Technical Report

NetApp and VMware View 5,000-Seat Performance Report

Chad Morgenstern, Chris Gebhardt, NetApp
August 2011 | TR-3949

ABSTRACT

In 2010 NetApp®, VMware®, Cisco®, Fujitsu®, and Wyse® published a joint 50,000-seat VDI reference architecture that provided customers an overview on design and architecture. This report provides a detailed performance analysis of a 5,000-seat pod, the building block of the 50,000-seat architecture. A mock scenario is used to demonstrate 2 weeks in the life of a desktop. This was done to illustrate the point that each day, the workload characheristics may change and thus have a differing effect on the architecture. Sizing storage for VDI is not just about steady state but about all workloads.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1 EXECUTIVE SUMMARY

In August of 2010, NetApp and a number of partners published a white paper describing the deployment of a 50,000-seat virtual desktop infrastructure (VDI) environment using NetApp storage, Cisco Unified Computing System™ (Cisco UCS™) and Cisco Nexus®, VMware software, and Fujitsu servers. This initial white paper focused solely on the high-level architectural design and the specifics of the technology that each of the partners brought to the table to allow for this deployment. This initial effort called for using NetApp FAS3170 storage controllers and defined modular units of storage and servers called "pools of desktops" (PODs), based on the hardware and software needed for deploying 5,000 seats into a virtualization infrastructure. The initial white paper defined a POD as follows:

- 60 ESX® 4.1 hosts (Cisco UCS or Fujitsu PRIMERGY)
- 1 FAS3170A high-availability (HA) cluster
- 96 15K RPM Fibre Channel drives
- 2 512GB Flash Cache cards
- 2 VMware vCenter™ Servers
- 3 VMware View™ Connection Servers running PC-over-IP (PCoIP) connection protocol
- 5,000 Microsoft® Windows® 7 virtual desktop virtual machines (VMs)

Shortly after the publication of the initial white paper, NetApp refreshed its midrange storage offerings to include the FAS3270 storage systems, which improved on the capacity and performance of the FAS3170 storage systems used in the initial white paper. As a result, for the tests described in this technical report, we used a FAS3270 storage system instead of the FAS3170 used in the initial white paper because it significantly improves the performance and scalability of the solution. In addition, subsequent early testing showed that being able to effectively support 5,000 VDI desktops required that we add 30 more servers so that adequate memory resources were available during the testing. As a result, we now define a POD as follows:

- 90 ESX 4.1 Servers:
  - Fujitsu PRIMERGY RX200-S5 with 2 Quad Core Nehalem CPUs with hyperthreading
  - 48GB main memory
- 1 FAS3270A HA cluster
- 96 15K RPM Fibre Channel drives
- 2 512GB Flash Cache modules
- 2 VMware vCenter Servers
- 3 VMware View Connection Servers running PC-over-IP (PCoIP) connection protocol
- 5,000 Microsoft Windows 7 virtual desktop VMs

Under this new definition, each NetApp FAS3270 controller supported 45 ESX servers and 2,500 Windows 7 persistent virtual desktops. Because of the large hardware requirements to create a full POD, we chose to limit these subsequent tests to using what we describe as half a POD or 2,500 virtual desktops being served by one of the two FAS3270 storage controllers. Because each FAS3270 storage controller is actually independently serving 2,500 virtual desktops, the performance measured for the 2,500 virtual desktops can simply be doubled to account for the full POD supporting 5,000 virtual desktops.

For these tests we used VMware View 4.5 and VMware vSphere 4.1 to deploy a scenario consisting of 2,500 virtual desktops. In addition, we used the VMware Reference Architecture Workload Code (RAWC) tool to generate a workload typical of what might be found in a VDI environment.

We tested with both the current General Availability (GA) release (Data ONTAP® 8.0.1) and Data ONTAP 8.1. We selected Data ONTAP 8.1 to use because it provides additional performance gains and spindle reduction benefits to the NetApp virtual storage tiering (VST) capabilities. VST allows customers to

benefit from NetApp storage efficiency and at the same time significantly increase I/O performance. VST is natively built into the Data ONTAP operating system and works by leveraging block-sharing technologies such as NetApp primary storage deduplication and file/volume FlexClone to reduce the amount of cache required and eliminate duplicate disk reads. Only one instance of any duplicate block is read into cache, thus requiring less cache than traditional storage solutions. Because VMware View implementations can see as great as 99% initial space savings using NetApp space-efficient cloning technologies, this translates into higher cache deduplication and high cache hit rates. VST is especially effective in addressing the simultaneous system boot or "boot storm" and login of hundreds to thousands of virtual desktop systems that can overload a traditional legacy storage system. With Data ONTAP in 8.1, deduplicated data blocks in main memory can be shared more efficiently than in previous releases. This allows for a larger working set to fit in the storage controller's main memory, resulting in faster access times and reduced CPU.

During our testing, we confirmed that there is much more to a VDI workload than simply what goes on during steady state when users are accessing their desktops during the course of normal business. Although it is important to understand the characteristics of this phase, there are also situations in which reboots, logins, profile creation, and manipulation of large numbers of virtual desktops place heavy burdens on the storage supporting the overall VDI environment. Failure to understand and plan for these workloads can have a serious negative impact on end-user experience and project success.

In the remainder of this report, we examine a number of different common VDI scenarios from the perspective of the fictitious Acme Corporation. Acme has deployed a 2,500-seat VDI environment using NetApp storage and wants to understand the different types of workloads that might be encountered during a typical workweek in which users typically log in in larger numbers, start and stop applications, and conduct routine tasks using the applications that have been provided to them. Additionally, there are times when routine maintenance requires that all of the virtual desktops be powered off, which then requires subsequent power-on and booting of large numbers of VMs simultaneously.

These scenarios and their outcomes as measured by user experience are defined at a high level in Table 1 through Table 6. The user experiences throughout this report are measured by Liquidware Labs Stratusphere UX, which uses a weighting algorithm to determine "goodness" of the user experience. The power-on and creation scenarios are reported in the elapsed time. Note that the application mixture selected is one that VMware advised would generate 12 input/output operations per second (IOPS)/desktop, the expected workload of "knowledge users."

Table 1) The creation scenario.

Test	Primary Measure of Success	Creation Time
Create the 2,500 VMs and measure how long it takes.	Length of time the creation takes to be ready for first power-on	3 hours from beginning of the clone process to 2,500 VMs ready for power-on

Table 2) The initial login scenario.

Test	Primary Measure of Success	User Experience
2,500 users log in for the first time over a half-hour period and begin working. This login triggers 2,500 profile creations.	User experience as measured by Liquidware Labs UX	• Data ONTAP 8.0.1: 62% of users had a good user experience, 38% fair. • Data ONTAP 8.1: 97% of users had a good user experience, 3% fair.

Table 3) The "Tuesday morning" or typical login scenario.

Test	Primary Measure of Success	User Experience
2,500 users log into VMs previously accessed and not rebooted since. At login, the users begin working.	User experience as measured by Liquidware Labs UX	• Data ONTAP 8.0.1: 100% of the users had a good user experience. • Data ONTAP 8.1: 100% of the users had a good user experience.

Table 4) The power-on scenario.

Test	Primary Measure of Success	Power-On Time
VMware vCenter controls the mass power-up of 2,500 virtual desktops. Neither profiles nor application DLLs must be loaded from disk.	Total time to power-on	• Data ONTAP 8.0.1: 36 minutes • Data ONTAP 8.1: 21 minutes

Table 5) The "Monday morning" or profile load login scenario.

Test	Primary Measure of Success	User Experience
2,500 users log into VMs previously accessed but since rebooted. At login, the users begin working. Profiles and application DLLs must be loaded from disk.	User experience as measured by Liquidware Labs UX	• Data ONTAP 8.0.1: 100% of the users had a good user experience. • Data ONTAP 8.1: 100% of the users had a good user experience.

Table 6) The steady-state scenario.

Test	Primary Measure of Success	User Experience
2,500 users have completed their logins and have opened and closed all applications at least once but continue their day's work.	User experience as measured by Liquidware Labs UX	• Data ONTAP 8.0.1: 100% of the users had a good user experience. • Data ONTAP 8.1: 100% of the users had a good user experience.

At a high level, through our testing we determined that each of the scenarios generated a workload unique to itself. The power-on workload was characterized as being read heavy with a high IOPS count. The "Monday morning" scenario as well as the initial login scenario generated workloads fairly balanced between reads and writes, although the intensity of the initial login workload was the greater of the two. The "Tuesday morning" login scenario as well as the steady-state workload scenario generated mostly write operations. The workload generated by these last two scenarios was far less than the other scenario; of the two, the login generated more than the steady state.

The characterization of each of these scenarios has an impact on end users and their overall experience. Power-on of VMs is not normally associated with end-user experience, except when events such as HA, disaster recovery (DR), or any unplanned maintenance occur. In this case, booting quickly is extremely important to the end users. For login, regardless of the day of the week, no user wants to wait 15 minutes until their desktop is ready for use. And while users are interacting with their desktops, they want performance as good as or better than that provided by a physical desktop. Failure to characterize these

different workloads and their impact on the infrastructure and the end user will ultimately affect the success of the project.

In addition, we found that these different categories of workloads exhibited characteristics that were sometimes vastly different from the anecdotal classification of a VDI workload as being small-block random reads. For example, we found the following:

- The disk writes are primarily small, in general 4kB, but may be much larger; 1MB write operations have been recorded during large file-save operations.
- Read operations ranged in size from sub-4kB to 1024kB and were fairly balanced in quantity between random and sequential.

In other words, although virtual desktop workloads do generate small random workloads, they also generate large random operations as well as small and large sequential operations. Our tests show that the industry-standard definition of a VDI workload (meaning just looking at IOPS) is much too simplistic and that correctly sizing storage for a VDI environment requires looking deeper.

The graphs in Figure 1 and Figure 2 demonstrate the differences in workload between the scenarios described as they might play out over a typical day or week in the life of Acme Corporation's VDI environment as displayed in the format "a day in the life of." In the original 50,000-seat VDI white paper, the environment was architected using state-of-the-art technologies. Although we did test the original configuration, many revisions of software and hardware have become available, so the newer versions were tested as well. Although the results of the original test validated the architecture, all results detailed in this paper were achieved with a FAS3270A, 15K drives, and Data ONTAP 8.0.1 and 8.1. For the purposes of this paper, this architecture represents our "golden" configuration: "golden" because this particular configuration produced the best user experiences, the highest throughput, and the lowest latencies, and it did so with the lowest cost/IO operation. These graphs showing throughput and IOPS are from the perspective of the storage controller.

Observe the clear delineation between the workload phases as measured in terms of both operations per second and throughput. Notice that the power-on scenario is dominated primarily by a relatively large number of read operations that drive the throughput past peaks of 700MB per second. In addition, the initial login scenario is fairly balanced in operations between read/write operations, with the read operations driving peak throughput of over 400MB per second. Finally, after users have powered on their VMs and logged into their desktops, the steady-state workload shows a significant drop in both IOPS and throughput compared to the power-on and login scenarios. Notice further that, as shown in Figure 1 and Figure 2, once steady state is achieved, the workload shifts toward an 80:20 distribution of write-to-read operations but toward a fairly balanced distribution in terms of throughput as more of the work associated with the initial read activity (for example, loading application libraries and opening/reading files) becomes cached in the VM guest operating system (OS).

Figure 1) Read and write operations per second.

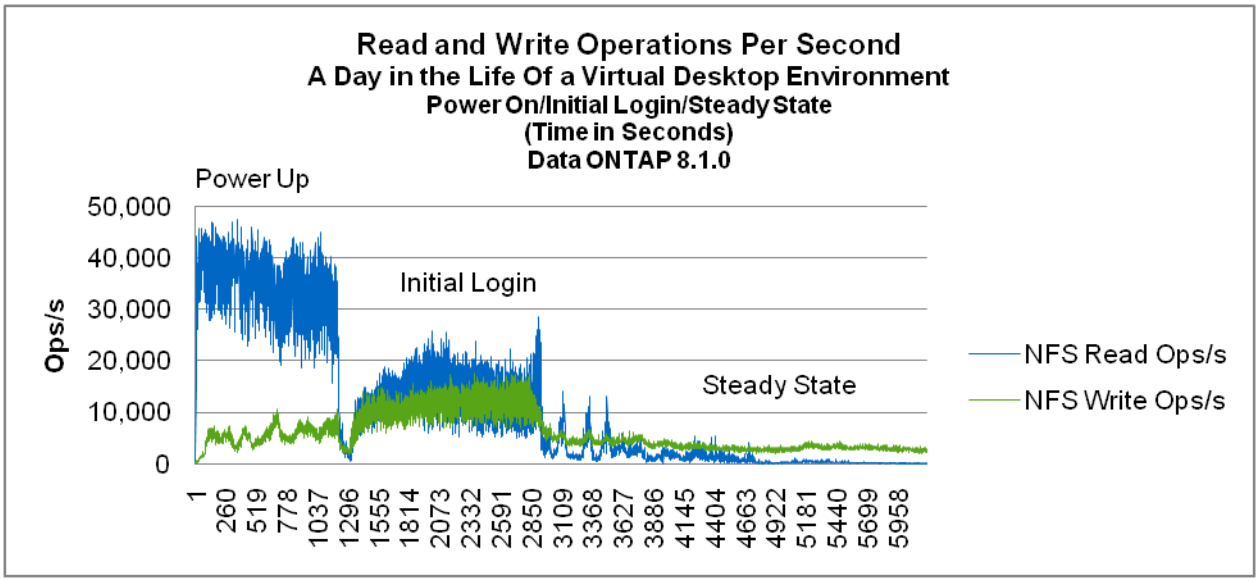
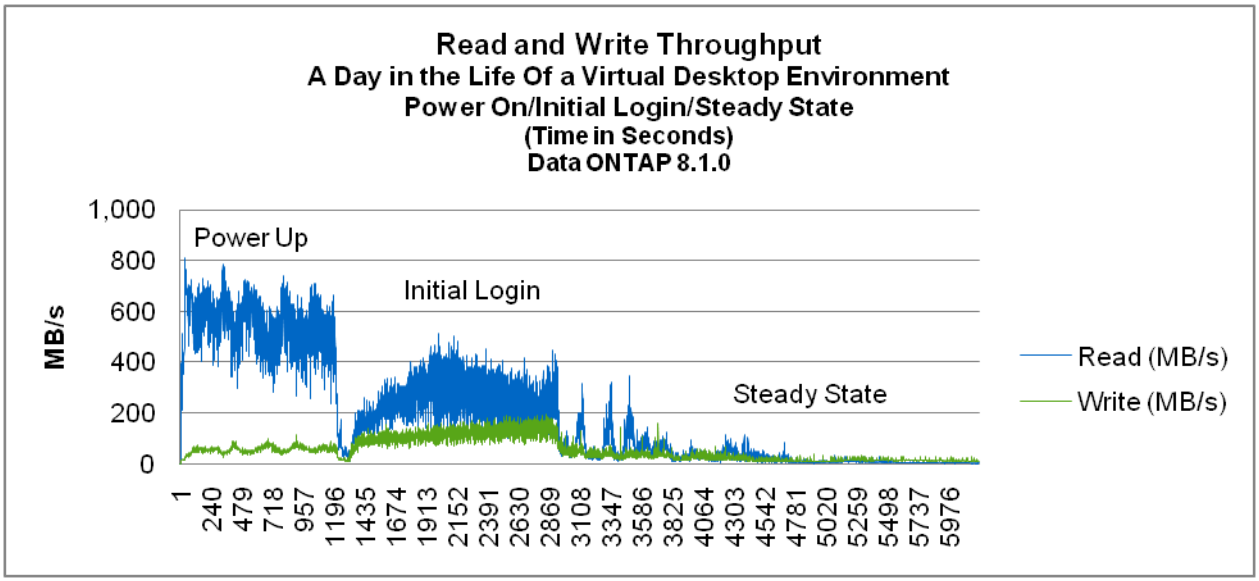
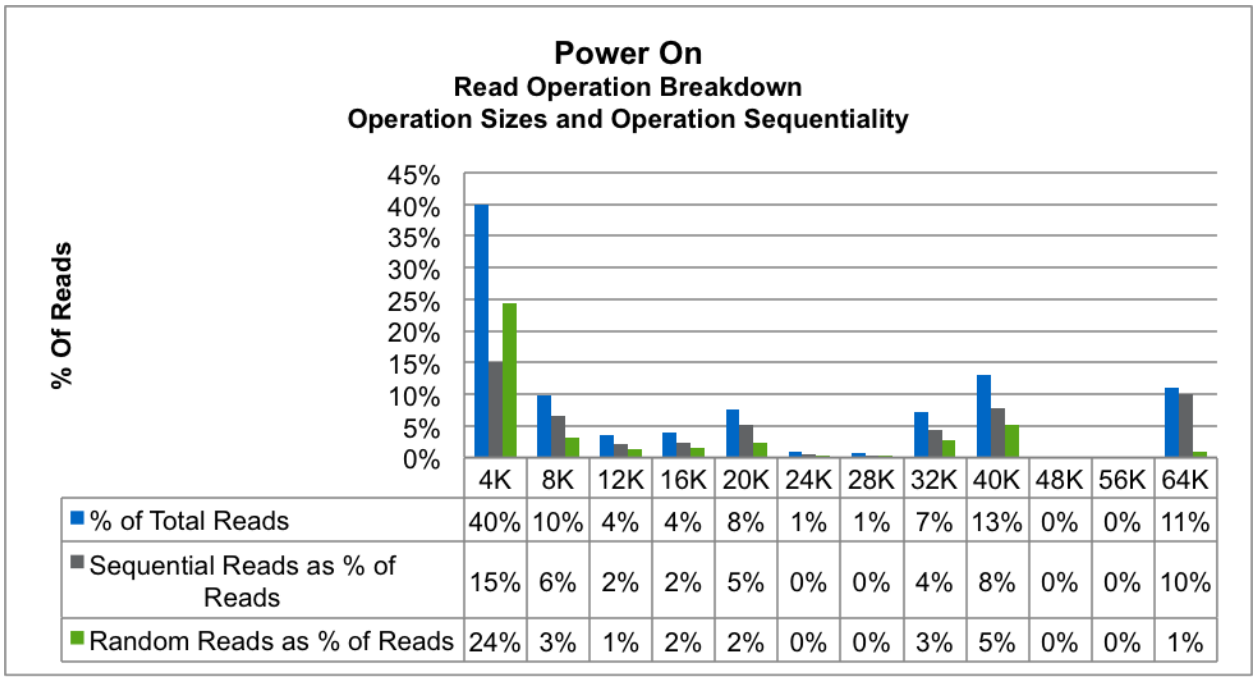


Figure 2) Read and write throughput.



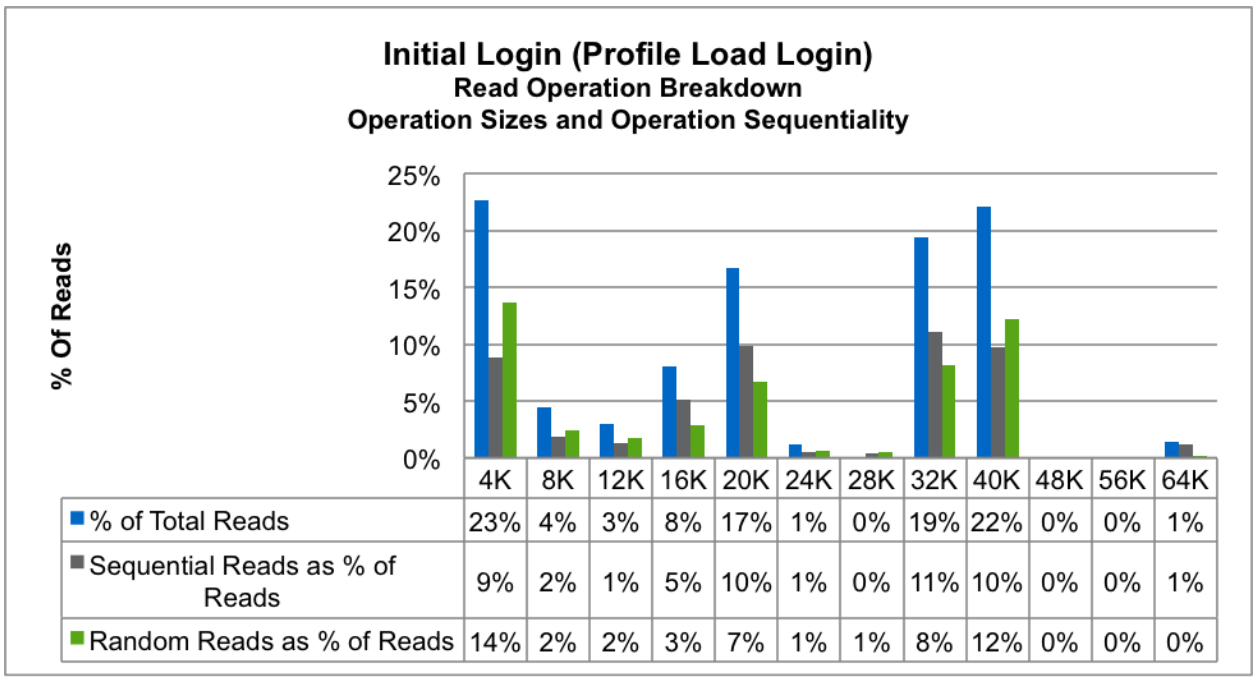
The graphs in Figure 3 through Figure 5 call out the operations sizes as reported by the storage controller. Notice that as the workload distribution and its quantity vary across the workload scenarios, the operations sizes and sequential mixtures vary as well. The constant across this set of workloads is that the percent of random/sequential read operations remains fairly balanced.

Figure 3) Read operation breakdown for power-on.



