HITACHI
Inspire the Next

TECHNICAL PAPER

Deploying Virtual Machines With VAAI and Deduplication in a Hitachi NAS Platform Environment

Tech Note

By Daniel Worden and Jeff Chen

August 2015

Hitachi Data Systems

Feedback

Hitachi Data Systems welcomes your feedback. Please share your thoughts by sending an email message to SolutionLab@hds.com. To assist the routing of this message, use the paper number in the subject and the title of this white paper in the text.

Contents

Deploying Virtual Machines With VAAI and Deduplication in a Hitachi NAS Platform Environment

Tech Note

Hitachi NAS Platform (HNAS), employed for VMware, is positioned and deployed as an enterprise NFS solution that delivers a level of scalable and predictable performance (IOPS and latency), capacity, efficiency, and data protection to meet the needs of VMware cloud environments in addition to other benefits. This paper is meant to address the following areas:

- Provide recommended configurations for certain virtual machine (VM) counts and expected performance (IOPS and latency) using current best practice deployments. During this testing, 500 VMs per cluster were tested.
- VM deployment times were measured while using and not using VMware vSphere Storage APIs Array Integration (VAAI)
  - To use VAAI, VM templates were stored on the VM deployment datastore. This is a VAAI requirement.
  - When not using VAAI, VM templates were stored on a separate VM deployment datastore.
- Testing was measured on before the first dedupe operation was ran and again during dedupe operations.

Table 1 lists the test cases that were verified, and this paper shows the following benefits of HNAS during this testing:

- VMs being deployed using VAAI saw a significant time savings compared to VMs deployed without the use of VAAI.
- There was not a performance impact on the VMs using VAAI.

Table 1. Verified Test Cases

Test Case	Pass/Fail Criteria	Result
Deploying a mixed workload VM environment both using VAAI and not using VAAI	VM deployment times should be less using VAAI than when not using VAAI	Pass
Mixed workload application performance was not impacted when deploying VMs with VAAI compared to when VAAI was used	All VMS should be able to reach and maintain the configured I/O target and there should not be a latency impact.	Pass

The following tests were run for this phase of the project:

- Deployed 250 VMs evenly across two datastores using VAAI and measured the deployment time
- Deployed 250 VMs evenly across two other datastores, not using VAAI and measured the deployment time.
- Ran 500 VM mixed workloads for a six-hour test and measured I/O performance and latency.

This document provides the following:

- Validation of the deployment of 250 VMs without the use of VAAI and the deployment of 250 VMs with VAAI.
- Validation of 500 VMs running on Hitachi NAS while a dedupe operation is in progress.

Note — Testing of this configuration was in a lab environment. Many things affect production environments beyond prediction or duplication in a lab environment. Follow the recommended practice of conducting proof-of-concept testing for acceptable results in a non-production, isolated test environment that otherwise matches your production environment before your production implementation of this solution.

Use Case Overview

During this phase of testing, 500 VMs were used which ran a variety of workload profiles. The workloads were divided into tiles of 50 VMs each. Each workload profile had a light, medium, and heavy component. The VMs in each tile were configured for an average of 25 IOPS. The workload profile within each tile is described in Table 2.

Table 2. Virtual Machine Workload Profile for Each Tile

Workload	Weight	IOPS	VMDK Size	Number of VMs	Total IOPS	Total VMDK Size
Microsoft® SQL Server®	Light	10	60 GB	3	30	180 GB
Web Server	Light	10	50 GB	6	60	300 GB
Exchange	Light	10	60 GB	3	30	180 GB
OLTP	Light	10	60 GB	3	30	180 GB
SQL Server	Medium	25	200 GB	1	25	200 GB
Web Server	Medium	25	80 GB	2	50	160 GB
Exchange	Medium	25	100 GB	1	25	100 GB
OLTP	Medium	25	100 GB	1	25	100 GB
SQL Server	Heavy	70	40 GB	1	70	40 GB
Web Server	Heavy	70	40 GB	2	140	80 GB
Exchange	Heavy	70	40 GB	1	70	40 GB
OLTP	Heavy	70	40 GB	1	70	40 GB
Total		25 per VM		25	625	1600 GB

The total VMDK size includes a 20 GB VMDK for the operating system. The remainder of the VMDK size was used as a VMDK for a data volume. Each VM had the following resources assigned:

- 1 vCPU
- 4 GB of RAM

Table 3 shows the I/O profile for each workload.

