HITACHI
Inspire the Next

WHITE PAPER

Hitachi Accelerated Fabric

Consider the Benefits of the Dynamic Node Fabric Designed for Hitachi Virtual Storage Platform 5000 Series

By Hitachi Vantara

October 2019

Contents

Executive Summary **3**

Introduction **4**

VSP 5000 Series Architecture **4**

Hardware Independence 5

ASIC-Less Design 6

Interconnect Switch 8

Conclusion **9**

Executive Summary

We know that you are under ever-increasing pressures to unlock value from your data so you can run operations more efficiently, improve your customer's experiences and create new services. However, traditional storage systems may not be capable of providing the proper foundation for DataOps initiatives. This is particularly true when current storage solutions can't scale, won't handle more than a single data type, operate in silos and require a lot of manual intervention to run. As a result, data repositories become fragmented and that limits your organizations' ability to gain insights. At Hitachi, we believe that an holistic approach to data storage is the answer.

Hitachi Virtual Storage Platform (VSP) 5000 series, powered by Hitachi Storage Virtualization Operating System (SVOS) RF 9, is designed to deliver the foundation for data-driven infrastructures. Faster than any other storage system, VSP 5000 series is capable of delivering data in real time, even as datasets grow to multi-petabyte scale. It's agile enough to handle any data type and can be integrated into application-optimized converged solutions. It enables you to adopt new technologies when they become available and when you're ready. It scales to store more data to power your business operations and insights. And, it offers unparalleled levels of protection to ensure data is always available and accessible to the right people and never available to those who shouldn't have it.

- **Leading Speed:** A maximum of 21 million IOPS allows for high consolidation of applications. New hardware architecture and intelligent I/O management software boosts the speed of applications with a minimum 70 microsecond response time.
- **Future-Proof Agility:** Support for both serial-attached SCSI (SAS) and NVMe flash as well as traditional hard disk drive (HDD) ensures a smoother transition between technologies purchasing risk that comes from buying into a new technology too soon. Cloud tiering, container integration and support for open systems plus mainframe ensures you are ready for whatever workload is needed.
- **Best-in-Class Scale:** 69PB of raw flash capacity allows a single system to store more data, eliminating stranded storage, cutting new system acquisition costs and reducing the complexity that comes with deploying multiple systems. It also centralizes data for analysis and driving value from data.
- **Trusted Protection:** A superior range of business continuity options and powerful security features eliminate risk of downtime or data loss. Meet the strictest of service level agreements (SLAs) around uptime and data access. And with the industry's only 100% data availability guarantee, you know Hitachi has your back.

In this white paper, we want you to understand the engineering innovation Hitachi is still bringing to the storage industry. Key to this is the development of a scale out and scale up storage controller block architecture that can intermix SAS and NVMe seamlessly, without performance degradation, especially when tiering between different media types. This allows for more applications to run faster while managing the cost of adding more capacity to existing workloads. What does this mean for you in the real world? Well, you can consolidate more onto a single platform. What does this unlock? Well, having more data in one place allows for faster analysis to unlock insights so you can really extract the value from DataOps.

Introduction

Data is used to design and deliver the products and services companies sell. The only way to meet this challenge of a multiprotocol environment with a diverse media portfolio is to accelerate the I/O between the controllers within the architecture. This is where Hitachi Accelerated Fabric comes into the equation. This innovative approach from Hitachi is built upon the 57 years of experience in the data center backed up with more than 350 individual flash technology patents.

With Accelerated Fabric in your business you have a platform that requires no vendor lock-in, and no forklift upgrades as new technology comes down the pike. We designed this technology to future-proof your data to allow you to make use of analytics, machine leaning and artificial intelligence to deliver real-world insights to your business.

What does Hitachi Accelerated Fabric bring to your organization?

- **Speed:** Store and access data in real time, even as datasets grow, to applications that drive business operations and analysis.
- **Agility:** Easily and quickly adapt to include new technologies and infrastructure models.
- **Scale:** Store more data for longer to power business operations and insights.
- **Protection:** Ensure data is available and accessible to the right people to drive operations running and compliant.