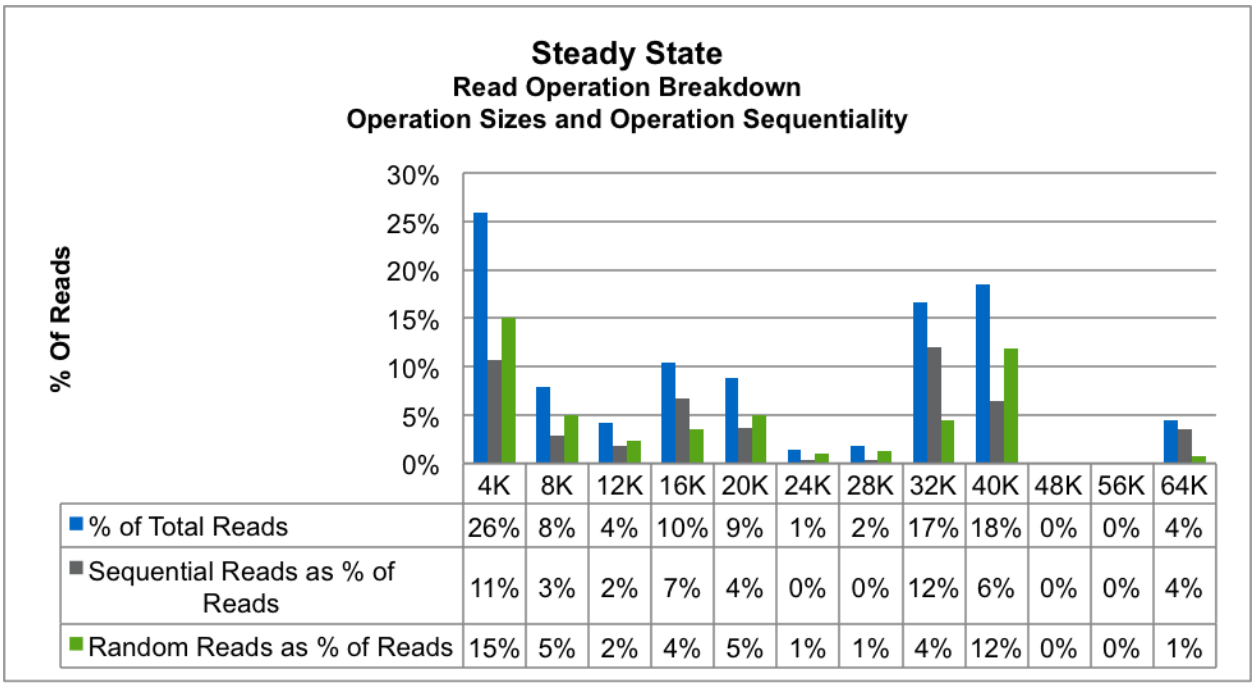
	4K	8K	12K	16K	20K	24K	28K	32K	40K	48K	56K	64K
% of Total Reads	40%	10%	4%	4%	8%	1%	1%	7%	13%	0%	0%	11%
Sequential Reads as % of Reads	15%	6%	2%	2%	5%	0%	0%	4%	8%	0%	0%	10%
Random Reads as % of Reads	24%	3%	1%	2%	2%	0%	0%	3%	5%	0%	0%	1%

Figure 4) Read operation breakdown for initial login.



	4K	8K	12K	16K	20K	24K	28K	32K	40K	48K	56K	64K
% of Total Reads	23%	4%	3%	8%	17%	1%	0%	19%	22%	0%	0%	1%
Sequential Reads as % of Reads	9%	2%	1%	5%	10%	1%	0%	11%	10%	0%	0%	1%
Random Reads as % of Reads	14%	2%	2%	3%	7%	1%	1%	8%	12%	0%	0%	0%

Figure 5) Read operation breakdown for steady state.



	4K	8K	12K	16K	20K	24K	28K	32K	40K	48K	56K	64K
% of Total Reads	26%	8%	4%	10%	9%	1%	2%	17%	18%	0%	0%	4%
Sequential Reads as % of Reads	11%	3%	2%	7%	4%	0%	0%	12%	6%	0%	0%	4%
Random Reads as % of Reads	15%	5%	2%	4%	5%	1%	1%	4%	12%	0%	0%	1%

With this goal in mind, these are the primary efforts of this paper:

- To document the testing methodology used
- To demonstrate the capabilities of NetApp technologies by comparing and contrasting various configurations
- To inform the reader of the characteristics of virtual desktop workloads
- Ultimately, to validate the performance documented in the 50,000-seat white paper

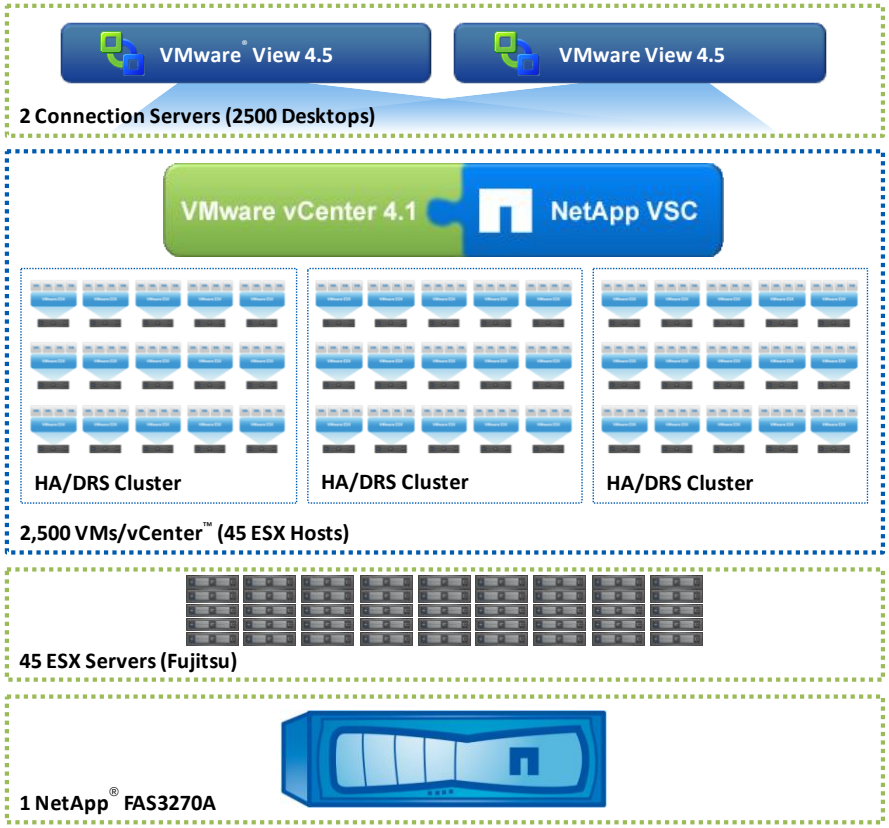
Combined, these efforts are intended to enable the reader to repeat our tests, to choose the best technology for their environment, to size correctly for their workload, and to better understand in general what VDI is.

In the end, our testing showed that delivering great performance in a VDI environment requires more than just the capability to support the steady-state workload (generally thought of to be 12 IOPS per virtual desktop). We found a number of other areas that must also be considered in order to create an overall good user experience. The remainder of this report provides the details associated with the test scenarios, including specifics of the various workloads along with the individual test results, including the user experience calculations.

2 ENVIRONMENT

NetApp conducted its testing in half-POD units of 2,500 virtual desktops spread across 45 ESX 4.1 servers and a single controller of a NetApp HA pair. One VMware vCenter VM managed the entire POD.

Figure 6) Half-POD architecture.



The virtual desktops were registered as dedicated desktops with a VMware View 4.5 connection broker pool made up of two connection brokers and 10 pools. Each pool used the default settings, including the PCoIP connection protocol.

To create the 2,500 VMs for these tests, we created a gold master image VM using 64-bit Windows 7 as the guest OS and applied a series of optimizations as defined in the VMware View Optimization Guide for Windows 7 as well the optimizations recommended in the VMware View Administrator's Guide. We then used the NetApp Rapid Clone Utility (RCU) to clone the 2,500 VMs. vCenter managed the VM power-on operations; logins and the virtual desktop workload were initiated by the VMware desktop RAWC.

NetApp FAS3270s were used for all testing, using variously solid-state drive (SSD), Serial Advanced Technology Attachment (SATA), and 15K FC drives. All tests were conducted using Data ONTAP 8.0.1 and 8.1.

3 TESTS, METHODOLOGY, AND TOOLS

The tests conducted for this report represent some of the most common workloads that a client might expect to experience.

3.1 SCENARIOS

The scenarios for the tests follow a typical two-week calendar at Acme Corporation.

INITIAL LOGIN SCENARIO

In this scenario, 2,500 users logged in over a half-hour period, doing so for the first time, thus testing the load on the storage controller as well measuring the user experience of this worst-case login event. (This is the worst case because this login scenario generates the most load of any scenario tested.)

These logins were accompanied by a profile creation in which the default 1.5MB profile was copied to the user's directory and a Windows Mail subdirectory was created containing 22MB of files.

Following login, the user of each virtual desktop began work: configuring Outlook, sending mail, creating files in Word and Excel, and reviewing documents in PowerPoint and Acrobat.

TYPICAL ("TUESDAY MORNING") LOGIN SCENARIO

As in the initial login scenario, 2,500 users logged in over a half-hour period. This scenario mimics a login scenario in which users log into virtual desktops that have been logged into previously and that have not been power-cycled since login so that they retain both application libraries and profile in memory. Because the profile and application libraries were already in memory, little storage input/output (I/O) was anticipated.

The users in this test ran the work as the "initial login" test with the exception that because Outlook had been configured previously, it was not configured again at this point.

Again, the purpose of this test was to measure the load on the storage controller as well as to measure the user experience.

POWER-ON SCENARIO

All 2,500 virtual desktops were selected within vCenter, power-on was selected, and the results were monitored and timed.

Rather than test for user experience, this test measured power-on time. As in the other scenarios, the load placed on the storage controller was measured.

PROFILE LOAD ("MONDAY MORNING") LOGIN SCENARIO

As in the other login scenarios, 2,500 users were logged in over a half-hour period. This scenario mimics the login scenario in which users log into persistent virtual desktops (as did the others) to which they had previously been logged in but that had been power cycled since, thus clearing each machine's memory. With memory cleared, the logins generated storage I/O by loading the user's profile from disk.

The users in this test ran the same work as in the "Tuesday morning" login scenario. This work generated I/O not seen in the previous scenario, however, because each application had to be opened without the benefit of its libraries having been previously loaded into memory.

As with the other login scenarios tested, the purpose of this test was to measure the load on the storage controller as well as the user experience.

STEADY-STATE SCENARIO

Steady state is defined as the state of the environment after all users have completed login and finished opening their start-of-day applications. In this scenario, each user performed the same work as that done post-login in both the "Monday morning" and the "Tuesday morning" scenarios.

The goal here, as in the login scenarios, was to measure the user experience and the storage controller load.

We performed this testing in response to observations that environments should be measured for "steady state" rather than for login or power-on scenarios. Therefore, we ran this test to determine just what that workload is.

CREATION SCENARIO

Before any test could be performed, the virtual desktops had to exist. To this end, NetApp used our own RCU to clone the desktops. Although this was not an actual test, we did track how long it took to create the virtual desktops from start to ready for power-on.

In order to relate the test results to a realistic user experience, we mapped the test cases into what we are calling "a week in the life of an Acme VDI user." For example, at the first of the week, large numbers of users might simultaneously log into their VMs, preparing for the week's work. This work might include starting a variety of applications and performing different levels of work. For the sake of clarity, all tests described in this paper are referenced to a corresponding date on the calendar, as if the administrator had scheduled the system for the events on these dates. Each test refers back to the calendar introduced in Table 7.

Table 7) Acme Corporation calendar.

Sun	Mon	Tue	Wed	Thurs	Fri	Sat
29 Deploy 2,500 desktops	30 8 a.m. 2,500 logins + Profile load (30 min) Typical workday	31 8 a.m. 2,500 logins (30 min) Typical workday	1	2	3	4
5	6 1 a.m. Network maintenance 2 a.m. Network outage 7 a.m. Reboot all VMs 8 a.m. 2,500 logins (30 min) (Post-power-on)	7 8 a.m. 2,500 logins (30 min) Typical workday	8	9	10	11

All tests were conducted as discretely as possible. Our testing showed that a VM that is powered on generates small amounts of disk I/O even when no additional workload is happening. We found this workload to be predominantly write oriented and generated primarily from System and, to a lesser extent, from the `svchost.exe` and `services.exe` processes. For purposes of this paper, we have termed this the "dripping water" effect, in the sense that each VM generates a small, yet constant workload that when generated from large numbers of VMs can be significant enough to affect overall performance. This is discussed further in section 5.7, "Observations and Lessons Learned."

All tests were conducted using SSD, SATA, and 15K RPM Fibre Channel drives on Data ONTAP 8.0.1. Tests conducted on Data ONTAP 8.1 were conducted with 15K RPM Fibre Channel drives alone because of the cost/IO operation (for the reasons mentioned in the executive summary). For simplicity's sake, the results of the tests performed with SATA and SSD are included in the appendix rather than the body of this report.

In order to return the environment to a baseline state, all virtual desktops were shut down between tests, and the aggregate was rolled back to a pretest Snapshot™ copy and/or the storage controller's memory was flushed.

The details of how we conducted each of these test cases are provided in the following sections.

3.2 CREATION TEST

This test had only one goal, to record the amount of time it takes to create 2,500 virtual desktops using the NetApp Virtual Storage Console 2.1 and Provisioning and Cloning Capability. This demonstrates how customers can easily and rapidly deploy or redeploy thousands of virtual desktops in relatively short periods of time. For example, Acme Corporation buys the Widget Corporation and wants to easily consolidate the new employees. Deploying in a rapid manner allows the company to save time onboarding the new employees. In the second case, customers patch the master template and decide to redeploy the infrastructure.

3.3 POWER-ON TESTS

Primarily, the goal of the power-on tests was to determine how long it would take to bring the environment back up after any event, such as an outage, maintenance, patching, or any other scenario that might require a rapid power-up. The power-on tests were deemed complete when VMware Tools had checked in on all virtual desktops and the Network File System (NFS) operations and CPU utilization on the storage controller had dropped to a low steady state.

The secondary objectives of these tests were as follows:

- To capture the workload profiles in terms of read/write and random/sequential mixture as well as their respective operation sizes
- To evaluate how the storage controllers behave when large numbers of VMs are powering on simultaneously in terms of resource utilization and response time
- To compare the total time taken to completely power-on all virtual desktops

While controlling power-on operations, vCenter throttles power-on to 128 machines at a time. As individual virtual desktops complete the power-on process, they are replaced in the service center with additional entities in an attempt to keep the service center full.

Note: Keep this 128 limit in mind as the divisor against which ops/sec and mB/sec are measured.

All 2,500 virtual desktops were selected within vCenter, with power-on selected and the results monitored and timed.

The conducting of the power-on tests was measured with the aid of vCenter logs, packet traces, as well as statistics collected from the storage controllers and encapsulated in perfstat.

3.4 LOGIN-WITH-WORKLOAD TESTS

In addition to isolating the characteristics of the power-on scenario previously described, we also looked at a number of login scenarios in which logins were followed immediately by users beginning their normal work functions. These functions include opening/loading a variety of applications, including Microsoft Office applications, Microsoft Internet Explorer, Outlook, and Adobe Reader, and writing a variety of documents using these applications.

The primary goal of these tests was to measure the users' experience by Liquidware Labs UX. The secondary goal was to understand the workload profiles of the different scenarios; to capture the workload profiles in terms of read/write and random/sequential mixture as well as their respective operation sizes, and to evaluate how the storage controllers behave during each scenario in terms of resource utilization and response time.

User login and workload scenarios were selected to represent typical "week in the life of" activities that might be encountered in a real-world environment. The login and workload scenarios were as follows:

SCENARIO 1: INITIAL LOGIN AND WORK

2,500 users logged in for the first time on Monday morning between 8 and 8:30 a.m. The users began work following login. This workload represents what might occur during initial login following a virtual desktop refresh or an initial deployment.

What makes this workload stand out is that the each login required a profile creation before completion. Profile creation involved at least two steps:

1. Copying the 1.5MB default profile from each VM's C: drive to the user's home directory
2. Creating a Windows Mail directory inside the new user's profile and populating it with approximately 22MB Windows Mail files

Besides the profile creation, this workload is unique among those tested in that each user configured Outlook as one of the application workloads.

SCENARIO 2: TUESDAY MORNING LOGIN AND WORK

2,500 users logged in between 8 and 8:30 a.m. on Tuesday morning or perhaps a later day in the week. The users began work following login. In this scenario, the users had logged in and opened their applications previously. Because the profile and application libraries were loaded into memory previously, this scenario generates much less I/O than a first login or a login following virtual desktop login.

SCENARIO 3: MONDAY MORNING LOGIN AND WORK

2,500 users logged in on Monday morning between 8 and 8:30 a.m. following the completion of a reboot event. At the onset of this login scenario, each user's profile already existed on disk within their assigned VM but not in memory. Therefore, this login scenario required storage I/O to load profiles from memory. The users began work following login.

This scenario's workload was further increased because each application needed to load its own libraries into memory, a scenario involving I/O not seen in application interaction on the following days of the week or hours in the day.

This test resulted in lower consumed bandwidth and lower I/O than the initial login scenario, but more than the "Tuesday morning" login scenario.

All 2,500 VMs were shut down and restarted before beginning this test. This test shows what happens if users have to log into an environment that has been power cycled before the start of the day. This scenario might occur following any of the scenarios documented earlier in the "Power-On Scenario" section. Before beginning this test, we confirmed that system load had returned to normal, inspected vCenter logs to confirm that all VMs had been rebooted, and consulted vCenter to make sure the VMware tools had logged in all VMs. In other words, we took care to confirm that the power-on process was fully completed before beginning this test.

SCENARIO 4: USERS ALREADY LOGGED IN, STEADY-STATE WORKLOAD

Each of the 2,500 users had already completed the login process and had already turned on all of their applications for the day. This test measured the workload performed after the morning rush was over. Although at this point users were still opening and closing applications, reading/writing files, sending e-mails, and searching the Web, the storage workload was at its lowest because much of what was required was already stored in memory from earlier. The workload characteristics of this test closely resemble those of the "Tuesday morning" login scenario. Because of the similarities, the "Tuesday morning" scenario and the steady-state results are reported together in the following sections.

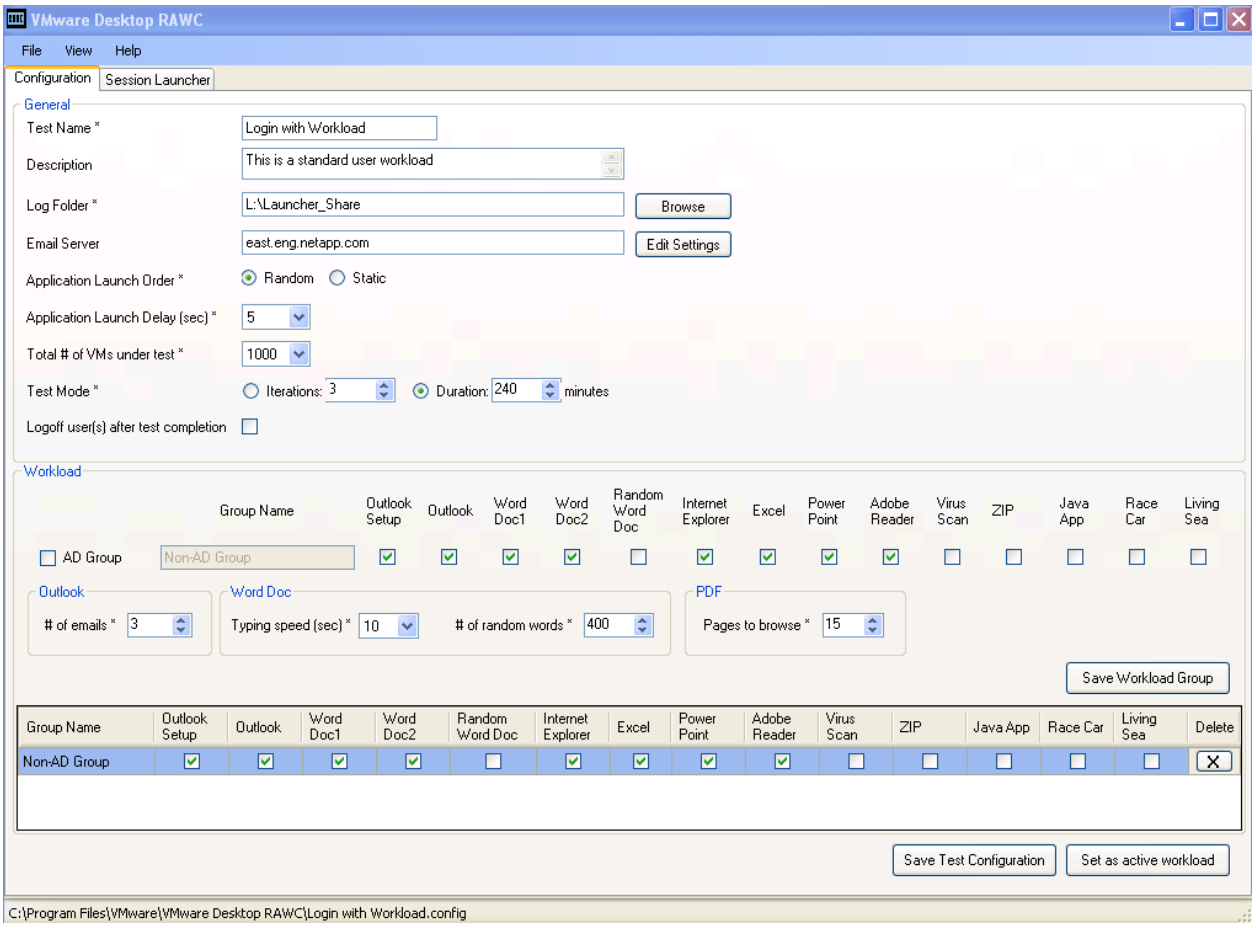
3.5 TOOLS

RAWC generated all workloads in the environment, as indicated by the screenshot in Figure 7. The selected applications were used by recommendation of VMware. VMware informed us that this application mixture would generate a “knowledge user” workload of 12 IOPS/desktop. The following information is also significant:

- The version of Microsoft Office was 2007.
- Microsoft Exchange was version 2010.
- The virtual desktops were set to run Outlook in cached mode.