Table 3. Workload I/O Profiles

Workload	I/O Size	Percent Random	Percent Read
Microsoft® SQL Server®	64 KB	100%	66%
Web Server	8 KB	75%	95%
Exchange	8 KB	80%	55%
OLTP	8 KB	100%	70%

Tested Components

Table 4 lists the specific hardware components used during testing.

Table 4. Tested Hardware Components

Hardware	Description	Version	Quantity
Hitachi Unified Storage VM	■ Dual controllers ■ 16 × 8 Gb/sec Fibre Channel ports ■ 64 GB cache memory ■ 128 × 600 GB 10k RPM SAS disks, 2.5 inch SFF ■ 8 × 3.2 TB FMD	73-03-06-00/00	1
Hitachi NAS Platform 4100	■ 2 × 10 Gb/sec Cluster ports ■ 4 × 10 Gb/sec Ethernet ports ■ 4 × 8 Gb/sec Fibre Channel ports	12.3.3826.03	1
Hitachi Compute Blade 500 Chassis	■ 8-blade chassis ■ 2 Brocade 5460 Fibre Channel switch modules, each with 6 × 8 Gb/sec uplink ports ■ 2 Brocade VDX 6746 Ethernet switch modules, each with 8 × 10 Gb/sec uplink ports ■ 2 management modules ■ 6 cooling fan modules ■ 4 power supply modules	SVP: A0170-D-8920 5460: FOS 7.0.2C VDX6746: NOS 4.1.2	1
520H B2 server blade	■ Half blade ■ 2 × 12-core Intel Xeon E5-2697 processors, 2.70 GHz ■ 256 GB RAM ■ 16 × 16 GB DIMMs	BMC/EFI: 01-29	4
Brocade 6510	■ SAN switch 48 × 8 Gb Fibre Channel ports	FOS 7.0.1a	2
Brocade VDX 6740	■ Ethernet switch with 40 × 10 Gb/sec ports	NOS 4.1.2	2

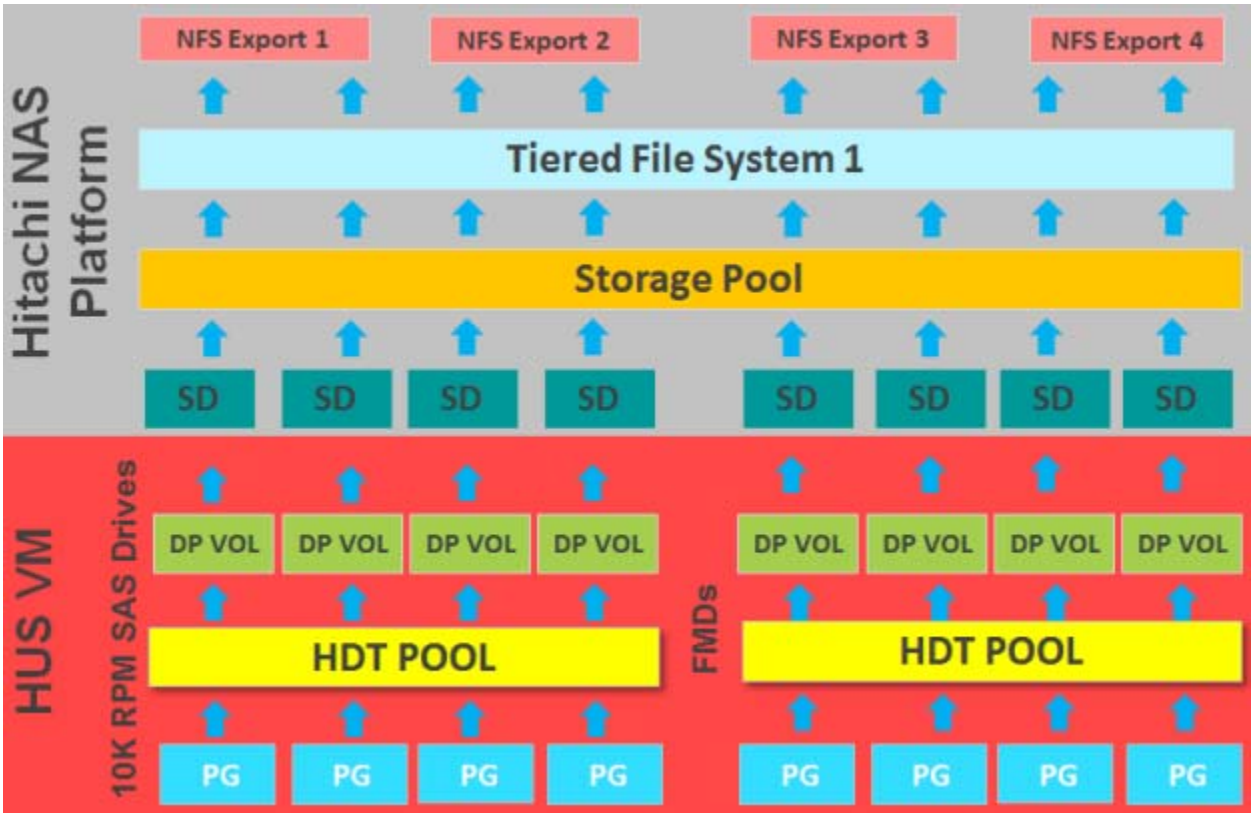
Table 5 lists the specific software components used during testing.

