPM Near line HDDs
2. 4U Dense Tray to support 48 x 3½" SAS 7200

The HUS 110 subsystem supports only the General-Purpose tray; the HUS 130 and HUS 150 support both types of expansion trays. HUS 150 supports up to 960 HDDs/SSDs, while HUS 130 and HUS 110 support up to 264 and 120 HDDs, respectively. See Table 1 for the maximum number of supported trays for each expansion type.

Table 1. Tray Support for Hitachi Unified Storage 100 Family Subsystems

Model/Tray Type	HUS 150	HUS 130	HUS 110
12 x 3½ " HDDs	40	19	9
24 x 2½" LFF (HDD/SSD)	40	10	4
48 x 3½ " HDDs	20	5	_

RAID Technique Levels

The possible RAID levels are described in Table 2.

Table 2. RAID Levels

RAID Levels	Minimum	Maximum
RAID-0	2	16
RAID-1	1D+1D	-
RAID-1+0	2D+2D	8D+8D
RAID-5	2D+1P	15D+1P
RAID-6	2D+2P	28D+2P

Please note that the larger the RAID Group, the longer the disk rebuild time is required. When a LUN is formatted, either a 64KB or 256KB (default) RAID chunk size may be selected.

HDD Types

The HUS 100 family subsystems support wide variety of HDDs with different capacities, and performance. The different models are shown in the Table 3**.**

Table 3. Supported HDD and SSD Types

Type	Capacity	Form Factor
HDD, SAS, 10K	300GB, 600GB, 900GB	2½" SFF
SSD, SLCMLC	200GB, 400GB	2½" SFF
HDD, SAS, 7.2K	2TB, 3TB	3½" LFF

See maximum configuration limits in Appendix 1, Table 6.
A typical, theoretical number of maximum I/O per second is described in Table 4.

Table 4. Maximum IOPS per HDD Type

HDD Type	Typical Physical Max. IOPS
3TB SAS 3.5" 7200 RPM	80
2TB SAS 3.5" 7200 RPM	80
900GB SAS 2.5" 10k RPM	150
600GB SAS 2.5" 10k RPM	150
300GB SAS 2.5" 10k RPM	150
200GB 2.5" SSD MLC type	5000
400GB 2.5" SSD MLC type	5000

Host Connectivity (block access)

HUS 150 supports up to 16 host paths (16 x 8Gbps Fibre Channel, 8 x 10Gbps iSCSI or intermix of 8 x 8Gbps Fibre Channel plus 4 x 10Gbps iSCSI), and HUS 130 supports similar connectivity types (16 x 8Gbps Fibre Channel or intermix of 8 x 8Gbps Fibre Channel plus 4 x 10Gbps iSCSI). HUS 110 supports up to 12 host paths (8 x 8GbpsFibre Channel, 4 x 1Gbps or 10Gbps iSCSI or intermix of 8 x 8Gbps Fibre Channel plus 4 x iSCSI). If the storage subsystem is used as unified storage, then each of the file access modules use two Fibre Channel ports on the block access module.

Host Connectivity (file access)

Each of the file modules supports two 10Gb Ethernet (10GigE) ports for file sharing, two 10GigE ports for clustering, six 1 GigE ports for file sharing, and five 10/100Mb LAN ports for management. Additionally, there are two 4Gbps Fibre Channel ports for SAN connectivity, and two 4Gbps Fibre Channel ports for direct connection to the block modules. The network ports can be intermixed; however, both file modules must have identical configurations. Two 10GigE ports for clustering are used as management port and "heart beat" port in a dual-node cluster configuration. The storage subsystem supports the following file protocols: NFS v2-4, CIFS SMB 1.0/2.0, FTP and SNMP protocols SNMP v1, SNMP v2c, SNMP v3.

Upgradeability

The HUS 100 family subsystems allow upgrading from HUS 130 to HUS 150 only. HUS 110 is not upgradeable.

Functionality

Hardware Extended Functionality

All HUS 100 family subsystems support full ***active-active symmetric*** front-end design that allows any host port to access any LUN managed by either controller, without the needs to change the LUN ownership. A host accessing a LUN through a Controller-0 can actually have most of the I/O request processed by Controller-1 with little intervention by the processor in Controller-0[1]. This function allows the use of host-based operating system path management and load balancing such as Microsoft Windows MPIO, Solaris MPxIO/traffic manager, IBM® AIX® MPIO, Linux Device Mapper, Veritas DMP or Hitachi Dynamic Link Manager with minimum overhead penalties. This feature is in particular useful in automated failover.

