Hitachi Unified Storage VM: All Flash Solution

By Leah Schoeb

July 10, 2013


**Introduction**

IT infrastructures are more demanding than ever on storage architectures. It has driven the demand for higher performing storage to handle the growing demand without greatly increasing data center footprints. The increased concentrated workload produced by virtualized environments has been the biggest accelerator for the storage industry to start making new disruptive architectures to keep up. Over 80% of performance issues reported from virtualized environments are about the storage. The ability to handle today’s tier 1 application environments within service level agreements (SLAs) can directly affect how much money a company can make in a second, an hour, or a day. There is a great economic value in accelerating applications to increase business operations and revenue.

All flash systems, a new solid state storage architecture, are now coming to market as new high density, high performing storage solutions to handle the demands of today’s IT infrastructures. Solid state technology changes the dynamics for accessing and using new all flash system technology as a modern architecture and is changing how performance demand in an efficient footprint is being met. Also using an all flash system as a fast layer of read cache in a dynamic tiering solution can be a cost effective way of meeting SLAs.

New flash controllers are optimized to handle the higher performance of solid state technology compared to traditional disk controllers that were designed specifically for the performance of hard disk drives (HDD). The average FC HDD performs at about 200 IOPS per drive while a flash module performs in the tens of thousands of IOPS per module. Traditional storage controllers that supported solid state devices (SSDs) are only able to support a few devices before quickly bottlenecking at the controller. Recently, some of these traditional controllers have been redesigned to handle the higher performance and unique characteristics of solid state devices. But for maximum efficiency a unique flash optimized controller architecture needed to be designed to handle the higher performance of flash modules and SSDs. In addition, flash controllers are designed to handle the unique housekeeping needs and behavior of solid state storage. For example, separating the housekeeping processing that consolidates data to be saved from blocks that are targeted for deletion creates additional data copy steps ahead of that erasure known as garbage collection.

Net effect for customers using all flash system technology improves performance for tier 1 applications, accelerates performance for virtual infrastructures and improves overall performance for any datacenter. All flash optimized systems can offer response times in the microseconds unlike traditional HDD systems that offer I/O response times in the milliseconds. This demonstrates that flash optimized systems are a magnitude faster than traditional HDD systems. The reduction in footprint allows for OPEX saving in power and cooling and the higher density makes for a smaller storage footprint in the datacenter. Additionally, less storage is needed to handle the same workload and all flash systems can reduce the total cost of ownership.
The Hitachi Unified Storage VM (HUS VM) All Flash System has been optimally designed to specifically accelerate and improve performance for tier 1 applications and virtualized infrastructures while offering a high level of storage efficiency and protection.

**Economics**

Solid state storage has been successfully applied in many areas to improve performance of applications and operating environments. Major usages have been in areas such as:

- Online Transaction Processing with utilization of databases
- Analytics software with immediate response time needs
- Collaboration software to permit faster resolution of access
- Server virtualization for increased virtual machine and virtual desktops

Solid state technology provides the most cost-effective benefits to application’s requirements. In general, the technology has minimized response times and permits the greatest number of I/Os. Today’s storage systems are often a shared resource both from multiple accesses from a farm of server and/or high availability requirements.

Due to the advent of new technologies like server and desktop virtualization, IT departments are finding increased challenges meeting SLAs with traditional HDD storage systems. There have been several approaches to meet these challenges, such as: using server side SSDs, adding SSDs to existing storage systems, and adding more HDDs. All of these options will accelerate with limited performance gains.

- Server side SSDs – This can be an inexpensive short term solution, but there are limitations on how much load they can handle with the appropriate data protection and efficiency.
- Adding SSDs to existing storage systems – Since the controllers in traditional storage systems are designed for the performance of HDDs, there is a limited amount of performance that can be gained before bottlenecking at the controller.
- Adding more HDDs – Adding HDDs or expansion units to an existing system are still bound to the performance limitation of the existing controllers. Buying additional storage systems and controllers adds additional power, cooling, and a increased footprint that could lead to storage sprawl.