Table 5. Tested Software Components

Software	Version
Hitachi Storage Navigator Modular 2	Microcode Dependent
VMware vCenter server	5.5.0 U2
VMware Virtual Infrastructure Client	5.5.0 U2
VMware ESXi	5.5.0 U2
VDbench	5.04
Microsoft® Windows Server® 2008 R2 (Microsoft SQL Server® VMs OS)	
Windows Server 2012 R2 (Exchange Server VMs OS)	
SUSE Linux Enterprise Server 11 SP2 (OLTP and Web Server VMs OS)	

High Level Test Infrastructure

Figure 1 illustrates the storage configuration for the HNAS NFS storage testing.



HUS VM = Hitachi Unified Storage VM, HDT = Hitachi Dynamic Tiering, PG = Parity Group, HDP Pool = Hitachi Dynamic Provisioning Pool, DPVOL = Dynamic Provisioning Pool

Figure 1

Hitachi Unified Storage VM (HUS VM) was configured with 10K SAS drives and FMDs grouped and presented as the following:

- 26 RAID-6 (6D+2P) 10K parity groups (PGs)
- 1 RAID-5 (7D+1P) FMD PG
- 1 HDT pool was created from the 26 × RAID-6 (6D+2P) 10K PGs and 1 RAID-5 (7D+1P) FMD PGs
- 16 × 1 TB thin provisioned DP-Vols were created from the single HDT Pool. The Tiering Policy was set to **Level 1**, so blocks written to these DP-Vols were always written to Tier 1, which is the FMDs.
- 32 × 8 TB thin provisioned DP-Vols were created from the single HDT Pool. The Tiering Policy was set to **All**, so blocks written to these DP-Vols were dynamically placed.

Note — Due to the large capacity of the FMD dynamic tiering in comparison to the working data set used during testing, most of the working blocks were promoted from the SAS tier 2 to the FMD tier 1. The HDT cycle time was set to 30 minutes. This means that every 30 minutes, blocks that were seen to be more heavily accessed were promoted to tier 1.

The following configuration was used for HNAS:

- 32 system drives (SDs) were created from the 32 8 TB DP-Vols that were provisioned from the 10K SAS drives
- 16 SDs were created from the 16 HDP Vols from the FMDs
- One Storage Pool was created from the 48 System Drives.
- Two filesystems with Tiered Filesystems (TFS) were created from the Storage Pool
- 2 NFS Exports were created from each filesystem as mount points for the ESXi datastore

Thin provisioned DP-Vols as HNAS System Drives were introduced in version 12.1.3613.10. Since this version, thin provisioned DP-Vols are recommended for the following reasons:

- It is quicker and easier to expand an HNAS file system
- Data rebalancing is automatically performed by the HDP pool more efficiently
- Prior to the support of thin provisioned SDs, new SDs had to be presented to the HNAS, and the Storage Pool had to be expanded by a stripeset. Then the file system had to be rebalanced.

When using HDP thin provisioned volumes with HNAS, it is a best practice to not over provision more than three times the actual available space.

HNAS Super Flush was configured on all SDs with the setting of 3 wide by 128 Kb (3 × 128). These settings are best practice when using HNAS Super Flush with the HUS VM.