Another hardware feature is the ***Load Balancing*** of individual LUN workloads between the two modules (controllers). An I/O operation is processed by the controller that currently manages that LUN, which may cause one controllers being more loaded than the other. The HUS 100 family storage subsystems are automatically entering a *load balancing* operation when one controller has an average 70% busy rate while the other is below 40% busy over a period of time of few minutes. After detecting such operation, the load balancing detection mechanism moves to management of one or more of the most active LUNs to the underutilized controller, regardless of which host ports were accessing that LUN. This unique Hitachi feature (can be disabled) reduces significantly the need for tuning and reduces storage management attention and operational costs.

[1] The Tachyon processors directly send each incoming I/O request (one or more FC packets) to the proper Xeon processor from either controller. The Xeon processors know to which Tachyon to route a response to for each host I/O request.

Another unique feature is the ***Hitachi Cache Partition Manager*** *feature*. All known midrange storage subsystems use fix block-size caching algorithm (cache stripe size), resulting in inefficient use of cache and/or I/O operation. This feature optimizes the cache slot size to specific customer application requirements. The results are less cache required for specific workloads, a better hit rate for the same cache size and better performance due to the optimal disk read access size.

The ***Hitachi Cache Partition Manager*** feature optimizes the cache slot size to specific customer application requirements. This results in less cache required for specific workloads, a better hit rate for the same cache size and better performance due to the optimal disk read access size.

There are several more features that simplify storage management and operation, such as ***Hitachi Volume Migration,*** which enables internal copy of volumes (SAS to SSD e.g.) without using host resource or change of host configuration, mount or booting. ***RAID Group Expansion*** allows you to change the RAID group without impacting the data. This can be used, for example, for non-disruptive migration from RAID-5 to RAID-6.

Software Functionality

Despite the fact that HUS series is classified as midrange storage subsystem, it supports a rich portfolio of advanced features. In addition to the usual features, such as Hitachi Copy-on-Write" (CoW) Snapshot, Hitachi ShadowImage® Replication and Hitachi TrueCopy® remote replication, it also supports some very advanced features that are not supported by every midrange storage subsystem. Hardware enhancements in HUS 100 series allow increase in the number of mirrored pairs, more consistency groups and larger capacity of P-VOLs[2] in comparison to the predecessor AMS series, which enable better utilization of the replication features. In addition, ease of use of the mirroring functions were implemented as well; for example, the users don't need to prepare a Hitachi Dynamic Provisioning (DP) pool specifically for snapshots and to specify how much capacity must be allocated for the pool to support CoW Snapshot pairs.

In addition to the block access functionalities, the file module system software supports several advanced features: the BlueArc® JetMirror feature which is a high-speed object-based replication capability over WAN; BlueArc JetClone file clone feature, which creates writable snapshots; and BlueArc JetImage, which enables high-speed steaming backup from disk.

Some of the most important features are elaborated in the next sections.

[2] The data to be protected from the production server on *P-Vol* is mirrored by S-Vol.

Block Access Important Features

Hitachi Dynamic Provisioning (HDP), in addition to thin provisioning, also enables wide striping across all member LUNS in its pool for the creation of high-performance volumes. Striping the data among a large number of physical devices practically eliminates hot spots and allows parallel access to disks, which results in almost uniform performance.

Thin provisioning enables the allocation of virtual storage as needed without having to dedicate the full physical disk storage capacity up front. Additional capacity can be allocated without any disruption to mission-critical applications from existing or newly installed capacity. HUS 100 series, like its predecessor AMS series supports *Zero Page Reclaim.* The *Zero Page Reclaim* function returns unused storage blocks back to the storage pool and reclaims storage space, while *Automatic Dynamic Rebalancing* automatically re-stripes existing virtual volumes when physical volumes are added to the pool for workload rebalancing. Hitachi is the only midrange storage vendor that offers automatic rebalancing of the virtual volume pages through active re-striping in order to take advantage of new disks when the pool is expanded. Hitachi Dynamic Provisioning increases installed capacity utilization, defers upgrades, and saves storage management tasks, thus reducing CAPEX and OPEX.

The new feature, which was not supported by the AMS series, is called ***Hitachi Dynamic Tiering.*** One ability users require of storage is to simplify the building of tiered storage infrastructures. This requirement gains higher importance in times when IT budgets remain unchanged or get slashed. One of the questions when deploying a tiered infrastructure is to determine the appropriate storage tier for the data. An SSD is made up of flash memory that emulates an HDD. There is no dispute that the performance of enterprise solid-state disks is higher than that of traditional magnetic disks (no mechanical movements, no disk rotation and seek latency delays); however, the higher-cost factor (10 to 20 times higher) prevents their use as general-purpose disks. The optimal usage scenario for SSDs is data that is cache-unfriendly to read (random access, non-sequential). Typical candidates include hot database tables, database temporary areas, metadata, indices, control areas, etc. The question is how to identify such data? Placing a whole volume or LUN on an SSD may be not economical, and partial data placement cannot be manually controlled by operators.