VSP 5000 Series Architecture

Hitachi VSP 5500 is designed to scale up and out to meet different capacity, performance and technology requirements. At the core, it begins with what we call a controller block. Each controller block has two nodes, each of which has two controllers, to scale a VSP 5500 system out and up. Resources can fail over across node controllers, across nodes and across controller blocks with quad redundancy to minimize downtime. The first, or base, controller block also includes a pair of node interconnection switches, which provide the backbone of the node fabric. This backbone is the new Hitachi Accelerated Fabric, which enables a dynamic technology and protocol mix without any performance degradation, like InfiniBand. Media chassis are optionally available in any of the controller blocks.

- **Scale from 2 or 4 to 6 nodes:** Each node has dual controllers that provide high-performance data access.
- **Media chassis:** The media chassis stores HDD, SAS SSD, PCIe NVMe SSD, SCM media[1] or Hitachi's flash modules (FMDs). Each media chassis is connected to the two nodes in the same controller block for availability, scale up capacity and performance growth.
- **Two interconnect switches:** These switches create the paths for data between all the controllers in a system, regardless of their location in a node or controller block. This approach enables performance and capacity to scale up and out for efficient use and sharing of resources across the system. It also allows the tiering of data across controller blocks for improved price-performance.

VSP 5500 can grow up to three controller bocks (six nodes or 12 controllers) over time in any combination of SAS, NVMe or diskless controller blocks.

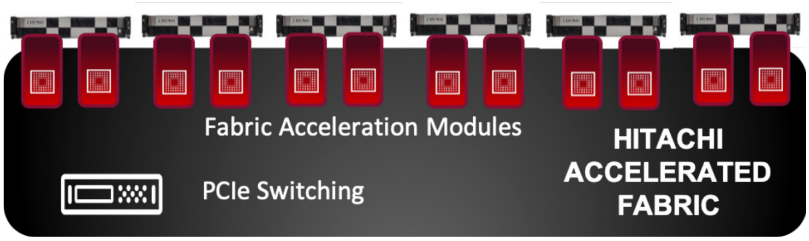
For organizations that want to start small, the VSP 5100 model is available in two-node, two-controller configurations that make use of the Accelerated Fabric technology for performance and capacity. If you need to expand the performance or require more capacity, you can nondisruptively upgrade to a VSP 5500.

[1] SCM media option will be available post-GA.

Hardware Independence

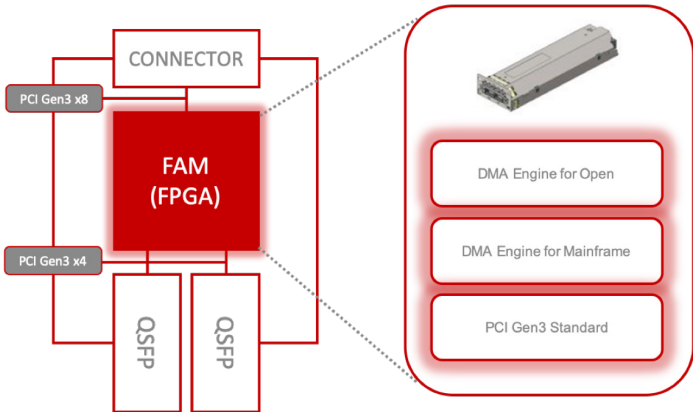
To cope with the demands of scale and also of mixing SAS and NVMe, Accelerated Fabric optimizes the interconnect path between the interconnect on the controllers with the infrastructure switch. This is a PCI-Express Gen3 4Lane link (4GB/s). Each controller has two fabric acceleration modules, each with two ports. Four interconnect paths link the controller to four separate infrastructure switch ports, providing quadruple redundancy for controller connections.