During this testing the HUS VM tier 1, served by FMDs, served the Hitachi NAS tiered file system (TFS) tier 0 metadata as well as VMware VMDK data.

All best practices were followed when configuring HUS VM and HNAS.

For information on Ethernet networking configuration recommendations, please see the reference architecture guide Deploy Hitachi Unified Compute Platform Select for VMware vSphere Using Hitachi NAS Platform With Hitachi Unified Storage VM.

For information on best practices when using Hitachi NAS Platform in a VMware environment please see the best practices guide Hitachi NAS Platform Best Practices Guide for NFS with VMware vSphere.

Test Result

HNAS NFS VAAI VM Deployment

The first test was deployment of 250 VMs without the use of VAAI compared to the deployment of 250 VMs using VAAI.

To use VAAI during deployment, the VMs or templates used as the source for the clones must reside on the same datastore as the VM clones destination.

- 20 tiles (500 VMs) were distributed evenly across four datastores.
- The 20 tiles were provisioned as Thin VMDKs.

Deploying 250 VMs without the use of VAAI, due to the source templates being stored on separate datastores from the VMs deployment destination resulted in the following:

- It took 5.5 hours
- 1.63 TB of disk space were used

Deploying 250 VMs with the use of VAAI, due to the source templates being stored on the same datastore as the VMs deployment destination resulted in the following:

- It took 20 minutes
- 325.31 GB of disk space were used

Deployment time was measured by documenting the time reported by vCenter of the start of the first VMs deployment and the completion of the last VMs deployment. Deployment time did not include power on or VM customization.

Note — The reason for the difference of 1.63 TB for the VMs deployed without VAAI and 325.31 GB for the VMs deployed with VAAI is due to HNAS fast clone technology. VMs deployed without VAAI are full copies. All blocks are copied for each new VM deployed from the template. VMs deployed with the use of VAAI use the template as the base image for all VMs deployed from that template and then create a delta file for all unique blocks. This reduces the creation of redundant identical blocks.

Figure 2 shows the comparison of time to deployment with and without the use of VAAI.

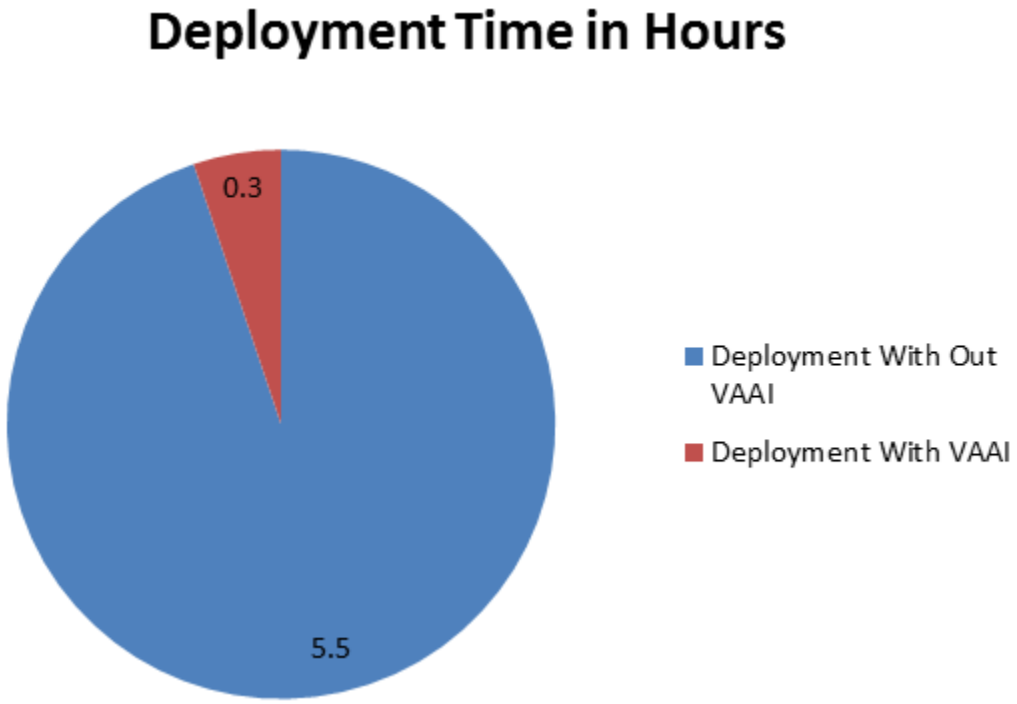


Figure 2

During the Hitachi NAS Platform testing, the following was observed:

- The deployment of 250 VMs without the use of VAAI
  - The average HNAS IOPS was 24,329
  - Deployment took 5.5 hours.
  - The HNAS CPU average percent busy was 7 percent
  - The HNAS FPGA percent busy was 70 percent
- The deployment of 250 VMs with the use of VAAI
  - The average HNAS IOPS was 7327
  - Deployment took 20 minutes.
  - The HNAS CPU average percent busy was 21 percent
  - The HNAS FPGA average percent busy was 42 percent

HNAS NFS Datastore Baseline

The second test was a baseline on the Hitachi NAS platform without a dedupe operation running.

- 20 tiles (500 VMs) were distributed evenly across four datastores.
- The 20 tiles were provisioned as Thin VMDKs. After the data VMDKs were 50% filled, 17.2 TB of disk space was used on the Hitachi NAS Platform before the dedupe operation.
  - 8.62 TB of disk space was used on the filesystem of the VMs deployed without using VAAI
  - 8.58 TB of disk space was used on the filesystem of the VMs deployed using VAAI

Note — The reason for the difference between the disk utilization for the virtual machines deployed with and without VAAI is because the operating system disks were not fully copied for each virtual machine due to VAAI, saving disk space.

The 20 tiles ran for 6 hours to establish a baseline for comparison, as described in Table 6.

Table 6. Second Test Case

Test Case	Description	Result
Mixed workload application performance without a Hitachi NAS Platform dedupe operation running to establish a baseline	Run workload on 20 tiles or 500 VMs for 6 hours	Passed
	Expected result: Virtual machines should be able to reach and maintain the configured I/O target.	

The tiles were distributed evenly across the ESXi host.

The tiles were evenly distributed across the four NFS datastores. The 250 VMs deployed without VAAI were deployed to datastores 1 and 2 which were on the first HNAS filesystem. The 250 VMs deployed using VAAI were deployed to datastores 3 and 4 which were on the second HNAS filesystem.

VMDKs were Thin Provisioned. Thin Provisioned VMDKs are the best practice when using Hitachi NAS Platform with ESXi hosts, and are the default VMDK type when a VMDK is migrated to, or created on, an NFS datastore.

All VMDKs were 50 percent filled and the test environment was limited to 20 percent of the VMDK blocks.

During testing of 20 tiles on the HNAS NFS datastores, 100 percent of the target I/O was achieved.

During the HNAS six-hour baseline testing, the average HNAS IOPS was 13,468.

During baseline testing, the HNAS CPU average percent busy was 10 percent and the FPGA percent busy was 17 percent.

During baseline testing the average application latency was 6 milliseconds.

HNAS NFS Datastore During Dedupe Operation

The third test was on Hitachi NAS Platform with a dedupe operation running.

The 20 tiles ran for 6 hours for comparison to the baseline test, as described in Table 7.

Table 7. Third Test Case

Test Case	Description	Result
Mixed workload application performance while Hitachi NAS Platform dedupe operation is running.	Run workload on 20 tiles or 500 VMs for 6 hours	Passed
	Expected result: Virtual machines should be able to reach and maintain the configured I/O target.	

The tiles were distributed evenly across the ESXi host.

The tiles were evenly distributed across the four NFS datastores. The 250 VMs deployed without VAAI were deployed to datastores 1 and 2 which were on the first HNAS filesystem. The 250 VMs deployed using VAAI were deployed to datastores 3 and 4 which were on the second HNAS filesystem.

VMDKs were Thin Provisioned. Thin Provisioned VMDKs are the best practice when using Hitachi NAS Platform with ESXi hosts, and are the default VMDK type when a VMDK is migrated to, or created on, an NFS datastore.

All VMDKs were 50 percent filled and the test environment was limited to 20 percent of the VMDK blocks.

During testing of 20 tiles on the HNAS NFS datastores, 100 percent of the target I/O was achieved.

During the HNAS dedupe testing, the average HNAS IOPS was 14,006 during the six-hour test.

During baseline testing, the HNAS CPU average percent busy was 17 percent and the FPGA percent busy was 27 percent.

During dedupe testing, the average application latency was 10 milliseconds. Figure 3 shows the comparison of the baseline testing without a dedupe operation and testing with a dedupe operation.