Hitachi Dynamic Tiering is a dynamically automated data placement feature that, based on actual usage, dynamically classifies and migrates data to the most suitable tier.

The high-end Hitachi VSP and HUS series are a tiered storage in a box: They support variety of technologies, including SSDs, high-performance SAS HDDs, high-capacity nearline SAS disks, and different RAID techniques. Hitachi Dynamic Tiering is a dynamically automated data placement feature that, based on actual usage, dynamically classifies and migrates data to the most suitable tier. This feature helps maintain peak performance under dynamic

conditions without manual intervention at optimal CAPEX and OPEX savings. The time of the data placement and the sampling frequency can be user-set.

According to the Hitachi storage economics database, more than 50 percent of Hitachi Universal Storage Platform® V and VM, as well as VSP users have deployed the Dynamic Provisioning feature, saving 30% to 40% of real capacity, on average.

File Access Important Features

Transparent Data Mobility

Data stored on the HUS file module is specified by different file attributes (such as age, data type, owner, etc.), which allows transparent relocation of data to an applicable storage "tier." This is a key feature of the BlueArc SiliconFS, which distinguishes the HUS 100 family from the majority of its competition. SiliconFS provides a policy-driven data migration mechanism, which allows data to be migrated transparently between several storage tiers. The ability to transparently move the data between the storage tiers while keeping a single file system presentation to the hosts, users and applications is possible through BlueArc Transparent Data Mobility, and it has several components.

Intelligent File Tiering and Data Migrator

The two major components in BlueArc Transparent Data Mobility are ***Intelligent File Tiering*** *and* ***Data Migrator***. Each file system contains metadata and user data. The metadata is characterized by relatively small capacity and small I/Os; it is accessed more frequently and requires fast access. The user data usually occupies larger capacity and is read or written by larger I/Os. The intelligent file tiering feature automatically separates metadata from user files, and places metadata on the highest available tier to ensure best performance.

Data migrator is the policy-based engine that allows storage administrators to implement their data movement policies. In a multitier storage system, the user data may be migrated to the appropriate tier-based policy setting. Metadata attributes, such as file type, file size, user or group ownership of file, last time of access, and other characteristics can be used to place the data on the optimal storage tier. Data movement may also be scheduled, running a policy check nightly, weekly, and monthly or whatever time period best suits the organizations' requirements.

When data is moved out of band, users must be notified of the new data location and applications have to be "reconnected," which is usually done by Information Lifecycle Management (ILM) host software on competitive products. However, this is not required by Data Migrator. As an embedded feature of SiliconFS, all file system functions (e.g., snapshots, replication, etc.) work seamlessly, as if the data were still on the original storage tier, and data integrity is maintained during the migration or recall. SiliconFS keeps track of where the data actually resides. Therefore, from an end-user and applications point of view, no migration was performed and they see the data as if it's still stored in the original location.

In simple words, this is automated and transparent, "in-box" ILM, which allows customers to build cost-effective, scalable and flexible storage solutions without compromising on performance.

HUS 100 series Local and Remote Copy Techniques

Despite being midrange unified storage subsystems the HUS series support enterprise-level replication techniques for both block and file access.

Block Access Replication Techniques

Like its predecessor, the HUS series supports ***Hitachi TrueCopy Synchronous*** remote mirroring technique and ***Hitachi TrueCopy Extended Distance*** (TCED), which enables hardware-controlled asynchronous remote copying between Hitachi midrange storage subsystems. TrueCopy is supported between two HUS series or between HUS and AMS[3] subsystems. Connections are made via Fibre Channel or iSCSI links. A primary volume (P-VOL) on the local array is copied to the secondary volume (S-VOL) on the remote side. Primary and secondary volumes may be composed of several LUNs.
TrueCopy manages operations spanning multiple volumes as a single group by defining "Consistency Groups[4]." All primary logical volumes of a group are treated as a single entity that allows TrueCopy operations to be performed on all volumes in the group concurrently. An error on any LUN of the group will cease the data transfer in order to maintain data consistency on the recovery site. The TrueCopy management (setup, operation and monitoring) is done by graphical user interface (Hitachi Storage Navigator Modular 2 GUI) or command line interface (Hitachi Storage Navigator Modular 2 CLI).[5] One of the HUS 100 series enhancements significantly eases the operation by simplifying and reducing the number of steps required setting a mirroring pair.