All logins with workload scenarios were conducted in the same manner. The users were logged in over PCoIP through RAWC. On average, 5 logins were initiated every 3 seconds, and all logins were completed within 28 minutes. To achieve this, 25 RAWC launchers were used. The launchers were each configured to log into one virtual desktop every 15 seconds. The launchers were logged in themselves every 3 seconds serially. After 75 seconds, all launchers were setting up new PCoIP sessions concurrently every 15 seconds. In other words, our testing used the default login rates as defined in the RAWC Administration Guide (available to VMware partners). The order in which the applications were run was random on each virtual desktop.

Figure 7) RAWC screen showing login with workload.



4 END-USER EXPERIENCE

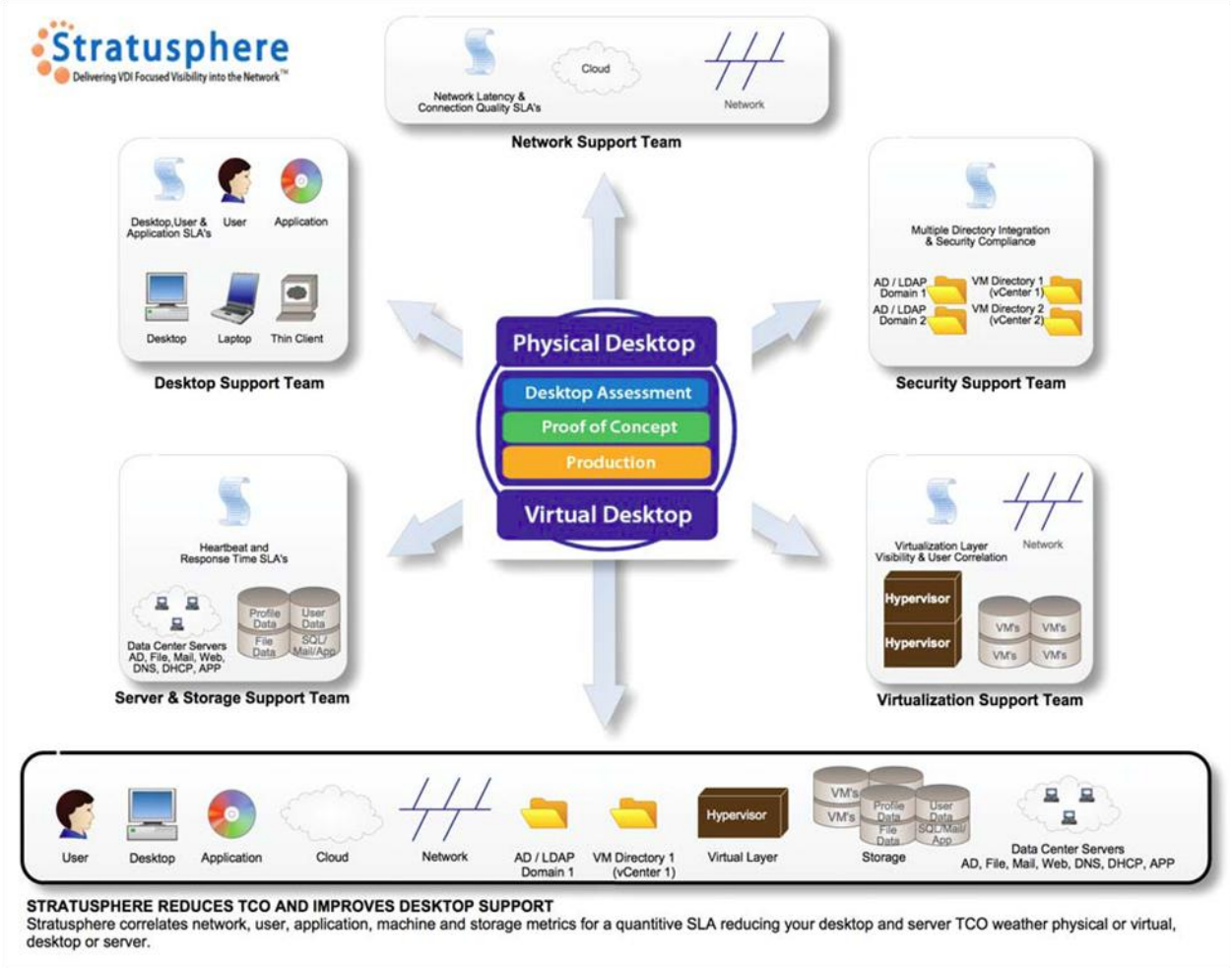
We used Liquidware Labs Stratusphere UX to report the user experiences in all login tests.

This application suite provides an accurate and methodical way to measure the user experience of the desktop and proof-of-concept (POC) environment. The application makes it possible to identify:

- Which VMs, users, and applications are causing I/O storms
- Slow network latency, network application, and protocol performance
- Login delays, application launch times, nonresponsive applications, and disk and CPU queues
- Traffic from endpoints, desktop, and servers in a multitenancy environment with named user and application correlation
- VM, user, and application workload: CPU/memory/disk/graphics/network, including disk IOPS and storage requirements

Figure 8 shows how the features of Stratusphere improve desktop support.

Figure 8) Stratusphere (graphic provided by Liquidware Labs).

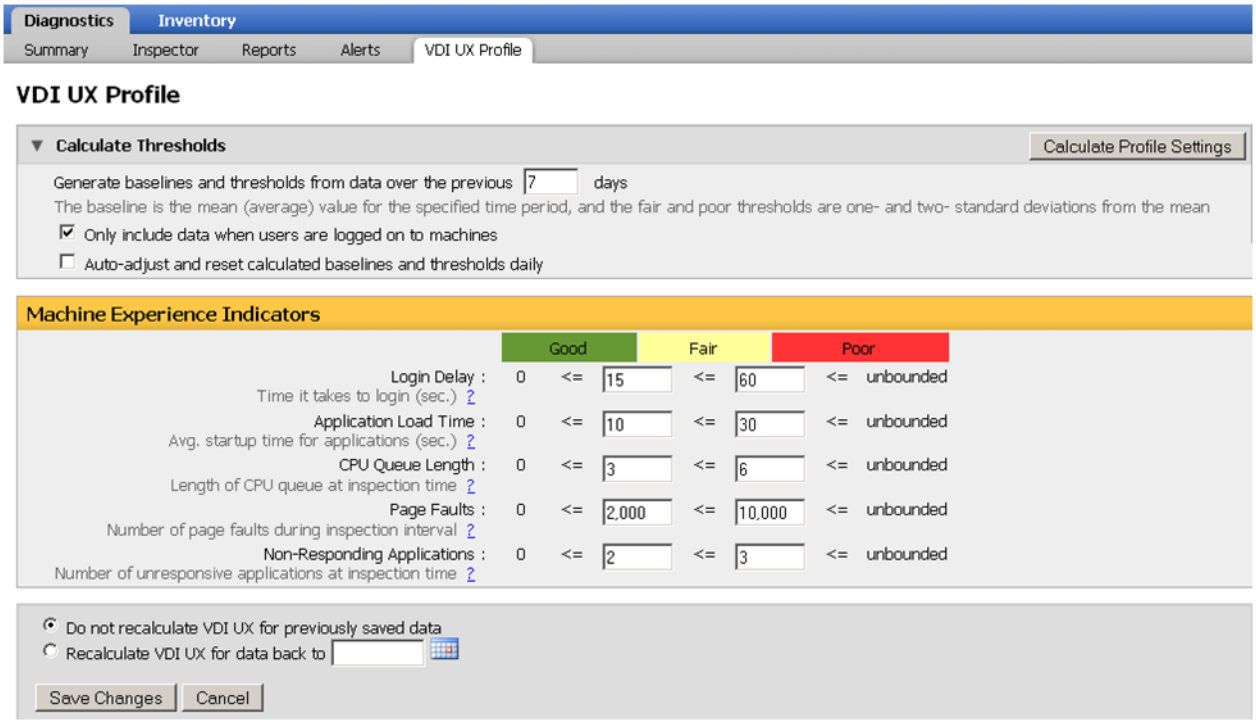


As the screenshot in Figure 9 shows, Liquidware Labs Stratusphere UX defines the user experience in terms of “good,” “fair,” and “poor.” Good for a login is defined as any login that takes less than 15

seconds, fair as any login that takes less than 60 seconds, and poor as any login that takes greater than 60 seconds. The following charts apply Liquidware Labs standards for user experience.

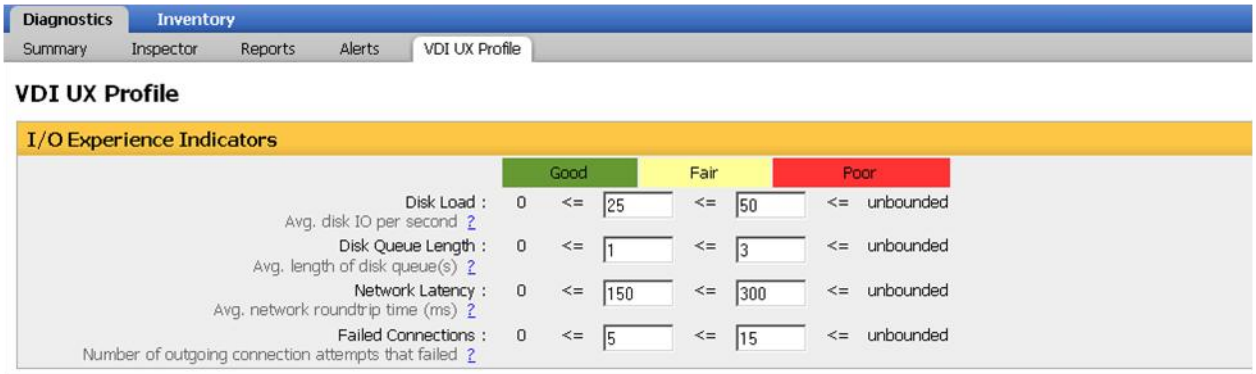
Note: The definitions of “good,” “fair,” and “poor” are given no flexibility based on type of login or size of profile. Thus, a time limit of 15 seconds or better is required for a classification of good, regardless of the size or state of the profile.

Figure 9) VDI UX Profile screen with machine experience indicators.



As the screenshot in Figure 10 shows, Liquidware Labs Stratusphere UX defines network latency thresholds as “good” for less than or equal to 150ms, “fair” for less than or equal to 300ms, and “poor” for greater than 300ms.

Figure 10) VDI UX Profile screen with I/O experience indicators.



We used the VMware vscsiStats utility on vSphere 4.1, taking advantage of its trace flag to characterize VM disk I/O workload from within the ESX servers. Of particular interest to our testing was vscsiStats with the trace options records I/O block size and command type. We used this data to confirm the operation mixtures and operation sizes encountered in the various tests. From this, we were able to document what

each application did on first use since reboot and compare this to what each application did on second use.

We also used Perfstat, a diagnostic data collection tool written by NetApp.

5 DETAILED TEST RESULTS

The remainder of this report focuses on the goals of demonstrating advancements and characterizing workloads from power-on to steady state. For the sake of thoroughness, the tools used for data collection/workload generation are discussed. There is also the following section on the creation of VMs using the NetApp Provisioning and Cloning and a section covering the characteristics of the individual Microsoft applications (see section 6.2, “Application Workloads” in the appendix).

The performance results are compared, adding the category of user experience where appropriate. There is also a focus on the observed concurrency, read/write mixture, and random sequential mixture for each workload.

For ease of reading, all NetApp testing is superimposed on a two-week calendar where all events are scheduled. The purpose of this overlay is to take the reader through a “day in the life of” a View administrator at the fictitious Acme Corporation. All events as reported occurred in NetApp testing; all timelines are factual; only the superimposed story of Acme Corporation is fictional.

5.1 CREATION PROCESS

Sun	Mon	Tue	Wed	Thurs	Fri	Sat
29 Deploy 2,500 desktops	30 8 a.m. 2,500 logins + Profile load (30 min) Typical workday	31 8 a.m. 2,500 logins (30 min) Typical workday	1	2	3	4
5	6 1 a.m. Network maintenance 2 a.m. Network outage 7 a.m. Reboot all VMs 8 a.m. 2,500 logins (30 min) (Post-power-on)	7 8 a.m. 2,500 logins (30 min) Typical workday	8	9	10	11

THE STORY

Acme Corporation’s Change Control board has approved the date of Sunday the 29th for the build-out of an additional 2,500 virtual desktops. By Sunday morning, the physical environment is in place, as are the View connection brokers, the requisite number of ESX Servers, and a new vCenter VM. The Windows 7 64-bit image has been optimized previously, according to VMware and Microsoft best practices, with all of the desired applications installed.

For the deployment, Acme Corporation has chosen to take advantage of NetApp cloning technology in preference to other choices. As of Sunday morning, the View administrator is ready to launch the creation process.

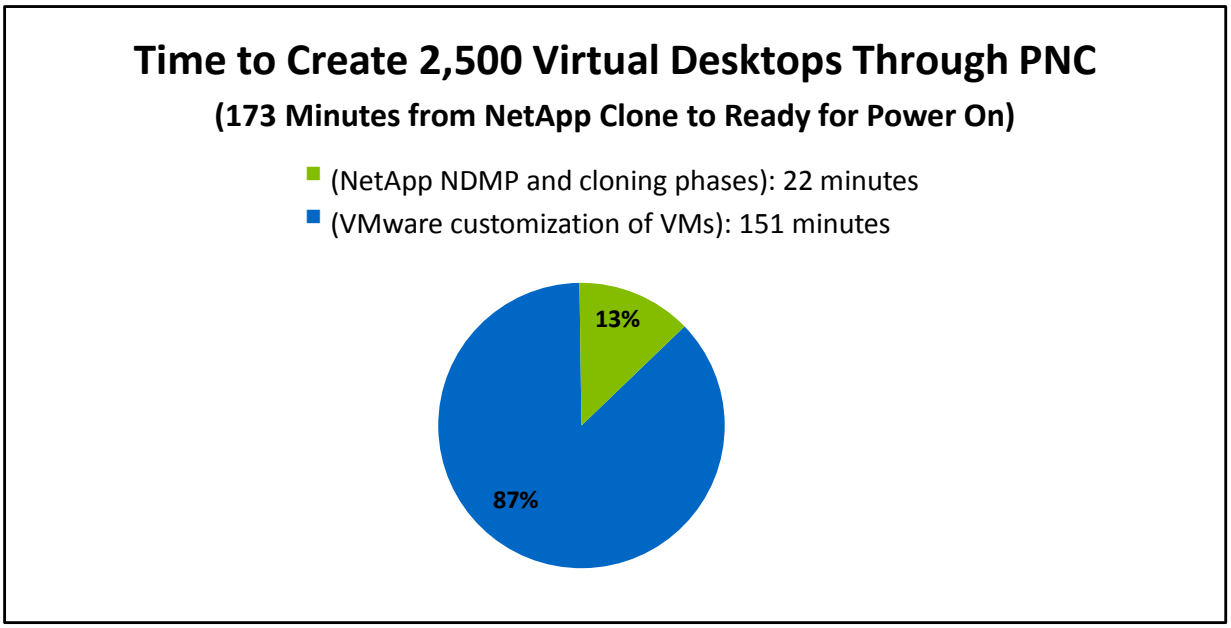
THE FACTS

The build-out of the environment took advantage of the NetApp Provisioning and Cloning Capability (PNC), a subcomponent of the Virtual Storage Console (VSC) suite. The environment’s 2,500 virtual

desktops were spread across 10 volumes (NFS datastores), with one additional volume serving as a "staging" volume. As the first step, VSC physically copied the template first to the staging volume created by VSC for this purpose. VSC then file-cloned the flat-vmdk newly located in the staging volume 249 times, and finally cloned the staging volume itself 10 times. The copying of the template to the staging volume was performed through Network Data Management Protocol (NDMP) in the form of a backup and restore.

As the graph in Figure 11 depicts, the cloning process was broken down into various stages. The NetApp NDMP and cloning of the flat-vmdk files were two of the earliest stages. At the completion of these stages, control was passed back to the virtual center for the pre-power-on customization specification of the virtual desktops, among other steps. Note that the NDMP process took 3 of the 22 minutes to back up and restore the 20GB template flat-vmdk (C:), the file-clone phase took 19 minutes, and the volume clone phase 16 seconds.

Figure 11) Breakdown of times for the stages of cloning virtual desktops.



5.2 INITIAL LOGIN

Sun	Mon	Tue	Wed	Thurs	Fri	Sat
29 Deploy 2,500 desktops	30 8 a.m. 2,500 logins + Profile load (30 min) Typical workday	31 8 a.m. 2,500 logins (30 min) Typical workday	1	2	3	4
5	6 1 a.m. Network maintenance 2 a.m. Network outage 7 a.m. Reboot all VMs 8 a.m. 2500 logins (30 min) (Post-power-on)	7 8 a.m. 2,500 logins (30 min) Typical workday	8	9	10	11

THE STORY

Acme Corporation's employees generally arrive between 8 and 8:30 a.m. each morning, and Monday morning is no exception. Because the 2,500 virtual desktops were freshly deployed, however, Monday morning the 29th was slightly different. On this morning, the View connection broker assigned each of the 2,500 users a new desktop, and because Acme Corporation had not yet implemented any profile management software, each initial login generated the creation of a new user profile.

THE FACTS

All users were logged in over a 28-minute period using the method described in section 3.5, "Tools."

At initial login, all users began their day's work. The VMware desktop RAWC controlled all of the workload in the environment, opening applications and standing in for human control at the keyboard. RAWC performed the following tasks on each virtual desktop:

- Configured Outlook and set up the Outlook client for cached mode
- Opened Microsoft Word and Excel for each user, creating and saving new documents in each
- Opened Microsoft PowerPoint as well as Adobe Acrobat Reader and reviewed existing documents in each
- Opened Microsoft Internet Explorer
- Wrote and sent three e-mails from Outlook client

RAWC randomly determined the order in which the applications were run on each virtual desktop.

USER EXPERIENCE

Table 8 reports the login times experienced by the user community on Monday morning during the initial login scenario.

Key Point
Upgrading from Data ONTAP 8.0.1 to Data ONTAP 8.1 decreased average user login time by 46% (from 46 down to 25 seconds).

The login times were captured with Liquidware Labs Stratusphere UX. As mentioned, Liquidware Labs defines a 15-second threshold for a good login user experience, less than 60 seconds for a fair user login experience, and more than 60 seconds for a poor user login experience.

Table 8) User experience of initial login (in seconds).

Configuration	Time to Log In: Average User Experience	Time to Log In: Maximum User Experience	Time to Log In: Login Standard Deviation for User Experiences
Data ONTAP 8.0.1	46 seconds	124 seconds	34 seconds
Data ONTAP 8.1	25 seconds	77 seconds	20 seconds

Note: FAS3270 with 48 450GB FC 15K, loop rate: 4Gbps.

Key Points
• Upgrading from Data ONTAP 8.0.1 to Data ONTAP 8.1 decreased the number of users having a poor experience at initial login (from 38% down to 3% of users). • The poor experience of users improved by 25% between Data ONTAP 8.0.1 and Data ONTAP 8.1 (from 86 down to 64 seconds).

Thus, the poor experiences became fewer, and those who still had a poor experience nevertheless had a better experience. Table 9 shows the login experiences of the users in the environment. For example, 45% of the users had a good login experience, and their logins took 5 seconds on average.

Table 9) User experience of initial login (in percentages of good, fair, and poor login time).

Configuration	Total Number of Users	% Users with Good Login Time (<= 15 sec)	% Users with Fair Login Time (<= 60 sec)	% Users with Poor Login Time (=> 60 sec)
Data ONTAP 8.0.1	2,500	27% (6-sec. avg.)	35% (35-sec. avg.)	38% (86-sec. avg.)
Data ONTAP 8.1	2,500	45% (5-sec. avg.)	52% (40-sec. avg.)	3% (64-sec. avg)

Note: FAS3270 with 48 450GB FC 15K, loop rate: 4Gbps.

In the following section, note that the storage controller's utilization is at or near capacity. At the same time, the user experiences as reported by Liquidware Labs Stratusphere VDI UX are acceptable for application load times.

SYSTEM EXPERIENCE

The storage controller behaved on a high level, as described in Table 10. Notice that the throughput increased by 140% after upgrading from Data ONTAP 8.0.1 to 8.1, whereas the latencies on the storage controller decreased 80% for a marginal increase in CPU utilization. Refer to the discussion on virtual storage tiering in section 1, "Executive Summary," for details on key contributing technology enhancement behind the improvement.

For the initial login scenario, the single 10GbE network interface card (NIC) was the chief bottleneck. The NIC sent transmit-pause frames for 5% of the packets received. This action resulted in client-side latencies far and above the latencies reported by the storage controller. To alleviate this chokepoint, NetApp recommends splitting the workload across two 10GbE NICs on separate cards.

Key Points

- A 140% increase in throughput directly affected the time required to complete the login phase for all 2,500 virtual desktops, which made for a better user experience individually and overall.
  - A decrease in storage controller latencies by 80%
  - Only a marginal increase in CPU utilization of less than 5%
- User latencies were well within the standards defined as "good" by Liquidware Labs UX (good <= 150ms, fair <= 300ms, and poor > 300ms).

Table 10 compares Data ONTAP 8.0.1 and 8.1, breaking out throughput, operations per second, and latencies.

Table 10) Data ONTAP 8.0.1 versus 8.1 during initial login.