The cost per gigabyte for solid state storage has been a misguided metric on the true TCO of the product. Finance departments using this as a metric for purchase decisions have been using the wrong metric to really measure the value and total cost of ownership of an all flash system. This has been an inhibitor to the adoption of flash-based storage devices. Getting 10x the performance for a fraction of the capacity begs for more accurate metrics that better describe the TCO of all flash systems. Performance per gigabyte demonstrates the much higher performance by capacity and even price-performance can also show a competitive metric. Combining both of these metrics to demonstrate price/performance/capacity can show the value of purchasing all flash systems.
All Flash System Considerations

All flash systems are a new architecture that can offer several key benefits compared to other flash technologies. They can be shared in a SAN or be used in a tiered solution in a shared infrastructure to support server and desktop virtualization to provide high availability. Compared with hybrid flash systems (supporting both flash and HDD capacity), all flash systems offer faster performance that is also more consistent and predictable. Unlike hybrid systems, their controllers can skip the overhead of caching and tiering algorithms that decide which data sets should be put on flash and the cycles required to actually move that data back and forth between the two storage areas. All flash systems aren't as susceptible to workload variability for highly dynamic environments, such as server and desktop virtualization. There's never the risk of a cache or tier "miss," where the requested data is not on flash as expected and applications are made to wait while the data is fetched from spinning disk and warmed in cache.

All flash controllers are specifically designed for the unique behavior and performance of flash modules and, therefore, the full performance of this solid state technology can be fully utilized. These controllers have the unique features of both efficiency and data protection. When considering an all flash system solutions, there are key features to look for and the efficiency of these features are critical:

- Storage Efficiency and Data Reduction Technology
- Flash media management features
- Integrated Management in virtualized environments
- Supporting advanced features (i.e. Replication)
- Performance and response time
- Works within standard management framework
- Storage virtualization option providing benefits to other storage
- Non-disruptive operations for enterprise applications

Storage Efficiency and Data Reduction Features

All flash systems typically include storage efficiency technologies such as deduplication, thin provisioning and compression. All these features work well in solid state storage.

Deduplication

Data deduplication is a data reduction technique that compares segments of data being written to disk storage with data segments that were previously stored. If duplicate data is found, an additional pointer is established to the original data as opposed to actually storing the duplicate segments, thus removing or "deduplicating" the redundant segments from the storage system. This can sometimes save upwards of 90% or more in capacity.
Deduplication can reduce cost by greatly reducing the amount of the data that is actually stored and is especially well-suited for all flash systems due to the high performance characteristics and efficiencies. Even though there may be a performance impact by implementing data reduction technologies like deduplication for a primary storage system, many manufacturers feel that all flash systems can handle the performance impact without passing the penalty onto users. The capacity savings with deduplication makes the higher cost of flash capacity very reasonable considering the high level of return with performance and has a better marginal return compared to HDDs when running reasonable data reduction ratios in return-on-investment calculations.

**Thin Provisioning**

Historically thin provisioning has been great at eliminating over provisioning capacity and wasting space with HDD. This becomes even more important with a higher performing media. At the cost of flash, administrators are even more sensitive to over provisioning high performance storage to ensure they get the most out of performance while being very efficient with the capacity.

---

Figure 1 – Thin Provisioning

© 2013 Evaluator Group, Inc. All rights reserved. Reproduction of this publication in any form without prior written permission is prohibited.
Business Continuity and Copy Technology

There are a number of reasons why you might want to utilize snapshot, replication, remote replication, and associated technologies, including the need to simplify and speed up restoration after data loss, data corruption or just have a need to save something. Other reasons include: making copies of data for test purposes (including software development, regression testing and disaster recovery testing), making copies of data for application processing, making copies to faceplate non-disruptive backups, and data migration. Managing replicated data is even more important both for replication and remote replication. Centralized management for all copies for reporting and monitoring is ideal.

Local Replication (or Snapshots)

The most popular snapshot technology currently on the market is copy-on-write. The Copy-on-Write Snapshot capability is software that creates a pointer-based point-in-time snapshot copy with space only consumed when a write occurs by copying the original data to a new location. The space efficient snapshot can be created quickly and provide a significant capacity saving over a full copy. Some implementations do provide a means for non-disruptive operations allowing applications to run without performance degradation or disruption.

Remote Replication

There are two main types of remote replication; synchronous and asynchronous replication:

- **Synchronous replication** continuously transfers all data that has changed and the operation waits until the transfer is complete. This means that data must be stored at the remote site before an application can resume. This can have a significant negative impact on performance and there are distance limits. In addition the application will stop if the I/Os cannot complete and the application will become stalled.