Hitachi Replication Manager software is a part of ***Hitachi Command Suite,*** which configures, monitors and manages Hitachi local or remote replication products across multiple storage subsystems. Replication Manager enables setup of all Hitachi replication products without requiring other tools, including communication port configurations. And it decreases management complexity and provides greater control through a single, consistent user interface.
The TrueCopy Extended Distance use data pool on the local side to store the changed data that accumulates before the next periodical update cycle. The data in the S-VOL following an update is complete and consistent. To avoid data inconsistency that may be the result of a

[3] Hitachi supports copy to the previous generations of AMS such as AMS 2000 (n-1) and AMS 1000 (n-2).
[4] Consistency group defines volumes belonging to the same application or dependant applications in order to keep a consistent image across the volumes in point-in-time replications.
[5] TrueCopy can be set up and perform all basic pair operations such as create, split, resynchronize, restore, swap, and delete from the GUI or CLI.

disaster while the secondary is updated, this consistent data is copied to the remote data pool before next update is to begin. TCED supports consistency groups as well.

As mentioned above, the HUS 100 series supports a ShadowImage point-in-time copy or a space-saving Copy-on-Write Snapshot copy. Larger cache and a new *Memory Management Layer (MML)* have been implemented in the HUS series, and these have significant impact on local and remote copy techniques. For example, the Copy-on-Write Snapshot enhancements include increase of mirrored pairs (from 2046 to 100,000), number of consistency groups (from 256 to 1,024) and the number of pairs in consistency group (from 2,046 to 4,096).

Hitachi Dynamic Replicator

Hitachi Dynamic Replicator is a heterogeneous software local and remote application-aware recovery. This is a technology that Hitachi Data Systems OEMs from InMage corporation. Dynamic Replicator employs a Continuous Data Protection (CDP) technique, which captures changes to data in real time as they occur and offers flexible recovery to any previous point in time. It continuously collects changes as they occur from multiple servers, and sends them (via LAN connection) to a local target called the Dynamic Replicator server. The Dynamic Replicator server (or appliance) transfers this data over an optimized WAN to second location. This approach allows very large data sets and applications to be protected by using limited network bandwidth. As data is captured, it is labeled to allow recoveries to occur from any previous point in time.

Application consistency is kept by Microsoft Volume Shadow Copy Service (VSS) integration on Microsoft Windows and native application-specific mechanisms on Linux and Sun Solaris systems. Application consistency is also enforced at the guest level in all virtual hypervisors. In addition to creating application-specific consistency points, Dynamic Replicator solutions include management of application failover and failback for major environments, such as: Microsoft Windows, SQL Server, SharePoint and Exchange, as well as Oracle, BlackBerry Server, SAP, Linux and Sun Solaris.

In addition to CDP, Dynamic Replicator can be used as simple, cost-effective disaster recovery solutions for long-distance requirements by using asynchronous IP-based replication which supports heterogeneous servers and storage.

File Access Replication Techniques

HUS 100 Series supports block-based replication, both synchronous and asynchronous, as well as file replication.

The BlueArc ***JetMirror*** software provides asynchronous (over IP) high-speed, object-based replication, which creates space-efficient, writeable copies of single files. The object based technique has significant performance benefits in comparison to standard file replication. It support up to 2.8 times faster transfers than full file replication, and up to 26 times faster than file based replication with incremental changes (the expected data transfer rate with file module is 700MB/sec). File clones can be created for both standard files and iSCSI volumes.

JetClone creates space-efficient, writeable copies of single files. File clones can be created for both standard files and iSCSI volumes. Because the file clones are initially created through the use of pointers to blocks, no write to new space or duplication are required at creation. Subsequent new writes are saved to the file system. New clones are created in seconds, regardless of the size of the file being cloned. By using clones, VMware administrators are able to quickly deploy new virtual machines, without consuming additional disk storage space. No limits on the number of file clones, as long as sufficient space exists on the file system to accommodate new data. This allows creation of “check points” and the ability to recover to the last uncorrupted image in the case of malfunction.

Availability

HUS 100 series, similar to Hitachi VSP, continue to maintain HDS's position as the industry’s storage subsystem availability leader. The HUS series provides full redundancy, ensures non-disruptive upgrades and maintenance, supports hot-swappable components, and allows non-disruptive online microcode changes. All components are installed in pairs (controllers, power supplies, fans, interface cards, etc.) and all paths are doubled to ensure full redundancy and avoid single points of failure. Symmetric active-active controllers ensure redundancy without needs to change LUN ownership or reconfigure.

Small-factor HDDs with smaller capacities allow for a fast disk copy or rebuild as well. Write operations are simultaneously duplexed in both cache-0 and cache-1 via the 6.4Gbps special crossover interface. In case of a main power loss or malfunction, the on-card batteries maintain power to enable flushing the outstanding write blocks to the internal flash memory. All HDDs and SSDs are carefully screened before installation.

Almost all microprogram features are proven codes ported from the AMS series promise less or zero “infant deceases” for the HUS series.