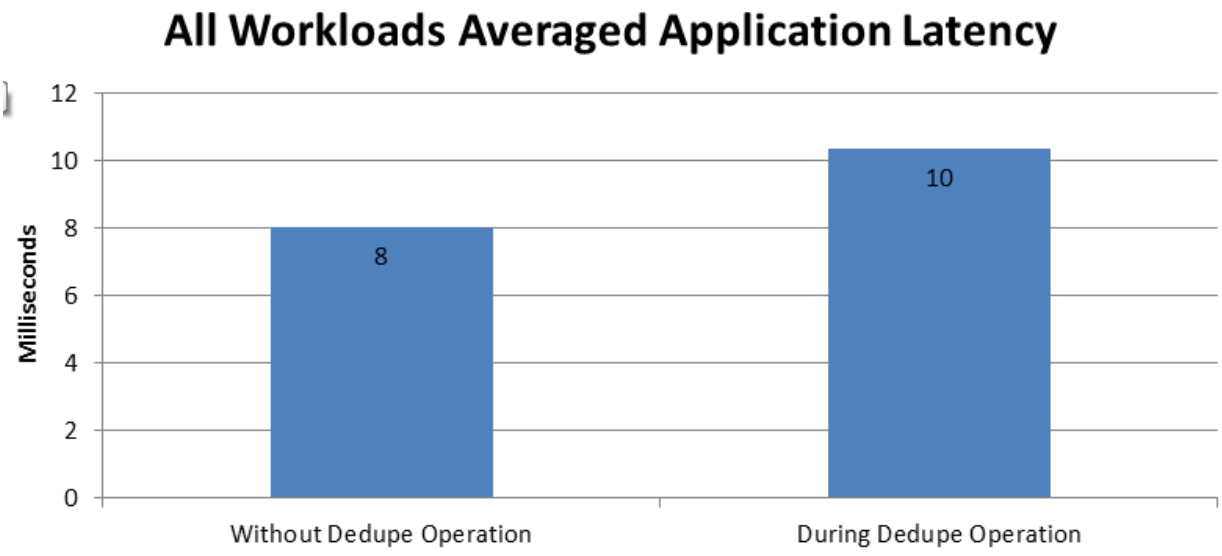


Figure 3

HNAS average CPU utilization during testing with dedupe operation was 17 percent. FPGA utilization was 27 percent. Figure 4 shows the comparison of CPU and FPGA with testing without and with dedupe operation.

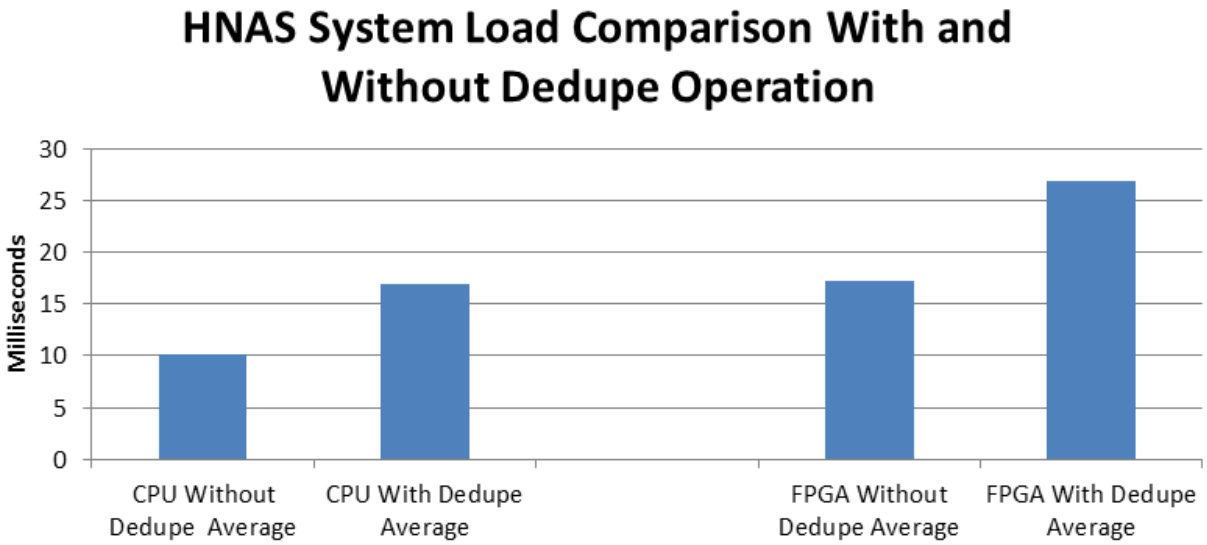


Figure 4

Conclusion

During testing of dedupe with Hitachi NAS Platform, deployment times were cut from 5.5 hours to 20 minutes for 250 VMs. Also, application I/O was consistently reached, but dedupe did impact application latency by 20 percent during this testing.