Figure 1. VSP 5000 Series Hardware Architecture



The power of the fabric acceleration modules comes from the field-programmable gate arrays (FPGAs) that are embedded in the interconnect on the controllers. FPGAs allow the controllers to offload processing functionality to them, so SVOS RF 9 can make use of flash-optimized code paths. This means that we use fewer CPU cycles to offer more I/O than anyone else in the storage market, with a peak of 21 million IOPS. For applications that need the fastest possible response times for retrieving critical data, these systems can reach as little as 70 microseconds of latency. It is only with Accelerated Fabric that we reach these performance milestones.

Figure 2. Overview of Fabric Acceleration Module Architecture



FAM = fabric accelerated module, FPGA = field-programmable gate arrays, DMA = direct memory access, QSFP = quad small form-factor pluggable

To accelerate the data processing functions that are offloaded, we built a new direct memory access (DMA) engine. It supports both open and mainframe workloads, allowing for I/O processing across nodes seamlessly, so you can mix SSDs, FMDs and HDDs in one solution! And we can also virtualize any third-party storage array for those of you who have invested in Dell EMC, Pure, HPE, IBM or just about any other vendor. The last thing we want you to do is have to throw this away or have to manage it separately. Virtualizing external storage from third parties or Hitachi provides you with a single management access point for greater efficiencies. It also extends the value-added features that come with the VSP 5000 series to your external storage. For example, the data stored in virtualized systems can be reduced via dedupe and compression to free up additional capacity and extend the life of those systems.

ASIC-Less Design

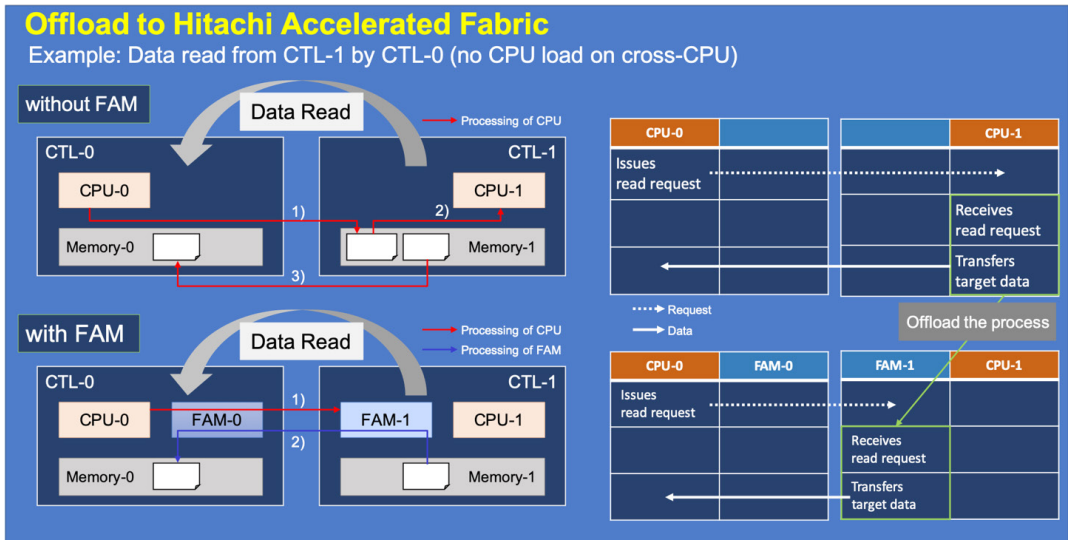
ASCIs, or application-specific integrated circuits, have been a staple part of storage engineering over the years; they have allowed companies to build hardware for a singular purpose. Over time, ASICs have become very complicated, with bespoke programming required to get the best from them and they can cost millions of dollars to develop. FPGAs are based on standard parts and programming logic. By eliminating ASICs from the design of the VSP 5000 series we have been able to lower the cost of production. This way Hitachi can invest our innovative engineering focus onto software (SVOF RF 9) to customize the programming to make the FPGAs deliver more value back to you. Table 1 shows the differences between our VSP G1500 and VSP F1500, and the new VSP 5000 series.