The file modules support High Availability (HA) active-active clustering two-node or four (HUS 150 only) node configurations. The clustered servers provide Non-Volatile RAM (NVRAM) mirroring for automated file-system failover. The communication between the

modules is done over a dedicated, out-of-band, High-Speed Serial Interface (HSSI). This interface is used to propagate the server's configuration, for NVRAM data mirroring and the "heart-beat" signals between the modules.

To ensure data integrity in the case of malfunction or loss of power the SiliconFS buffers data in NVRAM until it is written to disk. This data is mirrored to other NVRMAS of the cluster therefore in recovery operation when the surviving server(s) in the cluster assume the functions of the failed module; the mirrored NVRAM is used to complete outstanding data transactions that were not yet committed to disk.

Performance

Major Enhancements in Comparison to the AMS 2500

The HUS series hardware was completely refreshed from its predecessor, AMS. The major enhancements and its implications:

- Newer design and more powerful Intel processor with bigger local RAM increases the maximum throughput (I/O per Second or IOS) of random accesses
- Twice faster internal busses connecting Front-End and Back-End modules with DCTL increase the bandwidth for sequential operations
- Faster DDR3 data cache modules improve random IOPS performance
- Faster SAS links to disks (from 3Gbps to 6Gbps) and allow for more drives per link, faster access to SSDs and increase the bandwidth for sequential operations
- Redesigned DCTL chips have much faster internal paths and twice as many DMA ports[6] (internally) to manage parallel IO requests in-out of cache which increase both random and sequential operations.
- New cache design called ***Memory Management Layer*** (MML)

Early measurements show significant performance improvements in comparison to the AMS 2000 series. Table 5 shows the initial lab-tested performance and improvements in comparison to the AMS series. Note that the disks used in the benchmark were 10,000 RPM SAS HDDs. SSDs were not used in these measurements, and those performance results will come later.

[6] Direct memory access (DMA) is a feature that allows certain hardware within the computer to access system memory independently of the central processing unit (CPU).

Table 5. Performance Comparisons

	Hitachi Unified Storage 150	Hitachi Adaptable Modular Storage 2500	Improvements
RAID-1+0			
Random Read: 100% (IOPS)	77,100	77,382	-
Random Read: 50%, Write: 50% (IOPS)	72,800	55,866	30%
Random Write: 100% (IOPS)	53,600	37,389	43%
Sequential Read: 100% (GB/sec)	5,900	3,155	87%
Sequential Read: 50%, Write: 50% (GB/sec)	3,300	1,277	158%
Sequential Write: 100% (GB/sec)	3,000	1,152	160%
RAID-5			
Random Read: 100% (IOPS)	68,800	69,143	-
Random Read: 50%, Write: 50% (IOPS)	43,500	18,192	139%
Random Write: 100% (IOPS)	20,300	13,871	43%
Sequential Read: 100% (GB/sec)	8,700	3,152	176%
Sequential Read: 50%, Write: 50% (GB/sec)	3,500	1,089	221%
Sequential Write: 100% (GB/sec)	2,900	1,466	98%

As you can see in Table 5, the biggest performance improvements are in sequential data transfer, which fits the shift from structured to unstructured data trends and the growing requirements of applications accessing digitized audio, video, images, scanned documents, X-rays, MRI, mammography and cardiovascular diagnostic pictures, computer simulations, seismic data, etc. Such types of data are usually large files that require high bandwidth and sequential access to data.

In addition to simplifying the remote copy management such as reducing the steps in a pair creation, HUS brings improvements in scalability and performance of data replication. As mentioned above, these improvements are mainly contributed by completely new cache design called *Memory Management Layer* (MML) which is a part of the HUS series firmware.

One of the most significant benefits of the BlueArc NAS heads is the ability to provide sustained, predictable, and consistent performance under various loads regardless of I/O sizes and access patterns. To achieve maximum performance, SiliconFS use stripe set flushing, a technique designed to collate multiple writes into a single disk to reduce seek and latency times. File system metadata contain file attributes, permissions, access histories, etc.; therefore, it is accessed far more frequently than the user data. Separating the metadata from the user data and storing it on the highest storage tier (done automatically by the Intelligent File Tiering) reduces the overhead and improve performance. The proprietary HSSI interface is used to communicate between modules reduces I/O operations and overhead as well.

Hitachi Command Suite v7

Hitachi Command Suite combines advanced data management and protection capabilities in a simple, common (GUI and CLI) user-friendly centralized interface to manage the Hitachi dynamic storage infrastructure, whether it comprises virtualized heterogeneous storage or non-virtualized homogeneous storage. All other large storage vendors use different storage management applications for each platform as opposed to this single interface, which manages the complete Hitachi Data Systems storage portfolio, from the high-end Virtual Storage Platform to the Hitachi Content Platform. Users are able to see all their heterogeneous storage assets as a unified pool as opposed to managing disparate islands of storage, which dramatically improves productivity, reduces risks of losing data, and lowers storage management costs. Version 7 of Command Suite contains many enhancements that simplify usage and enhance functionality. Hitachi Command Suite includes:

Hitachi Device Manager is a core component of the new Command Suite that provides a common framework for configuring and provisioning storage on Hitachi storage subsystems and virtualized external storage.