- **Asynchronous replication** sends the data but the operation continues without waiting for the transfer to be complete. This means the application will continue to run non-disruptively even if the data is not completely stored at the remote site. There can be a delay in the writes to the remote side and this is usually called a window of “risk” but some implementations take precautions to mitigate this risk. The transmission of data can either be continuous or periodic to the remote site. Since asynchronous replication offers non-disruptive data migration for applications this is a more popular implementation choice.

Auto Tiering

Automated tiering has become a popular offering among most major storage vendors. The ability to automatically and intelligently move part of a LUN or volume to the appropriate storage media based on performance requirements or frequency of use has shown a performance and management value to some IT infrastructures. This is great for taking frequently accessed data and storing it in high performance storage, like solid state. The ratio of data that is frequently accessed is very small, on the average of 3-6% hot blocks for most environments. It’s like the 80/20 rules where 20% of the data...
accounts for 80% of the activity. Therefore, Sub-LUN granularity can be more efficient than volume level movement. Instead of moving entire LUNs or volumes, only the active part of a volume needs to reside on the higher performance tiers. Now there can be pools of Sub-LUNs (pages) represented from different tiers based on their activity level.

Since frequently accessed data tends to be accessed in small amounts, it would be great if there was a way to only access the needed pages and move them into high performance media, like solid state storage. This saves money in the amount of solid state or online media needed to meet high performance requirements so vendors have come up with a way to just move part of a LUN, or sub-LUN, instead of entire LUNS or volumes. Instead of moving entire LUNs or volumes, only the active part of a LUN/volume needs to reside on the higher performance tiers. Pools of these Sub-LUNs (or pages) can also be created and each one can represent different tiers based on their activity level.

**Flash Optimized Controller**

Unlike traditional storage controllers designed for hard drives, flash optimized controllers are specifically designed to handle the unique behavior and performance of flash storage. This includes the ability to manage usage characteristics of data and flash technology requirements.

Storage systems that contain all solid state technology are available and can be classified in two different architectures: one is an existing architecture system that has solid state devices installed that replace hard disk drives and the other is a new architecture targeted at exploitation of solid state technology. Systems designed for use of hard disk drives (HDDs) have covered the requirements for communicating to storage devices with low-level device protocols such as SATA or SCSI and have data transfer designs around the bandwidth and response time for disk drive access. These designs are very effective with HDDs but may not be optimal with the different characteristics of solid state technology. Use of solid state technology in the form of SSDs or flash modules with low-level device protocols will limit the amount of changes required to the design of the existing storage system.
Hitachi Unified Storage VM (HUS VM) All Flash Solution

What is HUS VM?

The HUS VM is targeted for usage in the entry enterprise space for customers who need either a SAN block storage, file storage, direct connect with high performance storage, or all of the above. The HUS VM brings many advanced features and capabilities and supporting software from the high-end HDS systems. Unified via Hitachi Command Suite management and Hitachi network attached storage file modules, the storage system uses hardware and software from the Virtual Storage Platform (VSP) and Hitachi NAS (BlueArc Mercury) to provide both block and file. The HUS VM includes NAS capability supporting CIFS, NFS, FTP, and HTTP access for file and object storage. The HUS VM file functionality is the same as offered with the other HUS systems.

The systems can scale up to 3PB and support 128TB LUNs and 256TB file systems with a single namespace. HUS VM supports several types of media with a choice of 2.5” or 3.5” hard drives, solid state drives, and now Hitachi Accelerated Flash modules. Reliability, availability and serviceability (RAS) characteristics include multiple concurrent RAID levels, disk scrubbing, dynamic FRU replacement, dynamic microcode upgrades, and remote maintenance with call-home capabilities. Additional RAS features include mirrored cache (with dual controller configurations), battery backup of cache, and global hot spare disk drives and redundant power supplies and cooling. Advanced features of VSP and Hitachi NAS Platform (HNAS) are available with HUS VM.