TABLE 1. THE MOVE FROM ASIC DESIGN TO FIELD-PROGRAMMABLE GATE ARRAYS

Number	ASIC Versus FPGA Features	VSP G1500 and VSP F1500	VSP 5000 Series
1	Open (command distribution)	ASIC	SVPS RF 9
2	Mainframe (command distribution)	ASIC	Controller CPU
3	DRR (parity calculation)	ASIC	SVOS RF 9
4	Intercontroller data transfer (mirroring, cross-access)	ASIC	FAM
5	Intercontroller communication (control communication)	ASIC	FAM + SVOS RF 9

Let us now discuss how this actually works when a data read operation is offloaded to Hitachi Accelerated Fabric. In Figure 3, you can see when a read request comes from the CPU between nodes without an Accelerated Fabric. It's the CPU that has to deal with not only the request but also the transfer of data. With the Accelerated Fabric in place, the CPU does not need to use cycles with read requests as the Accelerated Fabric can directly access the memory on the controller to reply to the CPU that initiated the request. This means the CPU can offload cycles, so other impacting technology like data reduction and encryption does not impact the performance of the array. Consider intercontroller communication with and without fabric accelerated modules in Table 2.

Figure 3. An Example Data Read With and Without Using Hitachi Accelerated Fabric



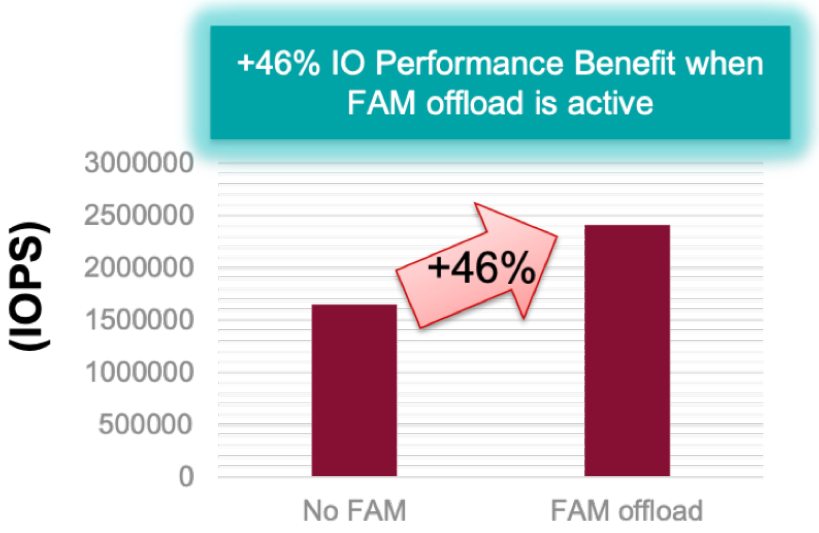
FAM = fabric accelerated modules

TABLE 2. EXAMPLE OF INTERCONTROLLER COMMUNICATION WITH AND WITHOUT FABRIC ACCELERATED MODULES

Number	Type of Communication	VSP G1500 or VSP F1500 (no fabric-accelerated modules)	VSP 5000 Series (fabric-accelerated modules)
1	Read memory on controller-1 from controller-0	Read request is sent to CPU on the target controller, which reads the memory and writes it to destination.	Fabric accelerated modules send the target data directly from memory.
2	Atomic access	Request sent to a CPU on the target controller, which processes atomic operation.	Fabric accelerated modules adjust the atomic operation.
3	Intercontroller transfers of user data	Intel DMA transfers the target to the other controller. After finished, the CPU on the source controller requests to verify it to the CPU on the destination controller, and it verifies the target data by T10DIF calculation.	Fabric accelerated modules transfer the target data to the other controller and simultaneously verify it by T10DIF calculation.

This enables us to offer a simple metric to measure performance gain of using the fabric accelerated modules on the VSP 5000 series or going direct to the controller CPU. From Figure 4, you can see that by offloading some of the intercontroller traffic we can drive a 46% performance increase from a solution. This approach enables us to drive higher I/O with incredibly low 70µsec latency to drive application performance.