Data ONTAP	Read Ops/s	Write Ops/s	Read MB/s	Write MB/s	Read Latency		Write Latency		CPU Utilization
					Controller	ESX	Controller	ESX	
Data ONTAP 8.0.1	11,522	9,783	216MB/s	98MB/s	5.0ms	64ms	4.3ms	64ms	93%
Data ONTAP 8.1	14,470	12,285	311MB/s	129MB/s	0.9ms	51ms	1.2ms	64ms	95%

Note: FAS3270 with 48 450GB FC 15K, loop rate: 4Gbps.

From this point on in this section, graphs detailing the Data ONTAP 8.1 configuration are displayed for brevity. We observed similar curves and metrics for 8.0 but with lower throughput, as described in Table 10.

Figure 12 shows the read and write throughput generated by both the initial logins of 2,500 users and their subsequent start-of-day workload. Users began working as soon as they logged in, so the application loads overlap the logins and the profile loads. Reads were responsible for 71% (avg. 311MB/s) of data passed over the network, and writes for the remaining 29% (avg. 129MB/s).

Figure 12) Read and write throughput at initial login.

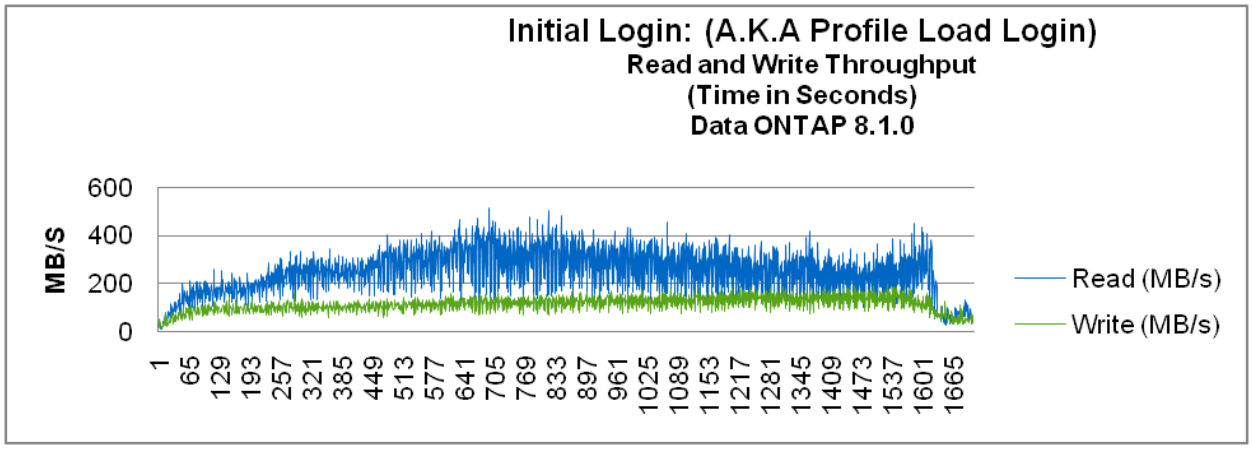


Figure 13 shows the read/write operations generated per second by both the initial logins of 2,500 users and their subsequent start-of-day workload. The read operations accounted for 53% of the NFS workload, and the write operations for 43%. Lookups (not displayed) accounted for the remaining approximately 2%.

Note: Users began working as soon as they logged in, so the application loads overlap the logins and the profile loads.

Figure 13) Read and write operations per second at initial login.

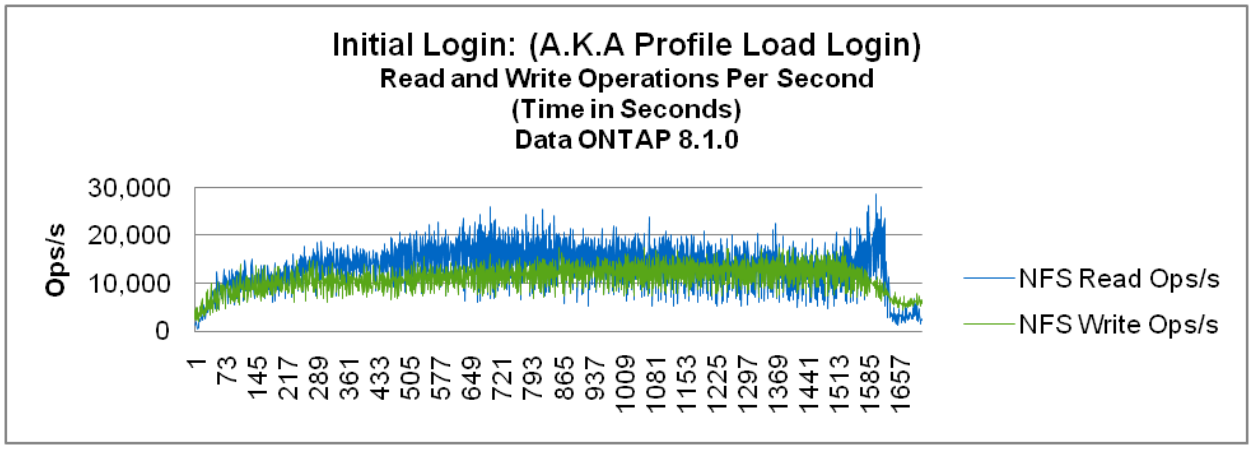


Figure 14 shows the latencies as reported by the storage controller for the NFS protocol. Both read (avg. 0.9ms) and write (avg. 1.2ms) protocol latencies are shown for the entire login time.

Figure 14) Read/write protocol latencies at initial login.

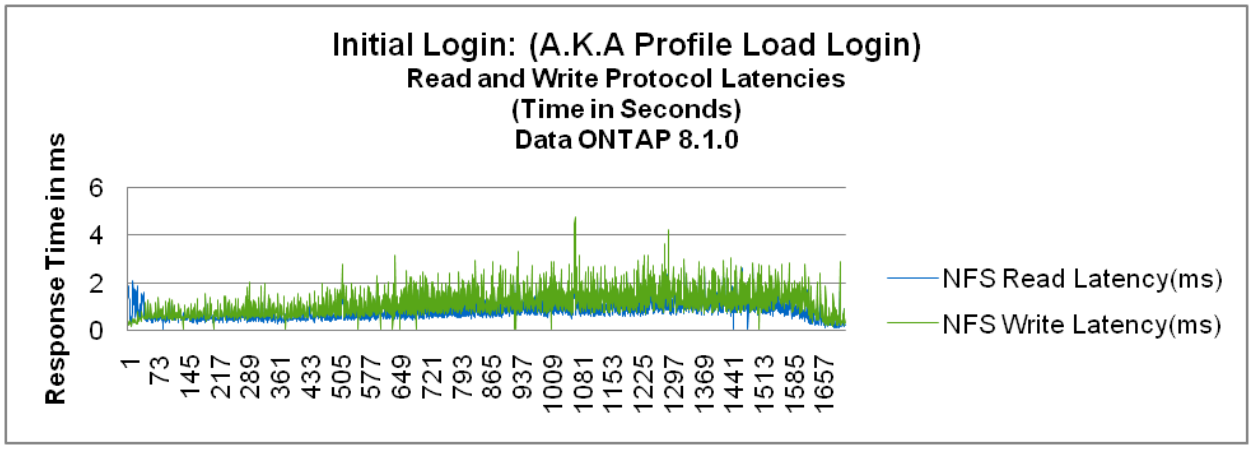
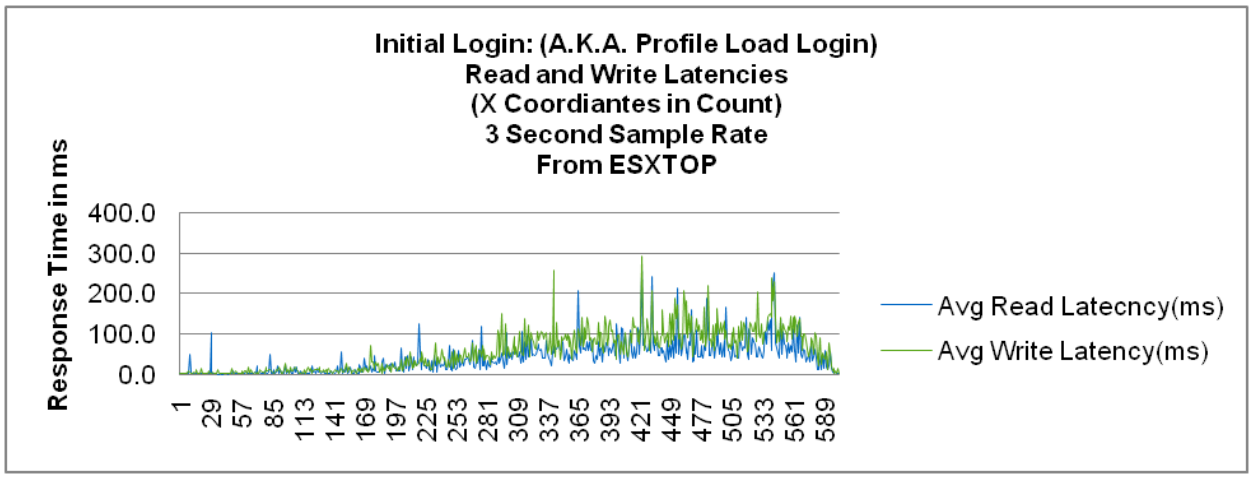


Figure 15 shows guest latencies as reported by ESXTOP batch mode and compiled by ESXTOP. ESXTOP batch mode has a 3-second minimum sample rate. The X-axis is the sample number. Because this graph represents 3-second sample rates, multiply the X-axis value by 3 to get the true run time.

Key Point

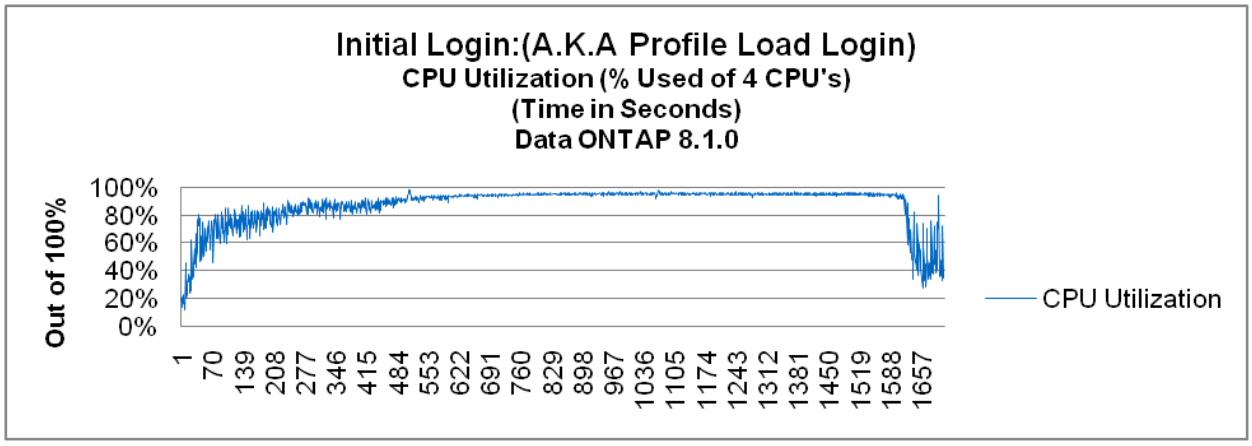
The latencies seen by the guest are significantly higher than those reported by the storage controller, although still within the bounds defined as “good” by Stratusphere UX. As stated earlier, the single 10GbE NIC was the chief bottleneck during this scenario. The NIC sent transmit-pause frames for 5% of the packets received. This resulted in the client-side latencies shown in Figure 15. (This issue occurred only during the initial login scenarios, in each of the Data ONTAP releases tested.) To alleviate this chokepoint, NetApp recommends splitting the workload across two 10GbE NICs on separate cards.

Figure 15) Read and write latencies at initial login.



During this time, as Figure 16 shows, on average 95% of the capacity of all four physical CPUs was consumed. Note that this CPU utilization was maintained while storage controller latencies remained low.

Figure 16) CPU utilization at initial login.



WORKLOAD CHARACTERISTICS

When this testing began, NetApp accepted the industry-standard sizing practice for virtual desktops, namely, to lump all users in buckets based on the number of operation the users were expected to generate. Furthermore, virtual desktop workload had been defined as containing no sequential I/O, and as having all of its I/O at either 4KB or 8KB. Furthermore, when sizing for virtual desktops, we traditionally assumed that 100% of the desktops would generate IOPS 100% of the time. However, packet traces, vscsiStats and storage controller stats, and ESXTOP have proven otherwise.

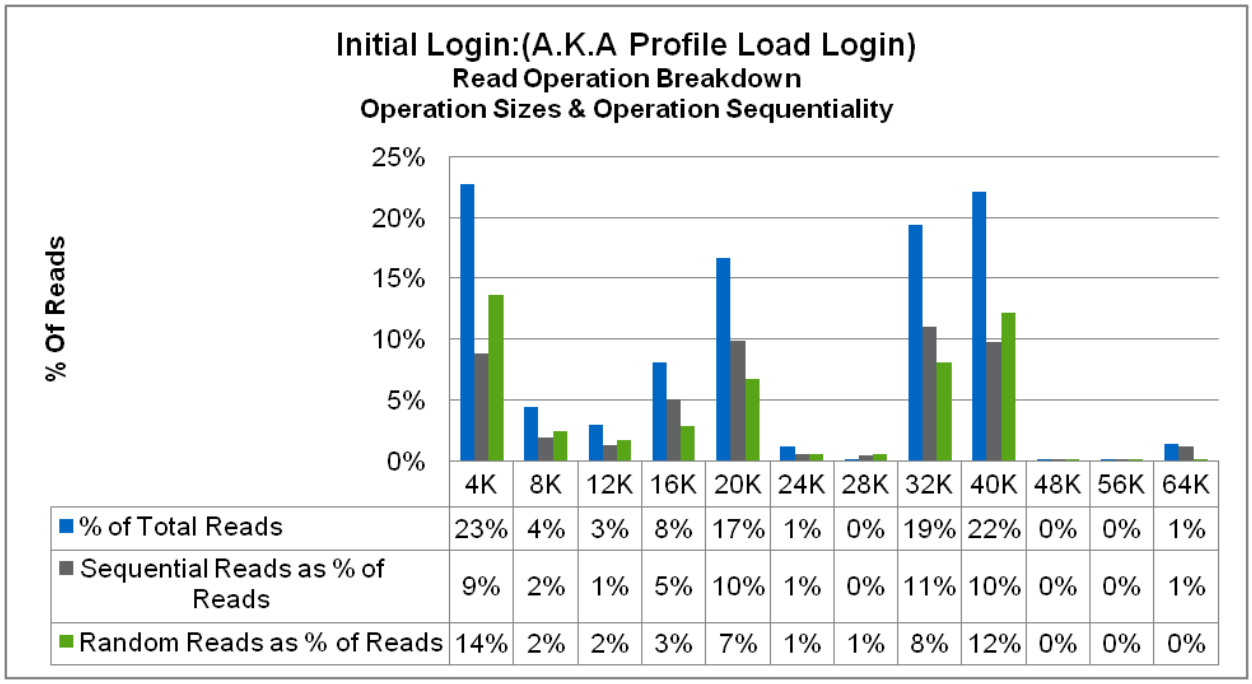
Each of the workload scenarios contains a section similar to this one that breaks out the operation sizes, random/sequential mixture, and concurrency of user operations. For the sake of this report, concurrency is defined as the number of virtual desktops generating storage-targeted I/O at the same time.

The read operation sizes and their respective natures, sequential or random, are documented in the chart in Figure 17. The statistics themselves were taken from counter manager read-ahead statistics captured on the storage controller during the test.

Key Points

- The workload is not all one size.
  - The graph in Figure 17 is broken down into operation buckets. Each bucket contains all operations from the size specified down to the next reported operation size.
- The workload is not all random: 50% of all reads are sequential.

Figure 17) Read operation breakdown for initial login.



	4K	8K	12K	16K	20K	24K	28K	32K	40K	48K	56K	64K
% of Total Reads	23%	4%	3%	8%	17%	1%	0%	19%	22%	0%	0%	1%
Sequential Reads as % of Reads	9%	2%	1%	5%	10%	1%	0%	11%	10%	0%	0%	1%
Random Reads as % of Reads	14%	2%	2%	3%	7%	1%	1%	8%	12%	0%	0%	0%

Table 11 documents the workload from the perspective of the virtual desktops themselves as captured in ESXTOP. During our testing, ESXTOP ran continuously in batch mode on all servers in the environment, with the lowest possible sample interval of 3 seconds.

We extracted the following information from ESXTOP output:

- The number of virtual desktops that were simultaneously performing read or write operations concurrently at different intervals during a specific test. This gave us a measure of the “concurrency” of desktops actively accessing the storage at any given time.
- The average read and write operations and throughput in megabytes per second generated by each active desktop.
- The average size of each read and write operation.

In Table 11, the charts generated by ESXTOP document the concurrency, I/O rate, and operation size from the perspective of the virtual desktops, as reported by ESXTOP. The purpose of this table is to illustrate that concurrency plays a major role and that individual working virtual machines may be very busy, but the sum of all virtual machines is much lower.

Key Points

- On average, only 20% of the virtual desktops generated reads, and 70% generated writes during any given second. Therefore, 100% concurrency was not achieved.
- On average, if concurrency is ignored, each virtual desktop generates 12 IOPS during the initial login scenario.
- ESXTOP confirms the data from the storage controller indicating that the IOPS are larger than the 4KB or 8KB initially assumed as common for VDI workloads.
- The average IOPS and throughput reported from ESXTOP closely track the values reported by the storage controller.

Table 11) I/O concurrency, rate, and size for read and write operations at initial login.

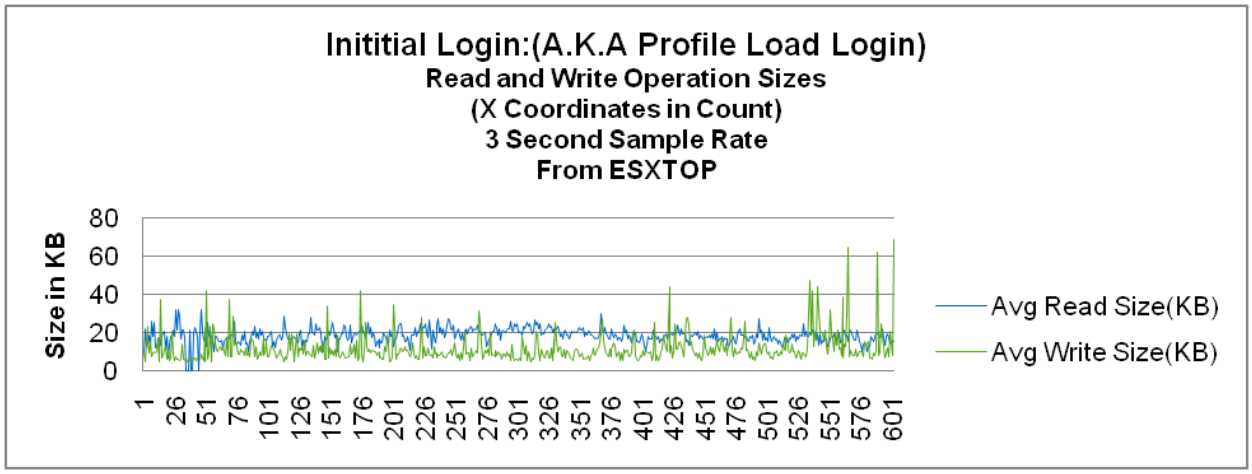
Subjects Measured	Values	
Total VM count	2,500	
Avg total IOPS	26,184	
Avg read IOPS	15,128	
Avg write IOPS	11,056	
Avg read throughput (MB/sec)	275	
Avg write throughput (MB/sec)	128.0	
# of reading VMs	550	
# of writing VMs	1,780	
Read size avg (KB)	19	
Write size avg (KB)	12	
Read latency avg (ms)	51.9	
Write latency avg (ms)	65	
	For One VM	For All VMs
Avg read IOPS per reading VM	28	6
Avg write IOPS per writing VM	6	4
Avg read throughput per reading VM (KB/sec)	512	115
Avg write throughput per writing VM (KB/sec)	58	53

In Figure 18, the graph generated by ESXTOP also shows the average read and write operation sizes as issued by the guests themselves throughout the entire login scenario.

Key Point

The graph in Figure 18 shows that the reads and writes fluctuate widely from sub-4K to greater than 60KB.

Figure 18) Read and write operation sizes at initial login.



5.3 TUESDAY MORNING LOGIN

Sun	Mon	Tue	Wed	Thurs	Fri	Sat
29 Deploy 2,500 desktops	30 8 a.m. 2,500 logins + Profile load (30 min) Typical workday	31 8 a.m. 2,500 logins (30 min) Typical workday	1	2	3	4
5	6 1 a.m. Network maintenance 2 a.m. Network outage 7 a.m. Reboot all VMs 8 a.m. 2,500 logins (30 min) (Post-power-on)	7 8 a.m. 2,500 logins (30 min) Typical workday	8	9	10	11

THE STORY

When the Acme Corporation employees left work on Monday evening after doing their day's work, they universally logged out, after which the virtual desktops remained mostly idle and ready for Tuesday morning, when, as they did every day, the users arrived and logged in between 8 and 8:30 a.m. Even though the users logged off Monday night, the memory of each virtual desktop had remained populated with user data (libraries, cached data, and so forth).

THE FACTS

The users were assigned virtual desktops at initial login, which occurred on a previous day. The users had logged in and opened their applications on previous days, and, although they had shut down their applications and logged off, the application libraries and much of each user's profile remained resident in memory. Therefore, this login and start-of-day generated a reduced storage workload. As in all of the login and workload scenarios captured in this report, the user login rate occurred at approximately 3 logins per second, with the final login being attempted 26 minutes after the first.

At login on Tuesday morning, all users began their day's work. The VMware desktop RAWC controlled all of the workload in the environment, opening applications and standing in for human control at the keyboard. RAWC performed the following tasks on each virtual desktop:

- Opened Microsoft Word and Excel for each user, creating and saving new documents in each
- Opened Microsoft PowerPoint as well as Adobe Acrobat Reader and reviewed existing documents in each
- Opened Microsoft Internet Explorer
- Wrote and sent three e-mails from Outlook client

RAWC randomly determined the order in which the applications were run on each virtual desktop by RAWC.