- Protocols supported - FC and iSCSI, NFS, CIFS, HTTP file, and object storage
- Monitoring and Tuning performance – Hitachi Tuning Manager, Performance Monitor and Hitachi Command Director
- Auto Tiering – Hitachi Dynamic Tiering and Hitachi Tiered Storage Manager
- Data Reduction technology - Thin provisioning (Hitachi Dynamic Provisioning) and file level deduplication
• **Business Continuity and copy technology** –
  - **Snapshots, local replication** (Hitachi ShadowImage Replication, Hitachi Thin Image, Hitachi Replication Manager),
  - **Remote replication** (Hitachi TrueCopy and Hitachi Universal Replicator)
  - **High availability** - their 100% availability guarantee
  - **Non-disruptive updates** - for both software and hardware components

• **Hypervisor support** – VMware VAAI and Microsoft Hyper-V

• **Storage virtualization** - virtualize and manage your multivendor storage

• **Clustered file option**

Like Hitachi Virtual Storage Platform (VSP), Hitachi Unified Storage VM is based on the Hi-Star crossbar-switch architecture. This architecture provides a shared resource system where all the primary components, including front-end ports, back-end ports, cache modules and processors, are connected through an internal switched network. Each controller utilizes two types of blades, a cache blade and a microprocessor blade. The cache blade provides the cache and all functionality other than running the system software, that being provided by the Intel Xeon processor on the microprocessor blade. There is a custom ASIC at the heart of each cache blade integrating many of the distributed elements within the VSP design. Should a blade fail, the system will continue to run at reduced performance using the surviving blade.
Figure 4 – HUS VM Controller Architecture

What’s New with HUS VM All Flash System?

The HUS VM now supports Hitachi Accelerated Flash modules and the Hitachi Base Operating System has been specifically re-architected to handle the unique behavior and performance of flash in order to build an all flash array integrated with the supported advanced features of the HUS VM. HUS VM with flash offers:

- **Hitachi Accelerated Flash Storage** – build an all flash system with HAF for high performance low latency storage
- **Dynamic Tiering** supports HAF and places frequently accessed data on HAF and data that is access less frequently on more cost effective HDDs. HUS VM is completely upgradable to a tiered storage system with hard drives, extends to third party storage with Hitachi external storage virtualization
- **Flash accelerated virtualization** – when external storage systems are virtualized their data may tiered or migrated to flash to improve performance while extending the life of the external storage array
- Unified – HUS VM may be upgraded to a unified platform to consolidate both block and file data.
- **Advanced Features of HUS VM** - As an all flash system it may be upgraded for more capabilities for consolidating file data, tiering data, disaster recovery, data mobility and analytics.
- **Flash Capacity** – Supports up to 12 1.6TB flash modules in a 2U flash storage enclosure and up to 8 flash storage enclosures per HUS VM system. This means a total of 96 flash modules are supported with a maximum raw capacity of 154TB in 16 rack units.
- **New configuration option** – Now supports an all flash configuration with Hitachi Accelerated Flash (HAF) storage.
- **Object storage support** - within the HNAS architecture helps to accelerate the object-based meta data

**Hitachi Accelerated Flash (HAF) and flash controller**

Each flash module is powered and managed by the Hitachi embedded flash controller. This is a new multi-core optimized flash controller that is designed to handle up to 128 flash chips and performs inline write avoidance to support faster formatting and extend the life of the supported MLC flash memory. The advantage of a multicore processor architecture in the flash controller is that background functions like garbage collection, data refreshes and ECC checking can be processed without affecting processing of concurrent host I/O’s. This inline write avoidance algorithm can demonstrate write reductions up to 94%. There is an adaptive data refresh feature to enhance data integrity and help optimize write endurance.
The management engine in the controller offers the management features, such as, Block/Page Mapping, Wear Leveling, Performance Manager, and Endurance Manager. Special optimized features are:

- **Adaptive Data Refresh** – A data refresh will read data and copy it to a different area before the data becomes unreadable. The data refresh does internal ECC checks on the whole flash module every 30 days and dynamically optimizes the page refreshed based on the applied error correction. This can sustain performance and flash cell longevity.

- **Contiguous Block Write Avoidance** – This feature offers real time inline write avoidance demonstrating up to 94% capacity savings depending on the data set. This can improve sustained write performance and preserves capacity for background tasks like garbage collection and wear leveling.

- **Formatting** – The optimized LDEV format allows format processing to be more than 4x faster for twice as much capacity than the LDEV format of SSDs.