The impact of dedupe on the workloads may be able to be mitigated by configuring various dedupe and snapshot deletion parameters, but this was not included in this round of testing.

Appendix A

Types of Hitachi NAS Platform Dedupe Operations

There are three types of HNAS dedupe operations. Table 8 lists and describes the three.

Table 8. Types of Dedupe Operations

Type of Dedupe Operation	Description
Full	Runs if a filesystem is formatted without dedupe enabled and then dedupe is later enabled on that filesystem.
Triggered Incremental	Is triggered after a configured amount of changes have occurred on the filesystem. The default is 1 TB.
Daily Incremental	Is triggered at a set time each day. As long as there are at least 20 GB of changes on the filesystem, this dedupe operation will run.

Platform Dedupe Operation Process

During an HNAS dedupe operation there are 4 steps:

- The dedupe operation is initiated
- A block-based snapshot is taken
- The dedupe operation runs until it is completed or aborted
- The snapshot is deleted

During testing there was an increase in NFS datastore and application latency. During the snapshot deletion process the datastore and application latency was more pronounced. The duration of the snapshot deletion impact depends on the length of the dedupe operation and snapshot size.

HNAS Dedupe Troubleshooting

During the day-to-day operation of HNAS with dedupe enabled, it may be convenient or necessary to know the status of dedupe jobs or snapshot creation and deletion. Table 9 describes commands used to monitor and manage HNAS dedupe.

Table 9. Dedupe Monitoring and Management Commands

Command	Command Syntax	Description
fs-dedupe-history	`fs-dedupe-history -a <fsname>`	List any current running dedupe operations and up to the last 5 previously running dedupe operations
dedupe-queue-add	`dedupe-queue-add --full -f <fsname>`	Manually triggers a full dedupe operation
dedupe-queue-add	`dedupe-queue-add --incremental -f <fsname>`	Manually triggers an incremental dedupe operation
snapshot-list	`snapshot-list -a --file-system <fsname>`	Lists current snapshots on the specified filesystem, if there are any

Table 9. Dedupe Monitoring and Management Commands (Continued)

Command	Command Syntax	Description
event-log-show	`event-log-show -o | grep -i dedupe`	When combined with the `grep` command all event log entries containing the dedupe keyword are listed
event-log-show	`cn all event-log-show -o | grep -i dedupe`	When working in a clustered environment, add `cn all` before the `event-log-show` command to list event log entries on all cluster nodes
dedupe-queue-status	`dedupe-queue-status`	Lists all 5 dedupe queues and the status of any jobs in those queues

HNAS Dedupe Customization

There are some HNAS parameters that can be customized to have dedupe operations better fit your environment. Table 10 lists some of these parameters and describes how to use them.

Table 10. Dedupe Configuration Commands

Command	Description
`dedupe-daily-run-scheduler-frequency-in-seconds`	Default value is 24x60x60 seconds, or 24 hours
`dedupe-daily-run-scheduler-poll-interval-in-second`	Set the frequency in which file systems are checked to see if a dedupe should be run. Default is 1 hour
`dedupe-threshold-max-factor`	Set the threshold that an incremental dedupe is run. The trigger is set at 2 TB/<value of the variable>. The default value is 2, so the default trigger is set at 1 TB.
`snapshot-max-unlink-blocks-per-checkpoint`	Default is 100,000

Hitachi Data Systems

Corporate Headquarters
2845 Lafayette Street
Santa Clara, CA 96050-2639 USA
www.HDS.com community.HDS.com

Regional Contact Information
Americas: +1 408 970 1000 or info@hds.com
Europe, Middle East and Africa: +44 (0) 1753 618000 or info.emea@hds.com
Asia Pacific: +852 3189 7900 or hds.marketing.apac@hds.com

AS-409-01 August 2015.