Hitachi Replication Manager is a software tool for management, consolidation and simplification of all replication operations.

Hitachi Storage Capacity Reporter reports both current and historical capacity views, forecasting and predictive analysis, and custom reporting capabilities.

Hitachi Tuning Manager is an automated, intelligent, and path-aware storage resource management tool that maps, monitors, analyzes, and reviews storage network resources from the application to the storage device. It helps storage administrators troubleshoot performance problems and preemptively identifies and alerts to potential problems.

Hitachi Tiered Storage Manager, as already mentioned, enables the non-disruptive movement of data across all Hitachi storage.

Hitachi Dynamic Link Manager Advanced provides load-balancing and recovery capabilities protecting critical data against various types of failure, such as disk drive malfunctions and other hardware and logical failures.

Integration with Server Virtualization, Major Applications

Storage virtualization amplifies the benefits of server virtualization.
In fact, users can quickly create hundreds of virtual volumes to support any virtual server platform and optimize the placement of virtual applications without wasting time, money or disk space. Tight integration of storage with server virtualization allows organizations to deploy robust, agile, and effective IT infrastructures

VMware

Hitachi Data Systems and VMware have been strategic partners since 2002. Today, Hitachi is a VMware Global Alliance Partner, an elite-level member of VMware's Technology Alliance Partner program and a participant in VMware Ready Partner programs for storage and infrastructure products. Together, Hitachi and VMware are committed to providing innovative, business-enabling end-to-end virtualization solutions for the data center. Hitachi storage and infrastructure solutions, including the newly released Hitachi Unified Storage platform, are catalog-listed as VMware Ready, which is the highest level of endorsement VMware offers.

Hitachi storage solutions are comprehensively tested and certified for interoperability with all versions of VMware. HDS participates in all VMware vStorage programs for storage vendor partners, and the engineers collectively collaborate on creating deeper integrations and new solutions to meet evolving customer needs.

The synergy between Hitachi and VMware enable most key native integrations, including:

- **vCenter Plugin for Storage Management** – allows for correlation between VMware datastores, ESX hosts, virtual machines and storage resources on Hitachi Unified Storage platform, along with all other Hitachi disk subsystems. vSphere administrators can easily manage Hitachi storage through the vCenter Console.
- **vStorage API for Storage Awareness (VASA)** - enables unprecedented coordination between vSphere/vCenter and storage subsystems. It provides built-in storage insight in vCenter to support intelligent virtual machine storage provisioning, bolster storage troubleshooting and enable new DRS-related use cases for storage.
- **vStorage API for Site Recovery Manager (SRM)** - allows customers to perform data center/site failover, ensuring data integrity. The unique combination of Hitachi storage-system-based replication technology and Site Recovery Manager simplifies and automates disaster recovery processes and significantly improves the ability of a customer to be resilient in the event of failover.
- **vStorage API for Array Integration (VAAI)** – enables the hypervisor to offload storage specific tasks to the storage subsystem. The customer benefit is directly proportional to the capabilities and performance of the storage subsystem. Hitachi supports VAAI features natively on its storage subsystems, without the need for 3rd-party plug-ins or special software, reducing complexity, support and management overhead.

- **vStorage API for Multipathing (VAMP)** works with the Hitachi Unified Storage Active-active Controller, as it provides hardware based algorithms for load balancing. This removes complexities for manually managing load within the controller and ESX Host as dual controller subsystems are susceptible to ESX Host workload imbalances that require administrators to spend time manually diagnosing and mitigating.
- **vStorage API for Data Protection (VADP)** – allow for the backup and restore of virtual machines without copying data to a proxy server. This significantly improves backup speed, alleviates large amounts of resources on the proxy server, and allows for many more virtual machines to be backed up simultaneously. During a restore, virtual machines are restored directly to the appropriate ESX Server and datastore, again without the need for staging on the proxy server, providing for much faster restores.

Microsoft SharePoint

SharePoint is gaining momentum, particularly for dispersed workplace organizations. SharePoint Server enables quick and easy development of new applications by individuals, small work groups and departments within these organizations. Over the past few years, organizations that use SharePoint Server and SQL Server as their main data repositories have seen an unprecedented increase in the volume of unstructured content.