### Hitachi Accelerator Flash Module

![Hitachi Accelerator Flash Module Diagram](image)

The flash optimized microcode was re-engineering to create PCI “express” host I/O transfer for flash media by significantly reducing instruction processing time. It enables scalability to over 500,000
random I/Os per second with the use of HAF. This feature in the operating system is transparent to other storage features like Hitachi Dynamic Provisioning, Hitachi Dynamic Tiering, in-system replication or remote replication. It can be enabled or disabled non-disruptively.

**Customer Value**

The HUS VM All Flash System is designed for better use of system resources with an existing architecture, which allows for investment protection. This all flash system provides IT administrators good value by supporting advanced enterprise features as well as increased performance to better support SLAs for tier 1 applications and virtualized infrastructures.

HUS VM All Flash Array has good economic value for both CAPEX and OPEX:

- **Solid state density** - Hitachi has doubled their solid state storage density per 2U enclosure by supporting HAF (19.2TB) in the HUS VM vs. SSDs (9.8TB).
- **Power savings** - In addition HAF (18.4 W/TB) storage consumes less than half the power per TB than hard disk drives (46.9 W/TB). This is a 60% savings in power.
- **Price/Performance** – Based on 500K IOPS HAF has a 64% improvement in price performance over the currently supported SSDs.
- **Single point of management** – the HUS VM is managed by the same Hitachi Command Suite platform as all other Hitachi systems eliminating islands of management
- **High Availability** – the HUS VM offers high availability through offering non-disruptive operations both with software updates, hardware component replacement and other data protection features, such as, asynchronous replication.
- **Efficient use of capacity** – there is thin provisioning to prevent wasting capacity by over provisioning and file level deduplication to minimize the amount of actually capacity that needs to get stored saving not only capacity but CAPEX and OPEX as well.

**Performance**

There are many tier 1 applications that are sensitive to timely data access and are impacted by common I/O performance bottlenecks. The HUS VM All Flash System accelerates applications, such as, VDI, OLTP database acceleration, and messaging applications.

It takes 8 HAF modules to achieve over 500K IOPS with microsecond response times compared to 32 SSDs to achieve the same performance but with millisecond response times The HUS VM microcode has been completely re-engineered to handle faster host and target side I/O processing. This allows the HUS VM to perform more than 500K IOPS with a magnitude faster response times making it more than twice as fast as the HDD solution. HDS projects with a future release of the base operating system the HUS VM will scale to 1M IOPS.
Summary

The use of solid state technology in storage can have profound impact in the areas of performance, power, cooling, and floor space. This is a new and evolving area with many different solutions today and confident predictions that more will be developed. The different approaches used for introducing flash solid state technology into storage can have different performance and cost characteristics. Depending on the environment and application, some approaches may be more effective than others. Understanding the different approaches is critical in making an informed decision when selecting and deploying a solid state storage solution. The transition to all solid state technology and eliminating electro-mechanical devices for performance in storage will continue and selection of the solid state storage technology implementation are key choices along the way. HUS VM All Flash Solution meets the critical criteria for a real enterprise storage system for today’s applications and new infrastructure technologies and environments.
About Evaluator Group
Evaluator Group Inc. is dedicated to helping IT professionals and vendors create and implement strategies that make the most of the value of their storage and digital information. Evaluator Group services deliver in-depth, unbiased analysis on storage architectures, infrastructures and management for IT professionals. Since 1997 Evaluator Group has provided services for thousands of end users and vendor professionals through product and market evaluations, competitive analysis and education. www.evaluatorgroup.com Follow us on Twitter @evaluator_group

Copyright 2013 Evaluator Group, Inc. All rights reserved.
No part of this publication may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopying and recording, or stored in a database or retrieval system for any purpose without the express written consent of Evaluator Group Inc. The information contained in this document is subject to change without notice. Evaluator Group assumes no responsibility for errors or omissions. Evaluator Group makes no expressed or implied warranties in this document relating to the use or operation of the products described herein. In no event shall Evaluator Group be liable for any indirect, special, consequential or incidental damages arising out of or associated with any aspect of this publication, even if advised of the possibility of such damages. The Evaluator Series is a trademark of Evaluator Group, Inc. All other trademarks are the property of their respective companies.