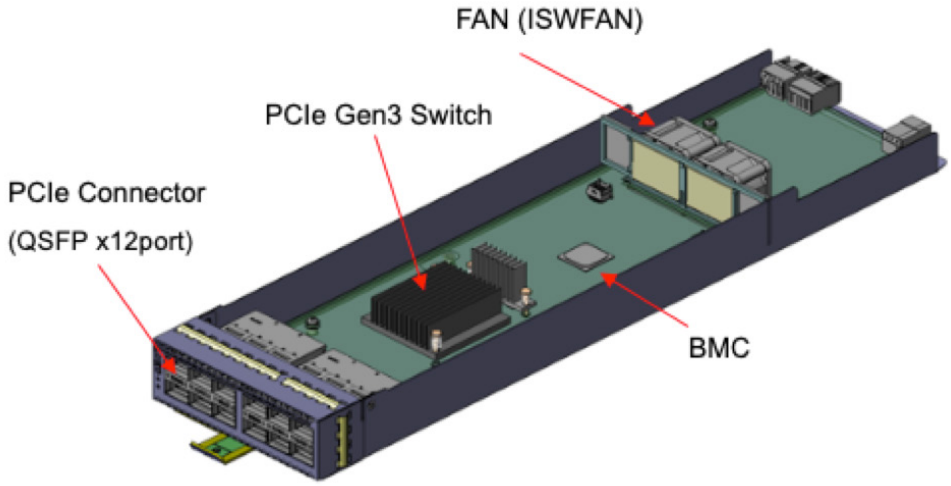
Figure 4. Performance Gain With Fabric Accelerated Modules Turned On



Interconnect Switch

The interconnect switch (Figure 5.) is a key component in this architecture. It is based on a PCIe-Express Gen3 switch module that connects to the fabric accelerated module on each controller board via four-lane link. This switch has 12 PCIe downstream ports for endpoint connections, which are integral to how we route I/O between the nodes. The switching architecture is designed to offer complete redundancy: If you were to suffer a failure, an individual switch in the system is designed to provide full resiliency at all times. The PCIe Gen3 switch is controlled via the root complex port of BMC chip on the board. This allows SVOS RF 9 to directly access the FPGAs to accelerate traffic between the fabric accelerated modules on the controllers to be at the right place at the right time.

Figure 5. Interconnect Switch



VSP 5000 Series Delivers More

The new architecture in the VSP 5000 series is backed up by the industry's first 100% data availability guarantee to help you sleep at night. Specifically, the VSP 5500 is designed to protect against a potential controller failure that may occur while another controller is being upgraded.

Conclusion

Overall, the unique design of Hitachi Accelerated Fabric, using FPGAs to offload I/O management from the nodes, delivers faster data transfers between controllers. It not only reduces latency between nodes in a scale-out storage design but also offloads I/O communication tasks to the fabric. It allows for a seamless technology intermix of SAS and NVMe, giving you the peace of mind to know this platform is truly future-proof. The FPGA modules accelerate processing of I/O communication tasks and PCIe switching, ultimately delivering improved performance.

The system design allows you to be end-to-end NVMe ready, launching with NVMe back end and high-speed 32Gb Fibre Channel front end. NVMe over Fibre Channel (FC-NVMe) will be available post-general availability with just an upgrade to SVOS RF 9 (along with server and host switch upgrades, of course).

With our latest generation of high-end enterprise storage platforms, it is clear that we have your Tier-1 workloads front and center in our design approach. Speed of data access to critical applications is driving revenue for you (and your competitors). Hitachi VSP 5000 series was designed with a fresh approach, as a new enterprise system architecture optimized to accelerate the most demanding workloads. How else could we deliver 21 million IOPS and as low as 70µ latency, in the fastest available high-end enterprise storage platform?

Hitachi Vantara

Corporate Headquarters
2535 Augustine Drive
Santa Clara, CA 95054 USA
hitachivantara.com | community.hitachivantara.com

Contact Information
USA: 1-800-446-0744
Global: 1-858-547-4526
hitachivantara.com/contact

WP-592-A BTD September 2019