Hitachi Data Discovery for Microsoft SharePoint offers significant improvements in scalability, performance and backup of SharePoint environments. It enables customers to tier and migrate SharePoint files to either Hitachi Unified Storage file module or Hitachi Content Platform system (see Hitachi storage portfolio section). Moreover, this advanced functionality provides the ability for archiving content, assigning retention policies and protecting content from modifications according to compliance regulations. When these files are migrated from HUS 100 to an HCP, a *stub* pointing to HCP will be automatically created in the file module database (transparent to users). This provides dramatic improvements in scalability, performance, and backup and recovery time for Microsoft SharePoint.

Hitachi Converged Platform for SAP HANA

The enormous amounts of data collected and stored by organizations, combined with the need to tap volumes of information for better insights, requires new infrastructure for fast data analysis. In November 2011, Hitachi Data Systems announced certified and optimized converged infrastructure solutions for SAP HANA™, which allows organizations to get faster analysis of information in purpose to speed up business decisions. ***Hitachi Converged Platform for SAP HANA*** provides enterprise-class storage performance with computing and processing power to enable SAP HANA customers to quickly harness newly found insights from their data assets. You can expect enterprise storage performance from HUS[7] and

[7] HUS series will support HANA in May 2012 in conjunction with the HANA sp3 release.

scalable Hitachi Compute Blade 2000 for a broad range of high-speed analytic capabilities for full exploitation of SAP HANA.

Sustainability

Many vendors talk about “green IT” or “green storage,” but such things do not exist. Each IT product consumes energy and takes up floor space, so the real differentiation is in how effective it is in comparison to other products. VSP and the HUS series use small form-factor HDDs and SSDs that consume less energy per capacity than large form-factor devices. The smaller 2½ inch HDDs use 50 percent of the power required for traditional 3½ inch drives. The 2U ultra dense drive drawer holds 24 drives of 2½ inch size.

Hitachi Data Systems Storage Portfolio

In addition to the HUS series subsystems Hitachi Data Systems offers an impressive range of storage solutions:

- **Hitachi Virtual Storage Platform (VSP)** is the leading enterprise high-end subsystems on the market.
- **Hitachi High-Performance NAS Platform (HNAS 3000 Series)** is based on the BlueArc Titan series and offers the most scalable, highest-performing clustered network storage system available in the market today. HNAS series scales from 2 (HNAS 3080) up to eight nodes (HNAS 3200) and more than 16PB of storage, while delivering up to 180,000 IOPS[8] per node, and up to 1,600MB/sec per node for sequential workloads.
- **Hitachi Content Platform (HCP)** is a content-addressable storage (CAS) active-archive solution for long-term retention of fixed content. HCP is available as a stand-alone appliance with integrated storage or as a diskless appliance solution supporting all Hitachi block and file storage platforms. HCP supports up to 40PB in an 80 node archive system. It supports options for digital signatures to ensure authenticity, data compression for efficient use of network bandwidth, in-flight data encryption, compression and data deduplication.
- **Hitachi Tiered Storage Manager** seamlessly (with a single click) migrates data across tiers and storage subsystems, at file and at object levels. HDS is the only vendor supporting such migration across discrete storage subsystems; other vendors use host-based software to move data across platforms.

[8] Based on SPECsfs_97 Benchmark.

- **Hitachi Data Discovery Suite (HDDS)** is a key enabler for Hitachi file and content solutions. HDDS is a content services software solution designed to search across the boundaries of all Hitachi storage subsystems or NetApp NAS devices (running Data ONTAP 7.3 running NFS v3 and above). It can use content search to trigger the movement of data among the storage tiers as if it were a single entity.

Hitachi storage subsystems are trusted by the global business community: 44 of the top 50 *Fortune Global 500* companies, the top 15 global commercial banks, the top 16 global property and casualty insurers, the top 10 global telecommunications companies, and the top10 global aerospace and defense companies use Hitachi storage. In 2011, Hitachi Data Systems revenue grew by 18,8% (in comparison on 2010) increasing Hitachi market share to 8.8% from 8.2% in 2011[9] . Hitachi Data Systems achieved also the highest "deliver to promised" ratio[10].

Competition

There are many storage subsystems on the market but only few can be considered as real midrange **unified storage**. The pioneer in unified storage was NetApp, which added SAN capabilities to their NAS head controller based on WAFL/ONTAP. The NetApp FAS 3000 and FAS 6000 series supports NAS, FC and iSCSI but they are optimized for file access. FAS 3270, which can be comparable to HUS 150, supports two controller configuration only as opposed to four on the HUS 150.

In January 2011, EMC announced its VNX and VNXe series, calling it: "Next generation platform that converges CLARiiON and Celerra or unified storage." In reality it is more "bolted together" than unified. In addition to Windows as a base operating system they use DART 30 in the CLARiiON controller and FLARE 6.0 in the Cellera controller. The both parts use different remote and local replications techniques. Different products are also used for data migration; Celerra Replicator is used to migrate the data from Celerra to VNX while SAN Copy migrates from CLARiiON. The VNXe (the e is for entry) represents much more unified storage than its big brother VNX with a unified operating system VNOX. The VNX series supports up to eight x-blades (NAS heads) but does not support single namespace.