USER EXPERIENCE

Table 12 reports the login times experienced by the user community on Tuesday morning during the typical login scenario.

Key Point
The average user experienced a 1-second login on Tuesday morning, regardless of the Data ONTAP version used.

Table 12) User experience of Tuesday morning login (in seconds).

Configuration	Time to Log In: Average User Experience	Time to Log In: Maximum User Experience	Time to Log In: Standard Deviation for User Experiences
Data ONTAP 8.0.1	1 second	1 second	0 seconds
Data ONTAP 8.1	1 second	1.5 seconds	0 seconds

Note: FAS3270 with 48 450GB FC 15K, loop rate: 4Gbps.

The Tuesday morning login scenario resulted in a 100% “good” user login experience (as defined by Liquidware Labs Stratusphere UX to mean less than 15 seconds).

Table 13) User experience of Tuesday morning login (in percentages of good, fair, and poor login time).

Configuration	Total Number of Users	% Users with Good Login Time (<= 15 sec)	% Users with Fair Login Time (<= 60 sec)	% Users with Poor Login Time (=> 60 sec)
Data ONTAP 8.0.1	2,500	100% (1 sec. avg.)	0%	0%
Data ONTAP 8.1	2,500	100% (1 sec. avg.)	0%	0%

Note: FAS3270 with 48 450GB FC 15K, loop rate: 4Gbps.

The user experiences as reported by Liquidware Labs Stratusphere VDI UX are acceptable for application load times.

SYSTEM EXPERIENCE

The storage controller behaved on a high level as described in Table 14. Notice that the amount of work passed between the virtual desktops and the storage controller was fairly insignificant and that it was dominated by write operations because the virtual desktops took advantage of the fact that large amounts of information remained in the guest OS cache. Besides the file-save operations, the writes came mostly from background processes such as System, `svchost.exe`, and services.

Key Point
As Table 14 shows, even though the operations are write dominant, without the required work of loading application DLLs or profiles from disk, the demands on the storage controller are decreased.

For further details, Table 14 compares Data ONTAP 8.0.1 and 8.1, breaking out throughput, operations per second, and latencies.

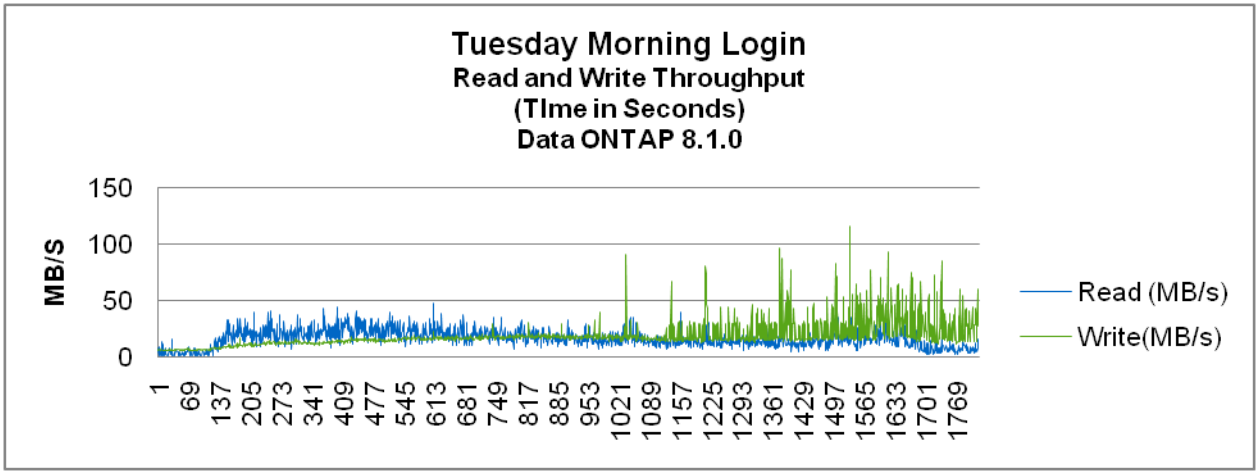
Table 14) Data ONTAP 8.0.1 versus 8.1 during Tuesday morning login.

Data ONTAP	Read Ops/s	Write Ops/s	Read MB/s	Write MB/s	Read Latency		Write Latency		CPU Utilization
					Controller	ESX	Controller	ESX	
Data ONTAP 8.0.1	860	3,794	26MB/s	18MB/s	0.9ms	1.7ms	0.4ms	0.9ms	27%
Data ONTAP 8.1	760	3,900	17MB/s	20MB/s	0.3ms	1.2ms	0.2ms	0.6ms	23%

From this point on in this section, graphs detailing the Data ONTAP 8.1 configuration are displayed for brevity. We observed similar curves and metrics for Data ONTAP 8.0 but with lower throughput, as described Table 14.

The graph in Figure 19 shows the read and write throughput generated by both the Tuesday morning logins of 2,500 users and their subsequent start-of-day workload. Users began work as soon as they logged in, so the application loads overlap the login and profile loads. Reads were responsible for 46% (avg. 17MB per second) of data passed over the network, writes for the remaining 54% (avg. 20MB per second).

Figure 19) Read and write throughput for Tuesday morning login.



The graph in Figure 20 shows the read/write operations generated per second by both the logins of 2,500 users and their subsequent start-of-day workload. The read operations accounted for 15% of the NFS workload, and the write operations for 77%. Lookups (not displayed) accounted for approximately the remaining 6%. The users began work as soon as they logged in, so the users' workload and the background work overlap with the logins.

Key Points
• As in the earlier cases, the first login began 123 seconds into the time shown on the graph. • Before the first login had been attempted, there were approximately 2,000 writes per second, with very few read operations. We investigated the source of these operations and found that System, `svchost.exe`, and `services.exe` running on the individual VMs were responsible for this I/O. More is said about this in section 5.7, "Observations and Lessons Learned." • The average write size remained at 4KB postlogin, even though the operations per second increased approximately 100%.

Figure 20) Read and write operations per second for Tuesday morning login.

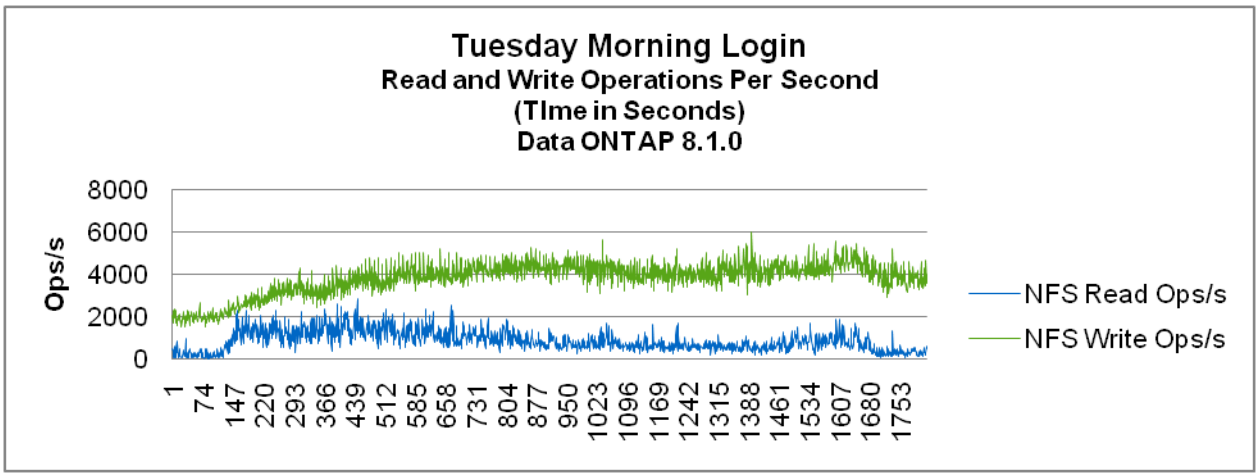
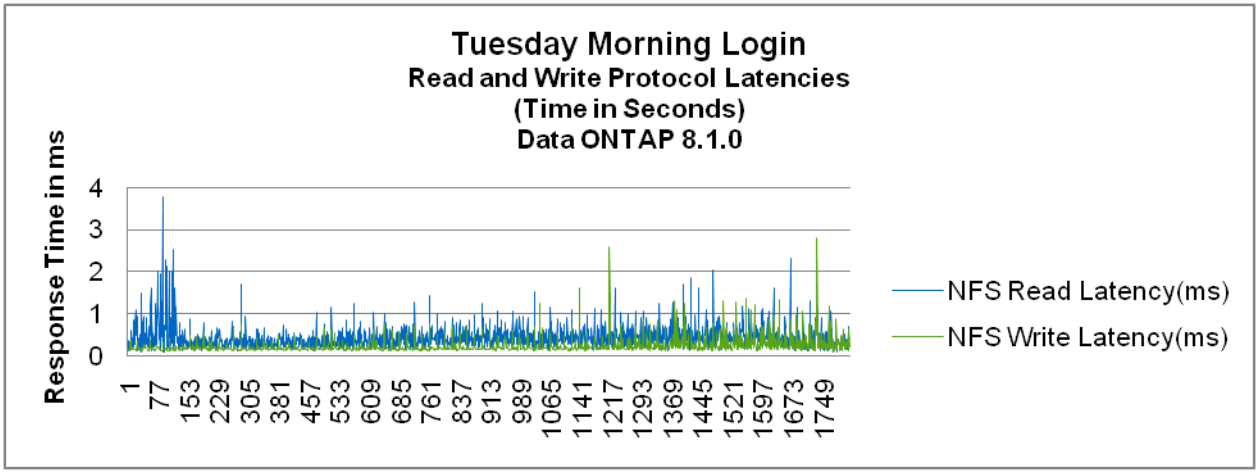


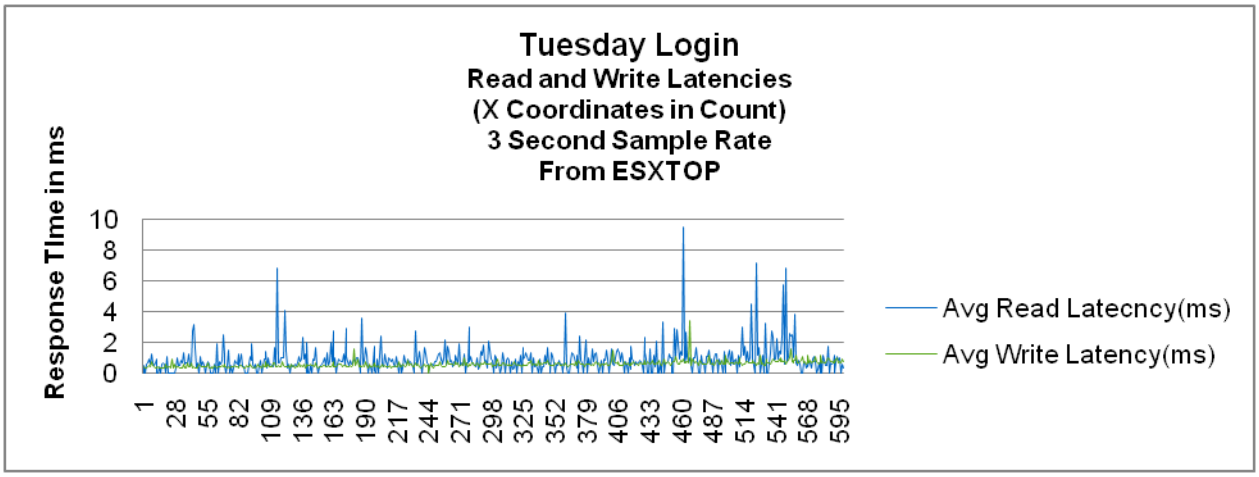
Figure 21 shows the latencies as reported by the storage controller for the NFS protocol. Both read (avg. 0.3ms) and write (avg. 0.2ms) protocol latencies are shown for the entire login time.

Figure 21) Read and write protocol latencies for Tuesday morning login.



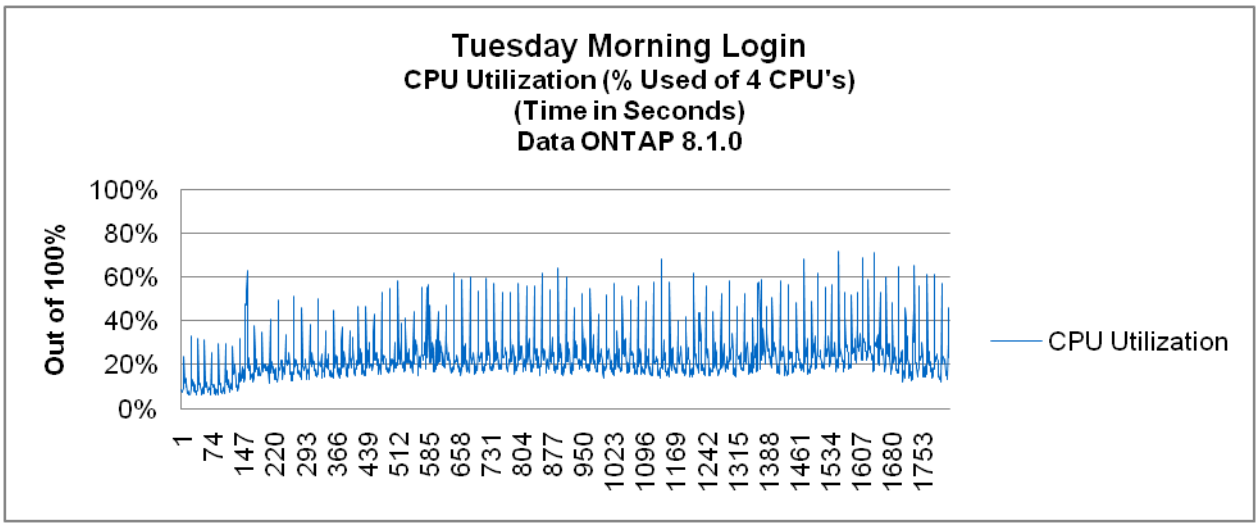
The graph in Figure 22 shows guest latencies as reported by ESXTOP batch mode and compiled by ESXTOP. ESXTOP batch mode has a 3-second minimum sample rate. The X-axis is the sample number. Because this graph represents 3-second sample rates, multiply the X-axis value by 3 to get the true run time. Notice that the latencies track along those on the storage controller (with no surprises).

Figure 22) Read and write latencies for Tuesday morning login.



During this time, as the graph in Figure 23 shows, on average, 23% of the capacity of all four physical CPUs was consumed.

Figure 23) CPU utilization for Tuesday morning login.



WORKLOAD CHARACTERISTICS

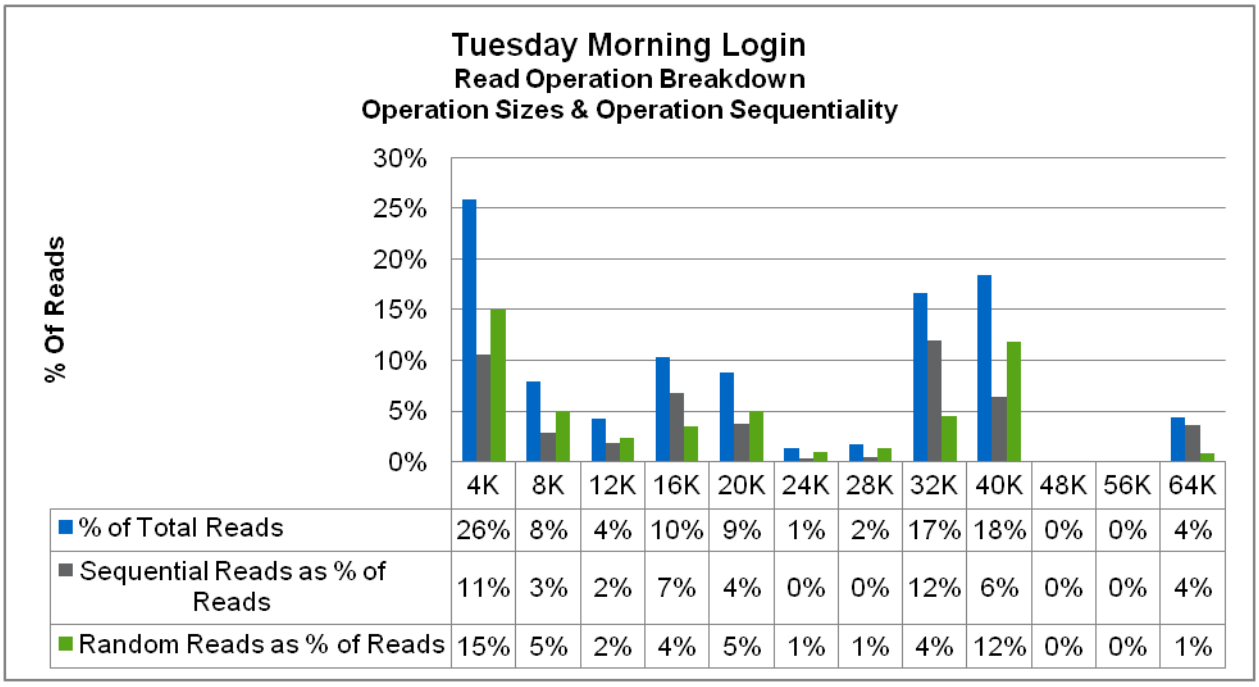
As with the other workloads outlined in this paper, the workload characteristics of the Tuesday morning login scenario also break with tradition in terms of I/O sizes, sequentiality, and concurrency. Recall that "concurrency," as used here, refers to the number of virtual desktops generating storage-targeted I/O at the same time.

The read operation sizes and their respective natures, sequential or random, are documented in Figure 24. The statistics themselves were taken from counter manager read-ahead statistics captured on the storage controller during the test.

Key Points

- The workload is not all one size.
  - The graph in Figure 24 is broken down into operation buckets. Each bucket contains all operation from the size specified down to the next reported operation size.
- The workload is not all random: 50% of all reads are sequential.

Figure 24) Read operation breakdown for Tuesday morning login.



	4K	8K	12K	16K	20K	24K	28K	32K	40K	48K	56K	64K
% of Total Reads	26%	8%	4%	10%	9%	1%	2%	17%	18%	0%	0%	4%
Sequential Reads as % of Reads	11%	3%	2%	7%	4%	0%	0%	12%	6%	0%	0%	4%
Random Reads as % of Reads	15%	5%	2%	4%	5%	1%	1%	4%	12%	0%	0%	1%

Table 15 documents the workload from the perspective of the virtual desktops themselves as captured in ESXTOP. During our testing, ESXTOP ran continuously in batch mode on all servers in the environment with the lowest possible sample interval of 3 seconds.

In Table 15, the charts generated by ESXTOP document the concurrency, I/O rate, and operation size from the perspective of the virtual desktops as reported by ESXTOP.

Key Points

- On average, 4% of the virtual desktops generated reads, and 62% generated writes during any given second. Therefore, 100% concurrency was not achieved.
- On average, if concurrency is ignored, each virtual desktop generated 2 IOPS during the Tuesday morning login scenario.
- ESXTOP confirms the data from the storage controller indicating that the IOPS are larger than 4KB or 8KB
- The average IOPS and throughput reported from ESXTOP closely approximate the values reported by the storage controller.

Table 15) I/O concurrency, rate, and size for read and write operations at Tuesday morning login.

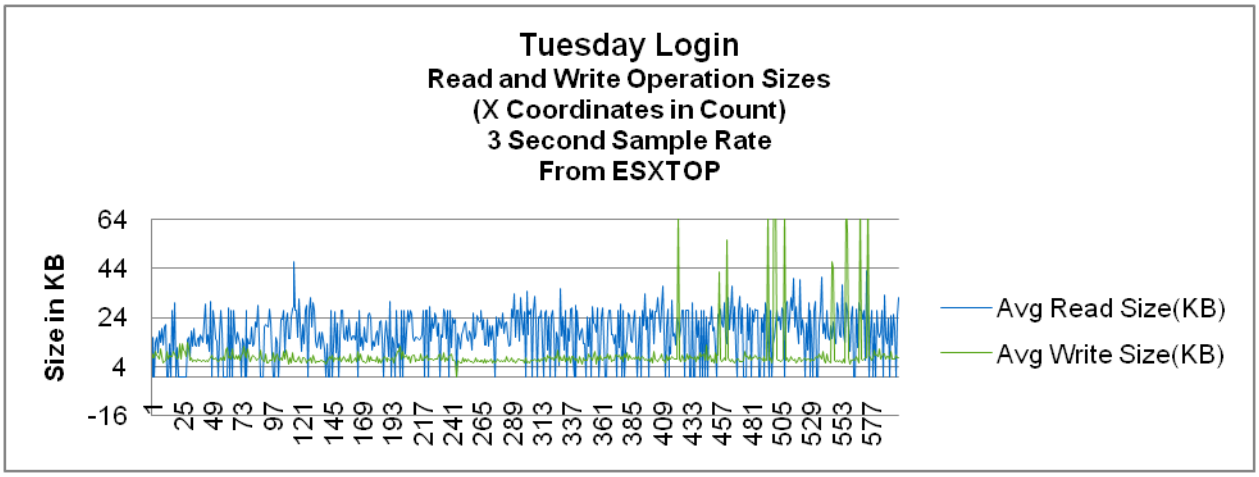
Subjects Measured	Values	
Total VM count	2,500	
Avg total IOPS	4,659	
Avg read IOPS	1,192	
Avg write IOPS	3,467	
Avg read throughput (MB/sec)	22	
Avg write throughput (MB/sec)	27.2	
# of reading VMs	92	
# of writing VMs	1,572	
Read size avg (KB)	19	
Write size avg (KB)	8	
Read latency avg (ms)	1.2	
Write latency avg (ms)	1	
	For One VM	For All VMs
Avg read IOPS per reading VM	13	.05
Avg write IOPS per writing VM	2	1.4
Avg read throughput per reading VM (KB/sec)	242	9
Avg write throughput per writing VM (KB/sec)	18	11

In Figure 25, the graph generated by ESXTOP also shows the average read and write operation sizes as issued by the guests themselves throughout the entire login scenario.

Key Point

As the graph in Figure 25 shows, the reads and writes fluctuate widely from sub-4K to greater than 100KB.

Figure 25) Read and write operation sizes for Tuesday morning login.



The workload characteristics of the Tuesday-morning login scenario are thus neither 4KB nor 8KB but a mixture of all sizes.

5.4 REBOOT

Sun	Mon	Tue	Wed	Thurs	Fri	Sat
29 Deploy 2,500 desktops	30 8 a.m 2,500 logins + Profile load (30 min) Typical workday	31 8 a.m. 2,500 logins (30 min) Typical workday	1	2	3	4
5	6 1 a.m. Network maintenance 2 a.m. Network outage **7 a.m. Reboot all VMs** 8 a.m. 2,500 logins (30 min) (Post-power-on)	7 8 a.m. 2,500 logins (30 min) Typical workday	8	9	10	11

THE STORY

A networking error occurred during the scheduled maintenance window at Acme Corporation early Monday morning. This outage affected communication between the storage controllers and the ESX servers. The outage ended at approximately 6:30 a.m., at which point the virtual desktop administrator had just over an hour to make sure that all 2,500 virtual desktops were available for use before the employees began arriving at 8 a.m. for the scheduled start of day. To confirm that the virtual desktops were fully available, a full reboot was chosen. All virtual desktops were first powered off, then confirmed down, and then powered up once more. The powering on of all 2,500 virtual desktops to the point at which all I/O settled down to the normal preuser login took just over 20 minutes. No users were affected.

THE FACTS

The power-on process was controlled centrally by vCenter by selecting all 2,500 virtual desktops and letting vCenter perform the power-ups. VCenter limits the concurrent power-on operations to 128 at a time, the remainder being placed in queue waiting to fill slots as individual virtual desktops exit the service

center after successful power-on. Thus, when sizing is determined for power-on operations, the total IOPS generated per desktop must be multiplied by the roughly 128 rather than by the full complement of virtual desktops in the environment. After successful mass power-up, the virtual desktops sat idling, waiting for the start of day.

SYSTEM EXPERIENCE

The storage controller behaved on a high level as described in Table 16. Notice that the workload was read dominant and the storage controller's CPU utilization hovered at full utilization throughout the power-on process. The goal of the mass power-on scenario was to power-on all of the virtual desktops as quickly as possible. Keep in mind that throughput and concurrency, not latency, are the dominant factors in the mass power-on scenario.

Key Point

Because of such technological enhancements as Virtual Storage Tiering that are new to 8.1, the time taken to power-on was decreased by 42% by upgrading from Data ONTAP 8.0.1 to 8.1 (from 36 down to 21 minutes).

For further details, Table 16 compares Data ONTAP 8.0.1 and 8.1, breaking out throughput, operations/sec and latencies, and power-on times.

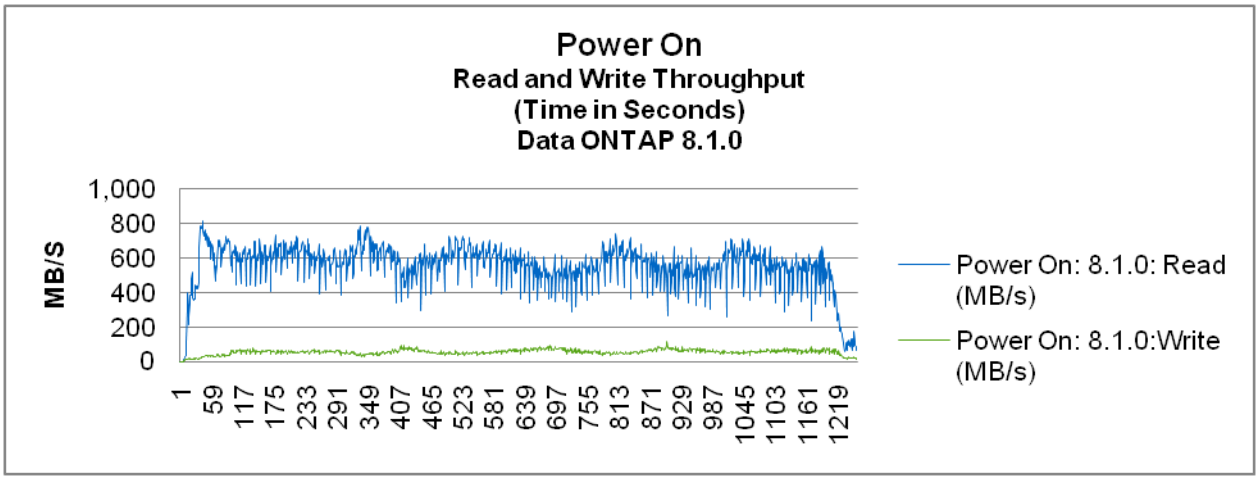
Table 16) Data ONTAP 8.0.1 versus 8.1 during reboot.

Data ONTAP FAS3270	Drive Type	Read Ops/s	Write Ops/s	Read MB/s	Write MB/s	Storage Controller Read Latency	Storage Controller Write Latency	CPU Utilization	Total Time for Power-On
8.0.1	48 4Gb FC15K	19,938	4,916	402MB/s	55MB/s	4ms	3ms	95%	36 minutes
8.1	48 FC15K	36,085	5,432	569MB/s	56MB/s	2ms	1ms	97%	21 minutes

From this point on in this section, graphs detailing the Data ONTAP 8.1 configuration are displayed for brevity. We observed similar curves and metrics for 8.0 but with lower throughput as described in Table 16.

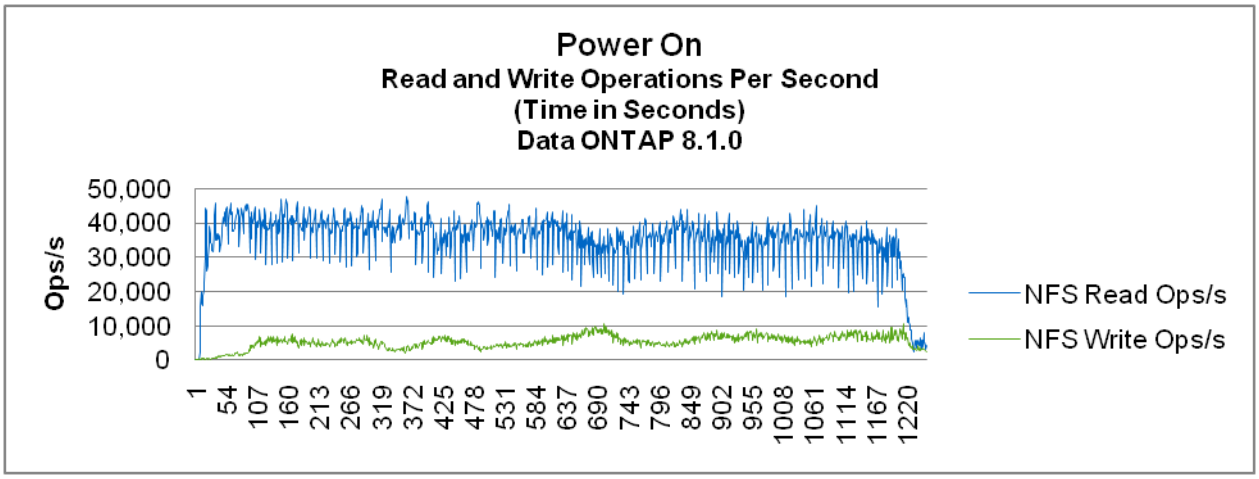
The graph in Figure 26 shows the read and write throughput generated by the power-on of the 2,500 virtual desktops. Reads were responsible for 91% (avg. 569MB per second) of data passed over the network, and writes for the remaining 9% (avg. 56MB per second).

Figure 26) Read and write throughput at power-on.



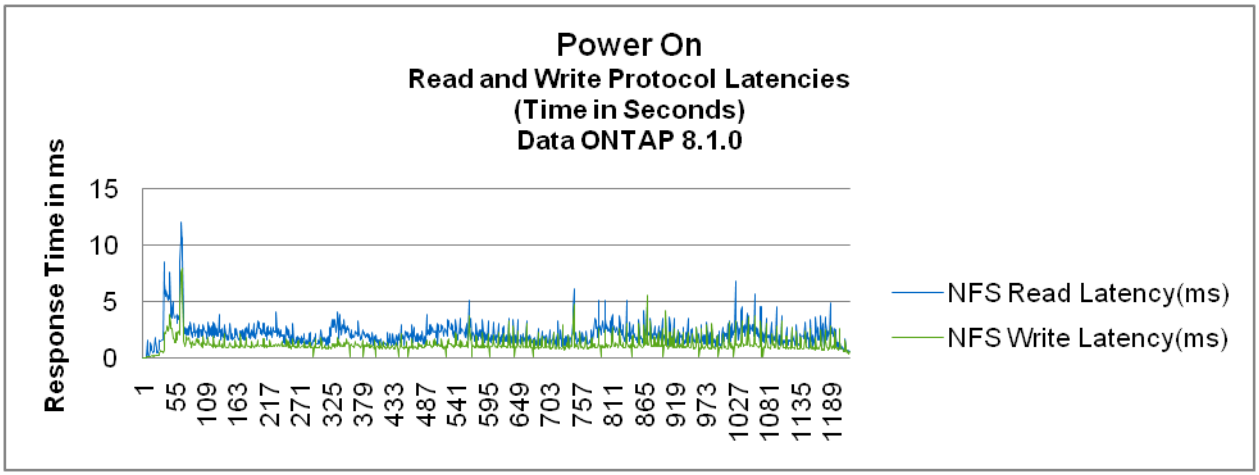
The graph in Figure 27 shows the read/write operations generated per second by the power-on of the 2,500 virtual desktops. The read operations accounted for 80% of the NFS workload, and the write operations for 10%. Lookups (not displayed) accounted for the remaining approximately 10%. Users began working as soon as they logged in, so the application loads overlap with logins.

Figure 27) Read and write operations per second at power-on.



The graph in Figure 28 shows the latencies as reported by the storage controller for the NFS protocol. Both read (avg. 2ms) and write (avg. 1ms) protocol latencies are shown for the entire power-on time. These low latencies were made because the VMs were all clones sharing many of the same blocks on disk, in the flash cache, and as of Data ONTAP 8.1 in RAM with the benefit of virtual storage tiering. These features of cloned VMs make it possible for extremely quick access.

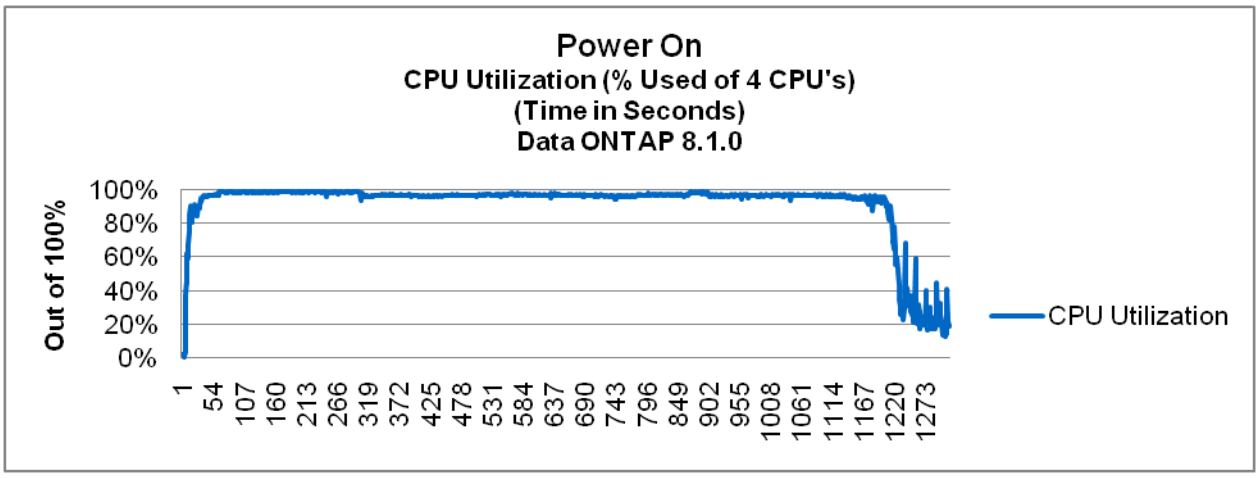
Figure 28) Read and write protocol latencies at power-on.



The power-on operations were not captured with ESXTOP, so no chart is available documenting the power-on workload from the perspective of the virtual desktops. Because virtual desktops first appear in ESXTOP after they begin their power-on operation, and because ESXTOP in batch mode prints a header row only once, we were left with no valid way of capturing the data and making its output meaningful.

During this time, as the graph Figure 29 shows, on average, 97% of the capacity of all four physical CPUs was consumed.

Figure 29) CPU utilization during power-on.



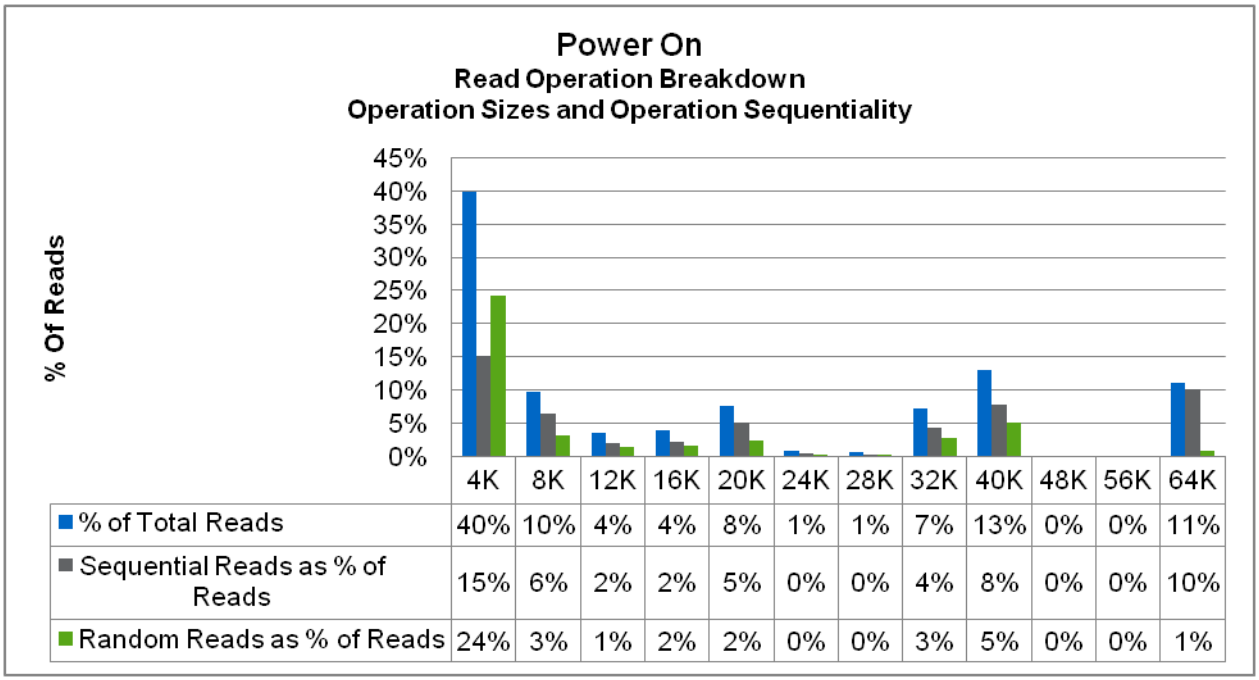
WORKLOAD CHARACTERISTICS

As with the other workloads outlined in this report, the workload characteristics of the power-on scenario also break with tradition in terms of I/O sizes, sequentiality, and concurrency. As defined here, “concurrency” refers to the number of virtual desktops performing power-on operations at once. Keep in mind the current vCenter limit of 128 simultaneous power-on operations, with the rest in queue.

The read operation sizes and their respective natures, sequential or random, are documented in Figure 30. The statistics themselves were taken from counter manager read-ahead statistics captured on the storage controller during the test.

Key Points
• The workload is not all one size: – The graph in Figure 30 is broken down into operation buckets. Each bucket contains all operations, from the size specified down to the next reported operation size. • The workload is not all random: 50% of all reads are sequential.

Figure 30) Read operation breakdown at power-on.



	4K	8K	12K	16K	20K	24K	28K	32K	40K	48K	56K	64K
% of Total Reads	40%	10%	4%	4%	8%	1%	1%	7%	13%	0%	0%	11%
Sequential Reads as % of Reads	15%	6%	2%	2%	5%	0%	0%	4%	8%	0%	0%	10%
Random Reads as % of Reads	24%	3%	1%	2%	2%	0%	0%	3%	5%	0%	0%	1%

Although ESXTOP may not be used for this particular workload, the following information may be gleaned nonetheless:

- The average read and write operations per second per desktop, considering that 128 virtual desktops underwent power-on at a time. This value fails to consider the operations generated by virtual desktops post-power on, but roughly serves our purposes:
  - Total operations per virtual desktop per second (for the 128): 323 ops/sec (reads: 281 ops/sec, writes: 42 ops/sec)
- Ignoring the 128 simultaneous power-on limit imposed by vCenter, if we were to divide the number of operations/sec across all 2,500 virtual desktops, the number of ops/sec per desktop is calculated in this way:
  - Total operations per virtual desktop per second (for the 128): 16 ops/sec (reads: 14 ops/sec, writes: 2 ops/sec)
  - This is not safe to do, however. As the number of virtual desktops in the environment decreases, this method of calculation results in an increased number of operations per second. The inverse is true as the number of virtual desktops decreases.

5.5 MONDAY MORNING LOGIN

Sun	Mon	Tue	Wed	Thurs	Fri	Sat
29 Deploy 2,500 desktops	30 8 a.m. 2,500 logins + Profile load (30 min) Typical workday	31 8 a.m. 2,500 logins (30 min) Typical workday	1	2	3	4
5	6 1 a.m. Network maintenance 2 a.m. Network outage 7 a.m. Reboot all VMs 8 a.m. 2,500 logins (30 min) (Post-power-on)	7 8 a.m. 2,500 logins (30 min) Typical workday	8	9	10	11

THE STORY

The power rebooting of the 2,500 virtual desktops cleared the contents of memory from each machine. Because of this, each user's profile required storage I/O, the same as when any of the users opened their applications for the first time in the day. As a result, the "Monday morning" login generated storage workload in excess of the "Tuesday morning" or typical login scenario but less than the initial login for which the profile had to be created and not just read from disk.

THE FACTS

As in all of the login and workload scenarios captured in this report, the user login rate was deterministic at approximately 3 logins per second, with the final login being attempted 26 minutes after the first.

At login on this Monday morning, all users began their day's work. The VMware desktop RAWC controlled all of the workload in the environment, opening applications and standing in for human control at the keyboard. RAWC performed the following tasks on each virtual desktop:

- Opened Microsoft Word and Excel for each user, creating and saving new documents in each
- Opened Microsoft PowerPoint as well as Adobe Acrobat Reader and reviewed existing documents in each
- Opened Microsoft Internet Explorer
- Wrote and sent three e-mails from Outlook client

RAWC randomly determined the order in which the applications were run on each virtual desktop.

USER EXPERIENCE

Table 17 reports the login times experienced by the user community on Monday morning during the post-power-on login scenario.

Key Point
The average user experienced 1-second logins on Monday morning, regardless of the Data ONTAP version used.

Table 17) User experience of Monday morning login (in seconds).

Configuration	Time to Log In: Average User Experience	Time to Log In: Maximum User Experience	Time to Log In: Standard Deviation for User Experiences
Data ONTAP 8.0.1	1 second	5 seconds	0.3 seconds
Data ONTAP 8.1	1 second	2 seconds	0.3 seconds

Note: FAS3270 with 48 450GB FC 15K, loop rate: 4Gbps.

According to Liquidware Labs Stratusphere UX, 100% of the users experienced a good login experience, with "good" defined as taking less than 15 seconds. Table 18 shows the user experiences for the Monday morning login.

Table 18) User experience of Monday morning login (in percentages of good, fair, and poor login time).

Configuration	Total Number of Users	% Users with Good Login Time (<= 15 sec)	% Users with Fair Login Time (<= 60 sec)	% Users with Poor Login Time (=> 60 sec)
Data ONTAP 8.0.1	2,500	100% (1 sec. avg.)	0%	0%
Data ONTAP 8.1	2,500	100% (1 sec. avg.)	0%	0%

Note: FAS3270 with 48 450GB FC 15K, loop rate: 4Gbps.

SYSTEM EXPERIENCE

The storage controller behaved on a high level as described in Table 19. Notice that the amount of work passed between the virtual desktops and the storage controller was read heavy in terms of both MB/s and IOPS. Overall, the reads accounted for 80% of the data and 60% of the IOPS passed to the storage controller.

Key Point
At the protocol level, the storage controller responded in less than 1ms on average to both read and write requests.

For further details, Table 19 compares Data ONTAP 8.0.1 and 8.1, breaking out throughput, operations per second, and latencies.