In October 2011, IBM joined the unified storage vendors, announcing the Storwize V7000 Unified with unified block and file storage available as new or upgradable from installed Storwize V7000 systems. It provides tightly integrated management console, which simplifies storage administration with a single user interface and common CLI. The Storwize V7000 also provides a virtualization layer for IBM or 3rd-party storage subsystems.

[9] *(Source: IDC Worldwide Disk Storage Systems Quarterly Tracker, March 1, 2012*
[10] Source: The INFOPRO Storage Study 1st half 2011

The VNX supports *FAST VP* and the IBM V7000 supports *Easy Tiering*, which are sub-LUN, automated storage tiering for blocks but don't support equivalent functionality for files without additional software

Summary and Conclusions

By definition, a *unified storage subsystem* should support block access (FC and iSCSI) and file access (NFS, CIFS) however the HUS 100 series is more *universal* than *unified* because it supports objects in addition to block and file accesses. By introducing this series, Hitachi continues with the tradition to deliver a state-of-the art storage subsystem. The HUS 100 series are mid-range subsystems with impressive reliability, functionality, performance, scalability and sustainability. These benefits, and in particular the advanced functionality, such as ***Load Balancing*** , ***Hitachi cache partitioning manager***, ***Hitachi Dynamic Provisioning , Dynamic Tiering*****, *Intelligent File Tiering, Data Migrator*** and the advanced replication techniques position the HUS series above all mid-range subsystems available today.

The HUS 100 series are mid-range subsystems with impressive reliability, functionality, performance, scalability and sustainability."

The HUS series is well integrated with other Hitachi storage products creating a complete portfolio of structured and unstructured data solutions by Hitachi Data Systems. This allows IT to keep data on the most cost-effective media for daily, mid-term and long-term data retention requirements.

Hitachi Command Suite storage management software delivers new, user friendly functionalities for the complete portfolio of Hitachi Data Systems storage solutions. HDS is the only storage company able to manage its full storage portfolio using one common storage management software (Hitachi Command Suite) using the same GUI/CLI.

Full redundancy, non-disruptive upgrades and maintenance, hot-swappable components, pre-emptive soft error detection and online microcode changes ensure high availability and data integrity. The advanced remote data replication techniques enable disaster recovery deployments and support of previous AMS generation, as secondary in TrueCopy, allows investment protection and lower costs in deploying such solutions. The larger cache and the new cache design enable better utilization of the local and remote data replication.

In summary, the HUS 100 series hardware and features position it on the summit of mid-range subsystems therefore each organization considering top quality enterprise mid-range storage should put it on the short evaluation list. Existing Hitachi AMS customers should evaluate the HUS series to get benefits from better performance, remote data replication enhancements and the new advanced functionality.

Appendix 1

Configuration Limits for block modules (Maximums)	*HUS 110*	*HUS 130*	*HUS 150*
RAID Groups	50	75	200
LUNs	2048	4096	4096
Port I/O Request Limit (total queued tag operations)	512	512	512
Port maximum transfer size per I/O	8MB	8MB	8MB
Maximum attached hosts through Fibre Channel virtual ports	1024	2048	2048
Max attached hosts through iSCSI virtual host ports (256 per port)	1024	1024	2048
LUN or DPVOL Size	128TB	128TB	128TB
LUNs or DPVOLs per Host Group	2048	2048	2048
LUNs per RAID Group	1024	1024	1024
HDP Pools	50	64	64
DP Vols	2047	4095	4095
Spare Disks	15	30	80
Cache Partitions	16	32	32
Host World Wide Names per Port	128	128	128

Table 6: Maximum configurations limits

File System Attributes	
Maximum volume size	256TB, dynamically scalable
Maximum virtual volumes	10,000
Maximum files per directory	Up to 16 million or more objects
Maximum file systems per Namespace	128
Maximum snapshots per file system	up to 1024 snapshots per file system one per second per file system

Table 7: File System Attributes

Appendix 2: About Josh Krischer and Associates GmbH

Josh Krischer is an expert IT advisor with more than 42 years of experience in high-end computing, storage, disaster recovery, and data center consolidation. Currently working as an independent analyst at Krischer & Associates GmbH, he was formerly a Research Vice President at Gartner, covering enterprise servers and storage from 1998 until 2007. During his career at Gartner he covered high-end storage-subsystems and spoke on this topic at a multitude of worldwide IT events, including Gartner conferences and symposia, industry and educational conferences, and major vendor events.

Tms

IBM and AIX

BlueArc®