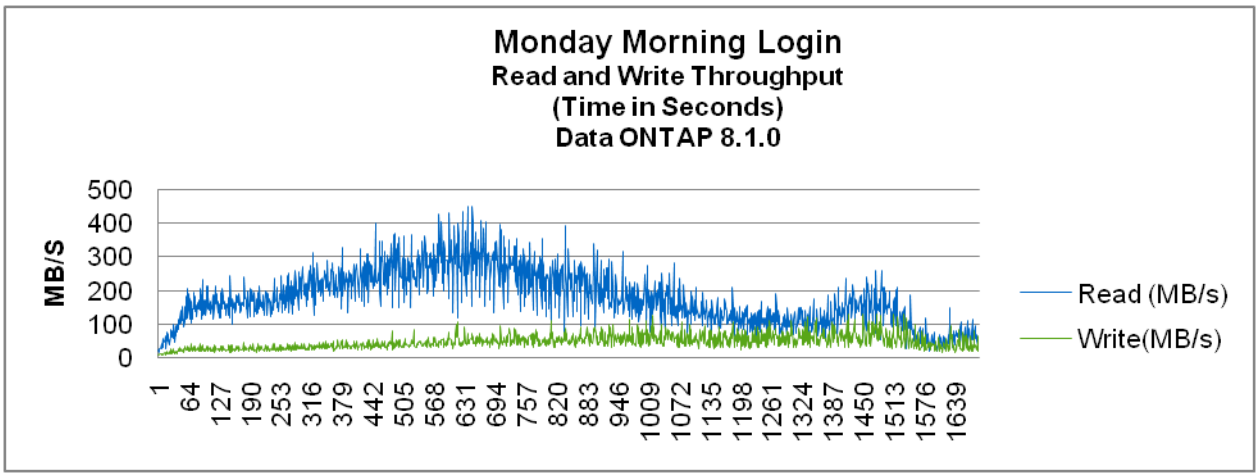
Table 19) Data ONTAP 8.0.1 versus 8.1 during Monday morning login.

Data ONTAP FAS3270	Read Ops/s	Write Ops/s	Read MB/s	Write MB/s	Read Latency		Write Latency		CPU Utilization
					Controller	ESX	Controller	ESX	
Data ONTAP 8.0.1	7,575	5,615	130MB/s	40MB/s	1.2ms	2.6	0.9ms	2.1	71%
Data ONTAP 8.1	9,782	5,978	180MB/s	49MB/s	0.4ms	2.2	0.5ms	2.1	73%

From this point on in this section, graphs detailing the Data ONTAP 8.1 configuration are displayed for brevity. We observed similar curves and metrics for 8.0 but with lower throughput, as described in Table 19.

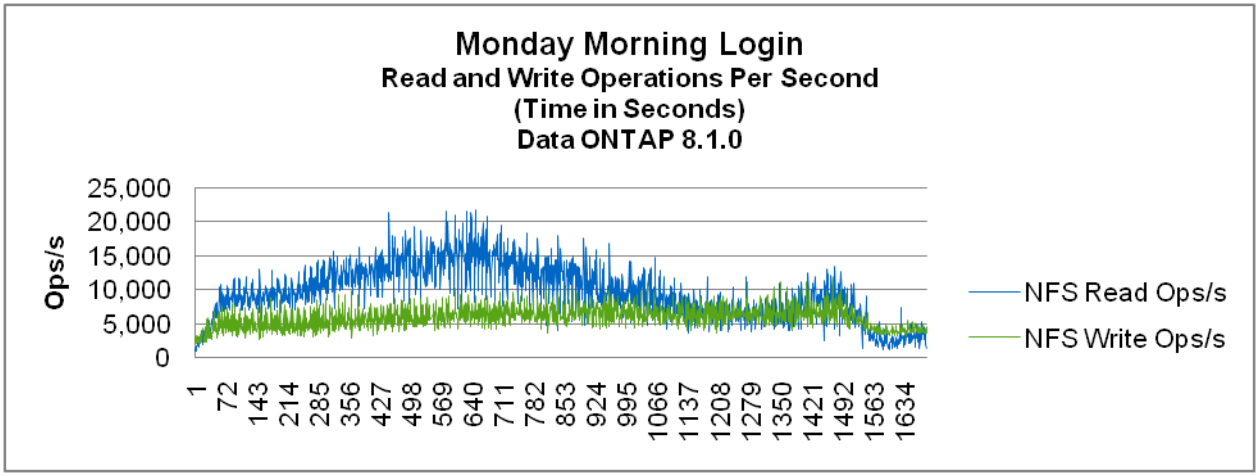
The graph in Figure 31 shows the read and write throughput generated by both the Monday morning logins of 2,500 users and their subsequent start-of-day workload. Users began work as soon as they logged in, so the application loads overlap the logins and the profile loads. Reads were responsible for 79% (avg. 180MB per second) of data passed over the network, writes for the remaining 20% (avg. 49MB per second).

Figure 31) Read and write throughput during Monday morning login.



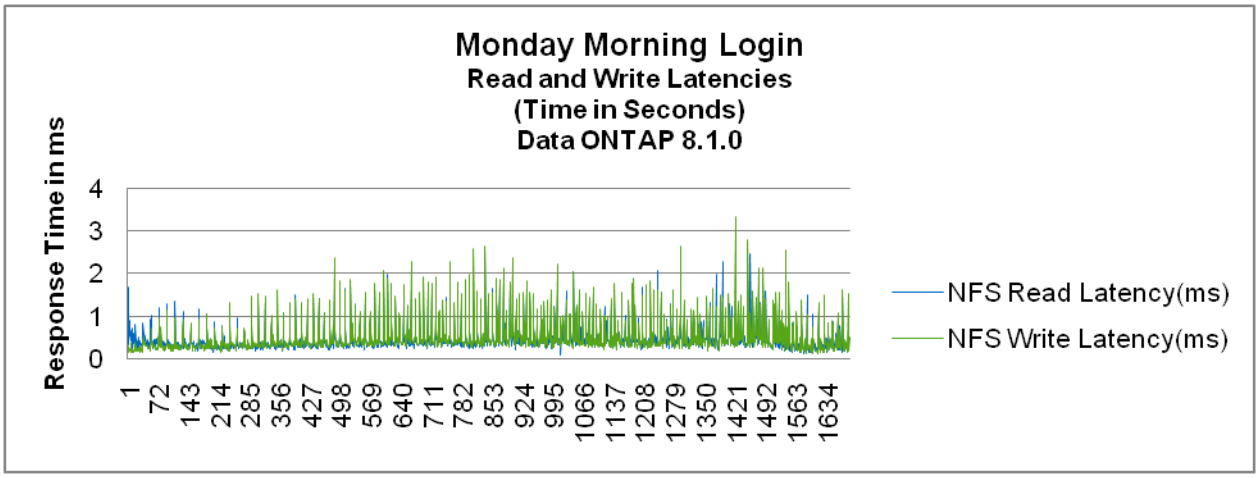
The graph in Figure 32 shows the read/write operations generated per second by the power-on of the 2,500 virtual desktops. The read operations accounted for 62% of the NFS workload, and the write operations for 37%. Lookups (not displayed) accounted for the remaining approximately 1%. As soon as users logged in, they began working; therefore, the user work and the background work overlap with logins.

Figure 32) Read and write operations per second during Monday morning login.



The graph in Figure 33 shows the latencies as reported by the storage controller for the NFS protocol. Both read (avg. 0.5ms) and write (avg. 0.5ms) protocol latencies are shown for the entire power-on time.

Figure 33) Read and write latencies (in seconds) for Monday morning login.

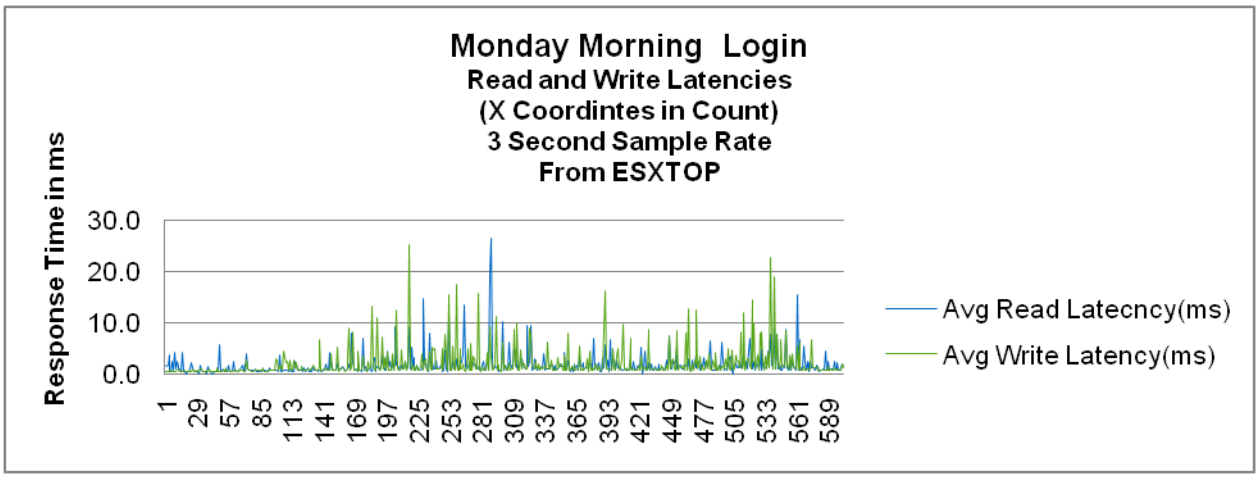


The graph in Figure 34 shows guest latencies as reported by ESXTOP batch mode and compiled by ESXTOP. ESXTOP batch mode has a 3-second minimum sample rate. The X-axis is the sample number. Because this graph represents 3-second sample rates, multiply the X-axis value by 3 to get the true run time. Notice that although there a few outliers, and they are well within the limits of "good" latency as defined by Liquidware Labs Stratusphere, these outliers are never sustained across sample intervals.

The average latency in this graph are:

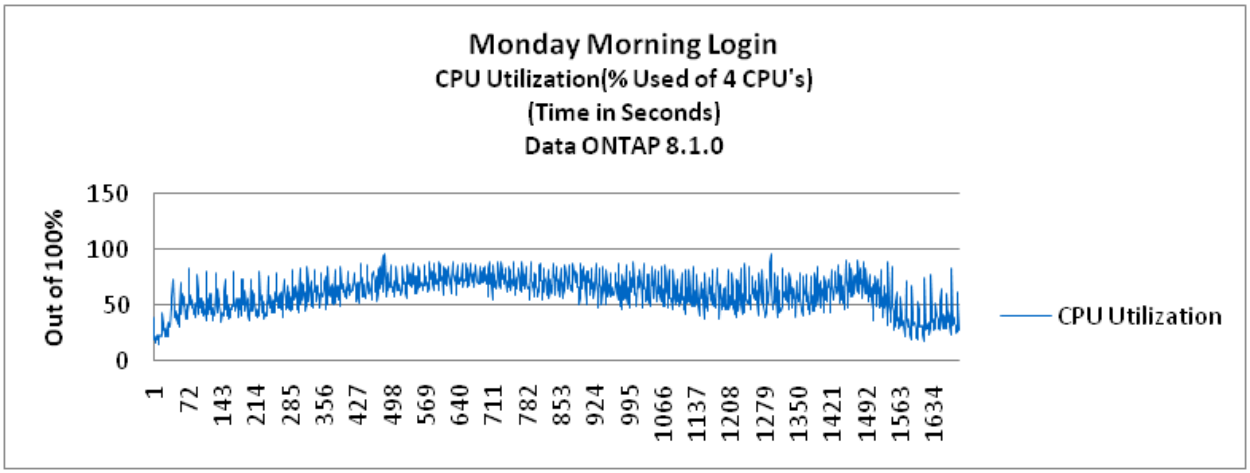
- Reads: 1.7ms
- Writes: 1.9ms

Figure 34) Guest read and write latencies for Monday morning login.



During this time, as the graph in Figure 35 shows, on average 61% of all four physical CPUs were consumed.

Figure 35) CPU utilization for Monday morning login.



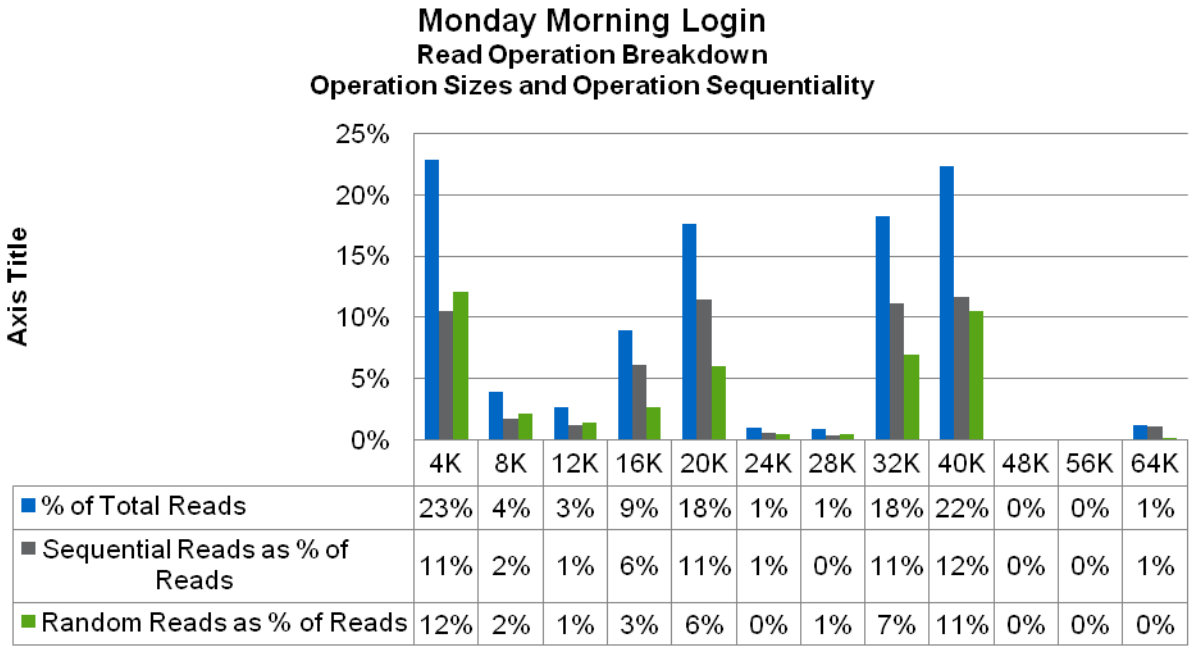
WORKLOAD CHARACTERISTICS

The Monday morning login scenario also shows a wide spread of operation sizes and contains a balanced amount of randomness/sequentiality in its I/O. Recall that "concurrency," as used here, refers to the number of virtual desktops generating storage-targeted I/O at the same time.

The read operation sizes and their respective natures, sequential or random, are documented in Figure 36. The statistics themselves were taken from counter manager read-ahead statistics captured on the storage controller during the test.

Key Points
• The workload is not all one size: – The graph in Figure 36 is broken down into operation buckets. Each bucket contains all operations, from the size specified down to the next reported operation size. • The workload is fairly balanced between random and sequential operations: 56% of all reads are sequential.

Figure 36) Read operation breakdown for Monday morning login.



	4K	8K	12K	16K	20K	24K	28K	32K	40K	48K	56K	64K
% of Total Reads	23%	4%	3%	9%	18%	1%	1%	18%	22%	0%	0%	1%
Sequential Reads as % of Reads	11%	2%	1%	6%	11%	1%	0%	11%	12%	0%	0%	1%
Random Reads as % of Reads	12%	2%	1%	3%	6%	0%	1%	7%	11%	0%	0%	0%

Table 20 documents the workload from the perspective of the virtual desktops themselves as captured in ESXTOP. During our testing, ESXTOP ran continuously in batch mode on all servers in the environment with the lowest possible sample interval of 3 seconds

In Table 20, the charts generated by ESXTOP document the concurrency, I/O rate, and operation size from the perspective of the virtual desktops as reported by ESXTOP.

Key Points

- On average, 11% of the virtual desktops generated reads, and 66% generated writes during any given second. Thus, 100% concurrency was not achieved.
- On average, if concurrency is ignored, each virtual desktop generates 7 IOPS during the Monday morning login scenario.
- ESXTOP confirms the data from the storage controller that the IOPS are larger the 4KB or 8KB.
- The average IOPS and throughput reported from ESXTOP closely approximate the values reported by the storage controller.

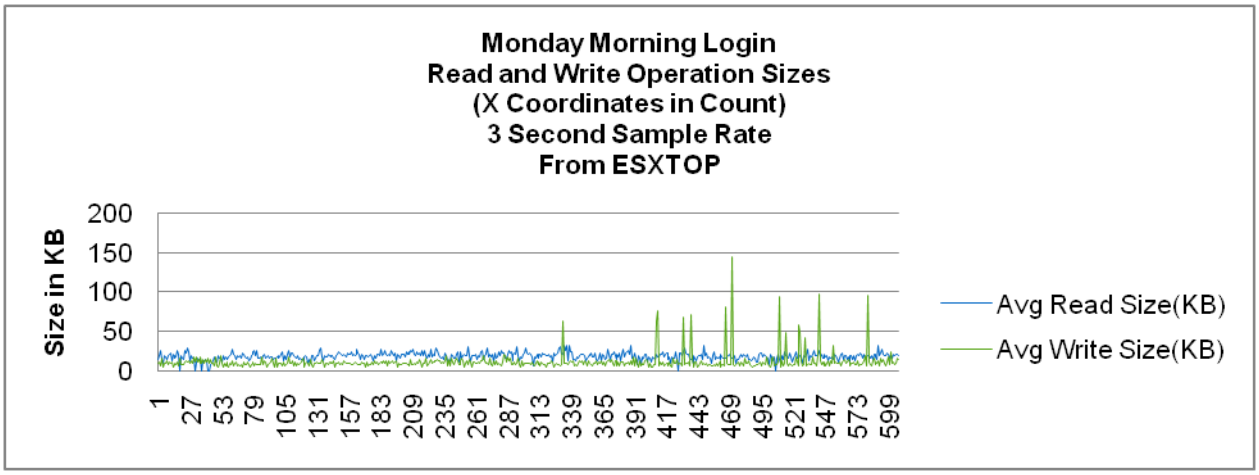
Table 20) I/O concurrency, rate, and size for read and write operations at Monday morning login.

Subjects Measured	Values	
Total VM count	2,500	
Avg total IOPS	16,118	
Avg read IOPS	10,621	
Avg write IOPS	5,497	
Avg read throughput (MB/sec)	192	
Avg write throughput (MB/sec)	55.9	
# of reading VMs	283	
# of writing VMs	1,662	
Read size avg (KB)	19	
Write size avg (KB)	10	
Read latency avg (ms)	2.2	
Write latency avg (ms)	2	
	For One VM	For All VMs
Avg read IOPS per reading VM	37	4.2
Avg write IOPS per writing VM	3	2.2
Avg read throughput per reading VM (KB/sec)	696	79
Avg write throughput per writing VM (KB/sec)	34	23

In Figure 37, the graph generated by ESXTOP also shows the average read and write operation sizes as issued by the guests themselves throughout the entire login scenario.

Key Point
The reads and writes fluctuate widely from sub-4K to greater than 100KB.

Figure 37) Read and write operation sizes for Monday morning login.



The workload characteristics of the Monday morning login scenario are thus neither 4KB nor 8KB but a mixture of all sizes. On average, the read and write sizes are 19KB and 12KB.

5.6 STEADY STATE

Sun	Mon	Tue	Wed	Thurs	Fri	Sat
29 Deploy 2,500 desktops	30 8 a.m. 2,500 logins + profile load (30 min) Typical workday	31 8 a.m. 2,500 logins (30 min) Typical workday	1	2	3	4
5	6 1 a.m. Network maintenance 2 a.m. Network outage 7 a.m. Reboot all VMs 8 a.m. 2,500 logins (30 min) (Post-power-on)	7 8 a.m. 2,500 logins (30 min) Typical workday	8	9	10	11

THE STORY

On any given day, all Acme Corporation employees have arrived at work by 8:30 a.m. and have begun their day's work. "Steady state" is defined here as the situation after most (or all) of the users have logged in and have already opened their applications and begun work. Users may continue to trickle in, but the expected time of logins has passed. Equally, during steady state, users may be expected to open applications not previously loaded on their VMs; however, they are not expected to do this all at the same time or in large numbers. Steady state is thus the time of day where bulk changes in workload are no longer expected.

THE FACTS

Each of our tests ran multiple RAWC iterations. The first iteration was run by RAWC soon after login. Each Iteration ran through each application in the work list, opening an application, doing work, and then moving on to the next application.

Note: For the first time the VM is logged into, either since creation or since the last power-on, the opening of each application requires the loading of application libraries from disk to memory. This differentiates iteration 1 from other iterations. If the applications have been run previously and the libraries have remained in VM memory, then the first iteration does not differ from the rest.

Our testing arrived at steady state after the final user had completed RAWC iteration 1 and all users were in their second or a later iteration.

The following applications were running during steady state:

- Microsoft Word, and Excel, creating and saving new documents in each.
- Microsoft PowerPoint as well as Adobe Acrobat Reader: documents were being reviewed in each; these documents had been reviewed in earlier iterations and were in memory on access.
- Internet Explorer
- Outlook client, sending three e-mails

In this test scenario, all logins have occurred previously, so user login times are not documented. Liquidware Labs has reported all user experience as being "good" during this time, regardless of which Data ONTAP version was used. The characteristics of this workload are very similar to those of the Tuesday morning login and workload scenario. Although the Tuesday morning login scenario included

2,500 logins in the workload, each user's profile had been loaded into memory before the start of the day. Because the user experience was entirely "good," according to Liquidware Labs, and because the steady state closely resembles the Tuesday morning login and work scenario, the charts in section 5.3, "Tuesday Morning Login," are relevant for the steady-state user experience, storage system load, and workload characteristics. The only characteristic that is not relevant is the login times because all users are logged in already.

5.7 OBSERVATIONS AND LESSONS LEARNED

MEMORY OVERSUBSCRIPTION

The 50,000-Seat VMware View Deployment white paper demonstrated that 600 ESX servers were used to support 50,000 Windows 7 VMs. Each Windows 7 VM was given 1.5GB of RAM, and each ESX server had 48GB of RAM. Although this was acceptable for demonstration purposes, when testing was done under a heavier load, this server count caused excessive VMware ballooning and negatively affected the storage. A safe oversubscription ratio in our testing was 1.5:1 machine to physical memory. We expanded our server environment to stay within the 1.5:1 memory-oversubscription threshold as physical memory consumed began to approach capacity. At 90 servers per 5,000 VMs, we achieved an oversubscription ratio of 1.2:1. Ballooning dropped to an acceptable range and storage performance returned to normal. For more on memory oversubscription, refer to the VMware paper Understanding Memory Resource Management in VMware ESX 4.1.

EVER-PRESENT ACTIVITY (THE DRIPPING WATER EFFECT)

We have pointed out that VMs generate work distinct from the user and distinct from scheduled tasks. Recall that we optimized our template VM, following the VMware View Administrator's Guide recommendations, as well as running the VMware Windows 7 optimization script found in the VMware View Optimization Guide for Windows 7. We went through the scheduled tasks and disabled remaining scheduled tasks that might otherwise run during our tests. Our goal was not just to have a VM environment optimized for maximum efficiency but to have an environment that would run the same way every time without unexpected workloads contaminating the tests.

These optimizations did not eliminate what can be called a "dripping water" workload. Dripping water is an appropriate name because each VM regularly generates a few write operations per second and every now and then a few read operations as well. This predominately disk-write workload was found to come from system, and to a lesser extent from `svchost.exe` and `services.exe` within each Windows 7 VM. On average in our environment, each desktop generated just less than 1 operation per second, every second. The write operations are approximately 4KB, as demonstrated by the two graphs in Figure 38 and Figure 39.

Figure 38) Dripping water workload (operations per second).

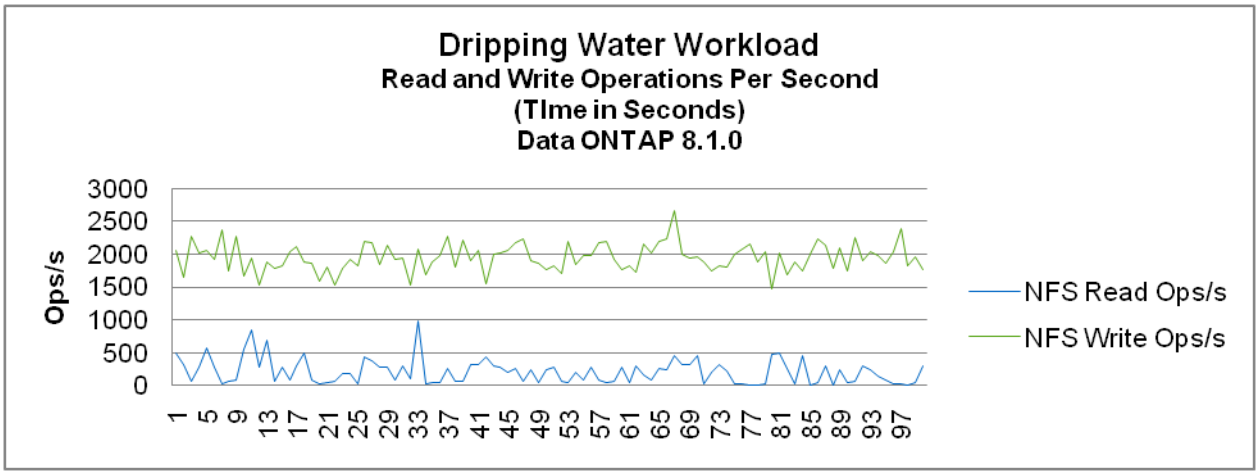
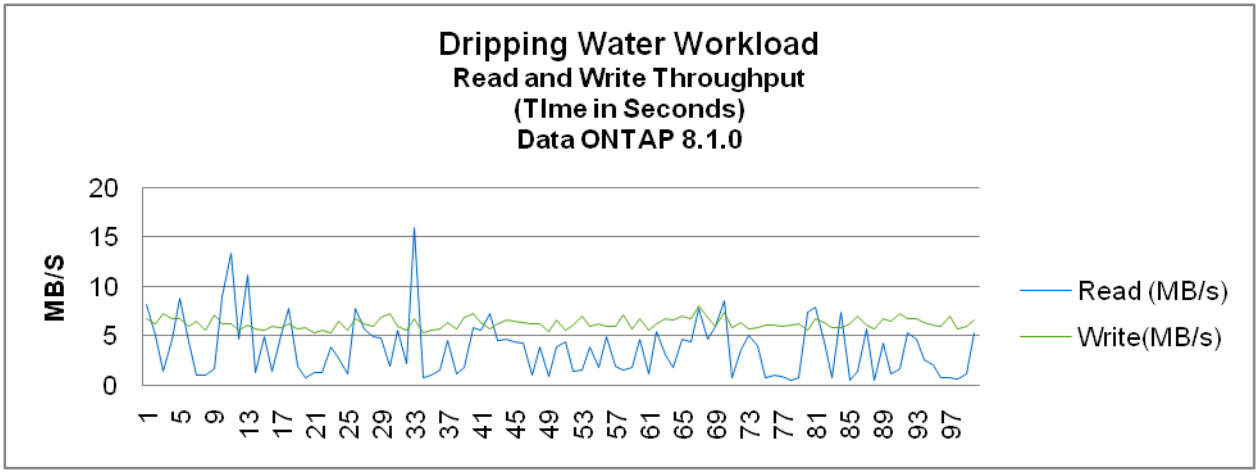
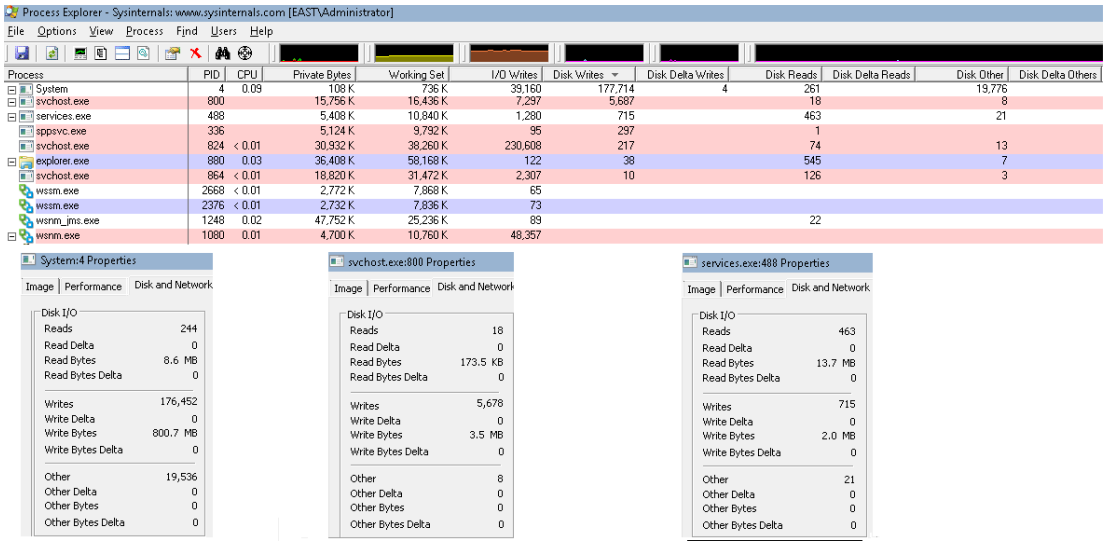


Figure 39) Dripping water workload (MB per second).



The screenshot in Figure 40 was captured from one of the VMs used to isolate the source of the “dripping water” workload. This is a picture of a process explorer that had been left running for just over two days. Notice that the system generated most of the writes. Notice also the delta of four writes observed in between screen refresh rate, which was set to every 1 second. The value is not always four; for example, it ranged from zero to seven in this case. This demonstrates that the average write ops/sec per VM is a normalized value. When divided against the total number of VMs, the average comes out to approximately 1 operation per VM.

Figure 40) Screen showing "dripping water" workload.



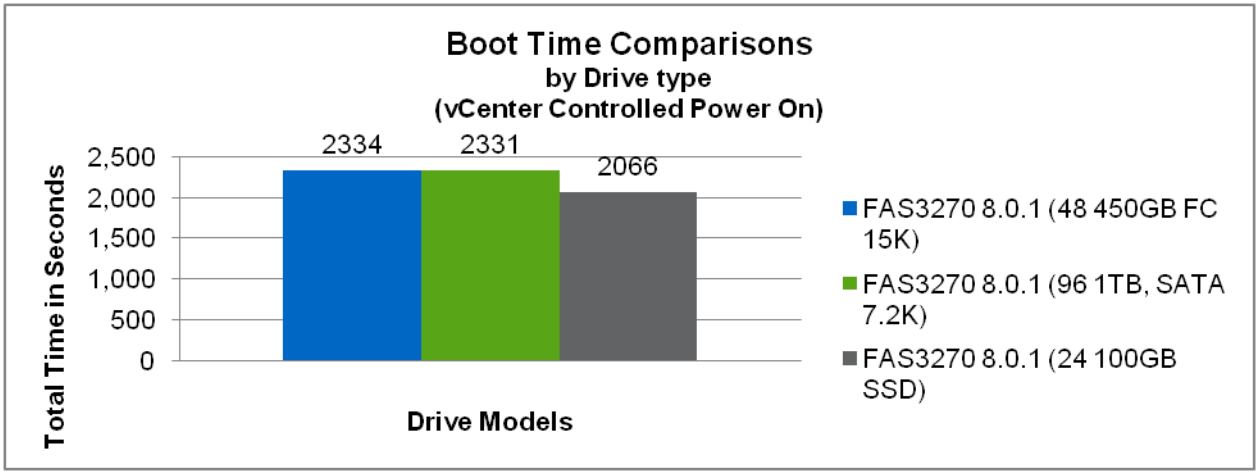
6 APPENDIXES

6.1 SSD AND SATA

The graphs in Figure 41 through Figure 43 compare the capabilities of SSD and SATA with the 15K drives referenced throughout this report. This testing was performed while the storage controller was running Data ONTAP 8.0.1, which was fairly early in the testing. The tests included power-on and isolated login tests of the full 2,500 VM environment. These login tests were limited in that user workload was not generated postlogin. The SATA configuration was made up of 96 7.2K 1TB drives, and the SSD configuration was made up of 24 100GB drives. We conducted these tests to help customers better understand what drive types (given performance, capacity, and price) are best for VDI workloads. We sized these configurations in such a manner that number of drives required would satisfy the workload.

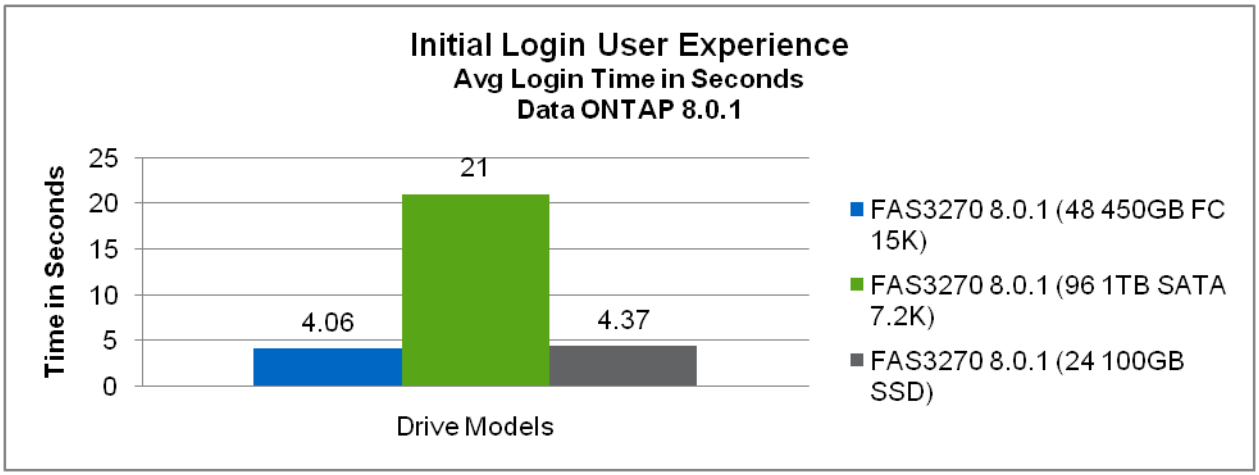
This first graphs show that the boot times for 96 SATA drives roughly matched the boot times encountered with 48 FC drives, at twice the drive count. The 24 SSD drives resulted in a boot time performance reduction of 4.5 minutes, but at an increased cost over the 48 FC drives.

Figure 41) Boot time comparisons by drive type.



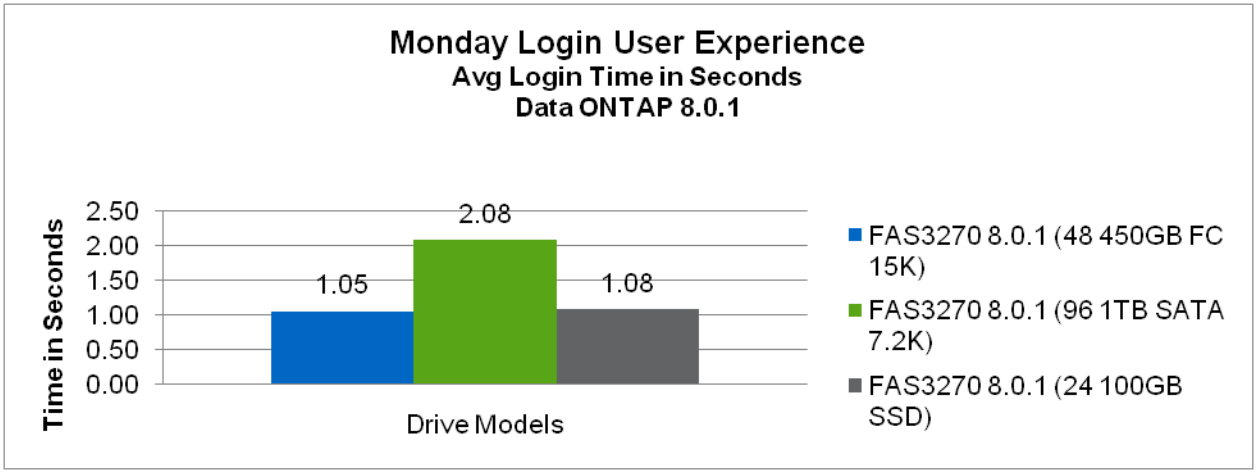
The graph in Figure 42 shows an isolated first-login test. 2,500 users logged in for the first time, triggering profile creation, but no RAWC workload was launched. In this test, SATA drives were at a disadvantage compared to the other two drive types. The SATA drives were affected by write performance, an experience not encountered with the other two drives.

Figure 42) User experience at initial login with SATA drives.



The graph in Figure 43 shows an isolated Monday morning login scenario in which 2,500 users logged into VMs that had previously been accessed but had since been rebooted. No RAWC workload was generated upon login. SATA drives were at a disadvantage again, compared to the other two drive types. The response time for the SATA configuration was slightly higher for both read and write operations compared to either SSD or FC.

Figure 43) User experience on Monday morning with SATA drives.



As the graphs show, our testing led us to determine that the 15KB drive performed the best for the money spent. The SATA drives did not perform as well in write scenarios as the FC and SSD. Additionally, at twice the drive count of FC, the price and environmental considerations weighed against the deployment of SATA. Further, the extra space available with SATA is unnecessary because of the NetApp deduplication/thin provisioning/cloning technologies, otherwise known as storage efficiency. As for the SSD drives, although only half as many spindles were necessary compared to the 15K drives, this quantity was cost prohibitive by comparison. The SSDs were able to complete the boot times more quickly than either FC or SATA, but the advantage was lost during the login phases.

6.2 APPLICATION WORKLOADS

This appendix section describes the interaction of IOPS and throughput in a virtual desktop environment. Operations alone do not adequately describe a workload.

The applications and their workloads were selected for this section because each demonstrates a unique point. Among the applications, the Microsoft Word workload dealt with the opening of the application without opening any files; this resulted in 30MB of data read the first time the application was opened. Windows Media Player was used to stream a movie, and it demonstrated large read operations. The Microsoft Excel workload dealt with reading a large file, searching the file, then saving and closing the file. This workload demonstrates that virtual desktops can generate large write operations.

This section also demonstrates how the workload generated by each application differs between first and subsequent invocations.

MICROSOFT WORD

This set of charts documents the workload encountered while the user opened and then closed Microsoft Word. No file was created after Word was opened. This scenario demonstrates the workload generated by the opening of the application in isolation from interacting with files. The first two charts show the workload generated the first time Word was opened since reboot. The second set of charts shows what happens when the user repeats the steps subsequently. The workload differs because the application's libraries stayed in the guest's cache, resulting in reduced overall usage of storage resources.

The first run resulted in far more work (throughput and operations) because the application libraries were loaded here. The charts are histograms: Each vertical bar represents a different operation size, and the height of the bar represents how many operations of that type were performed. These graphs were created from the output of vscsiStats captured on the ESX server.

The following breakdown is a high-level summary of the workloads documented in the graphs:

Run 1

- Data read: 29MB (99% of throughput)
- Data written: 0.3MB (1% of throughput)
- Read operations: 1,471 (96% of ops)
- Write operations: 64 (4% of ops)
- Average read size: 21KB
- Average write size: 5KB

Run 2

- Data read: 0.84MB (84% of throughput)
- Data written: 0.17MB (16% of throughput)
- Read operations: 34 (57% of ops)
- Write operations: 26 (43% of ops)
- Average read size: 26KB
- Average write size: 7KB

Notice that the total operations required by each run decreased between the two runs by 97%. Total data required from storage between the two runs decreased by 97% as well. These changes in required I/O were because the application's libraries had already been loaded into memory by the first run.

The graphs in Figure 44 and Figure 45 describe the workload generated by the initial opening and use of Word.

Figure 44) Read operations for first opening and closing of Microsoft Word.

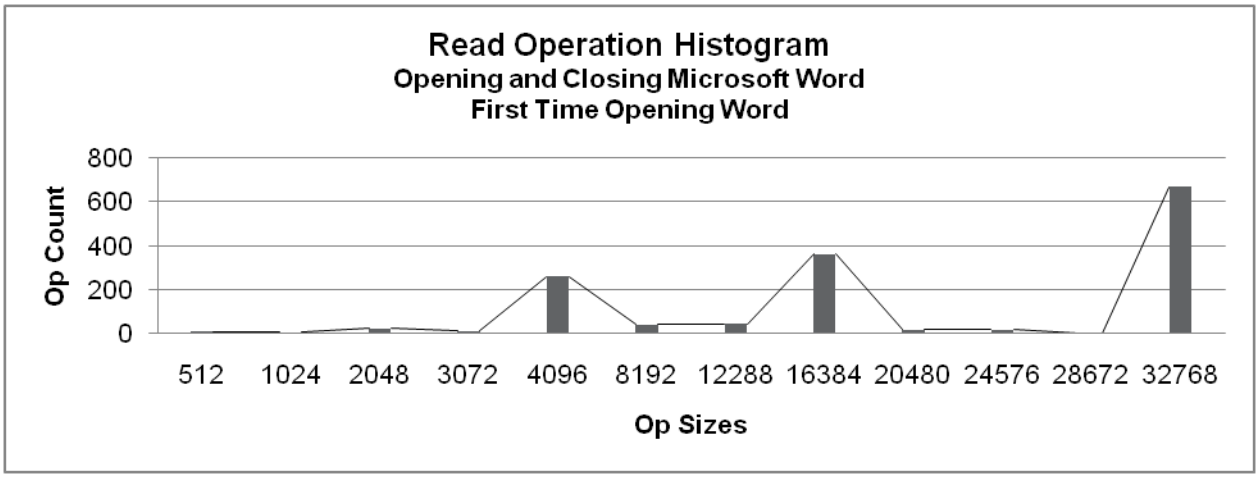
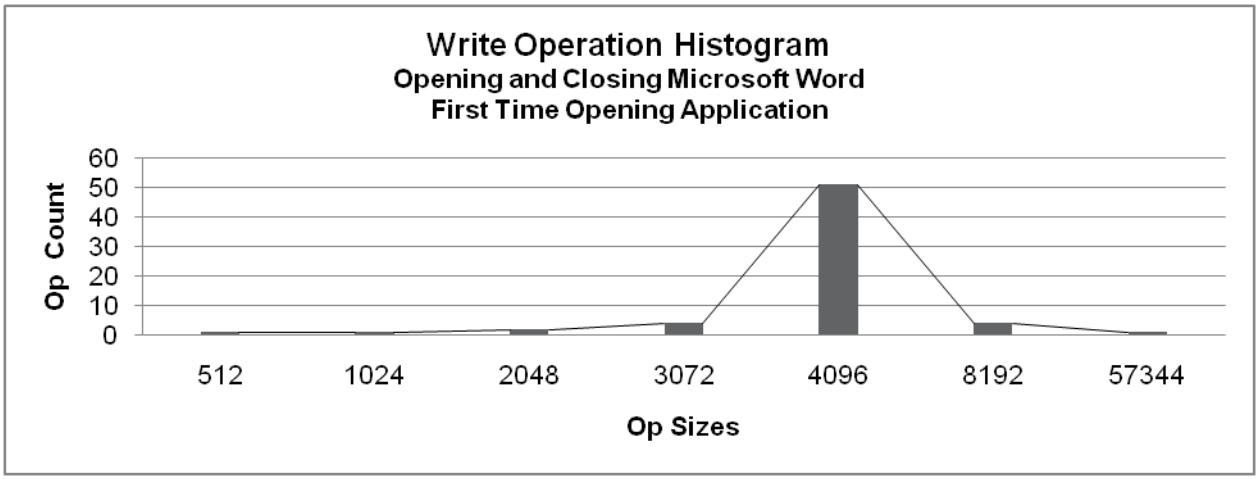


Figure 45) Write operations for first opening and closing of Microsoft Word.



The charts in Figure 46 and Figure 47 show what happens when the user repeats these same steps (that is, opening and closing Microsoft Word). Notice the marked decrease in operations and throughput.

Figure 46) Read operations for subsequent opening and closing of Microsoft Word.

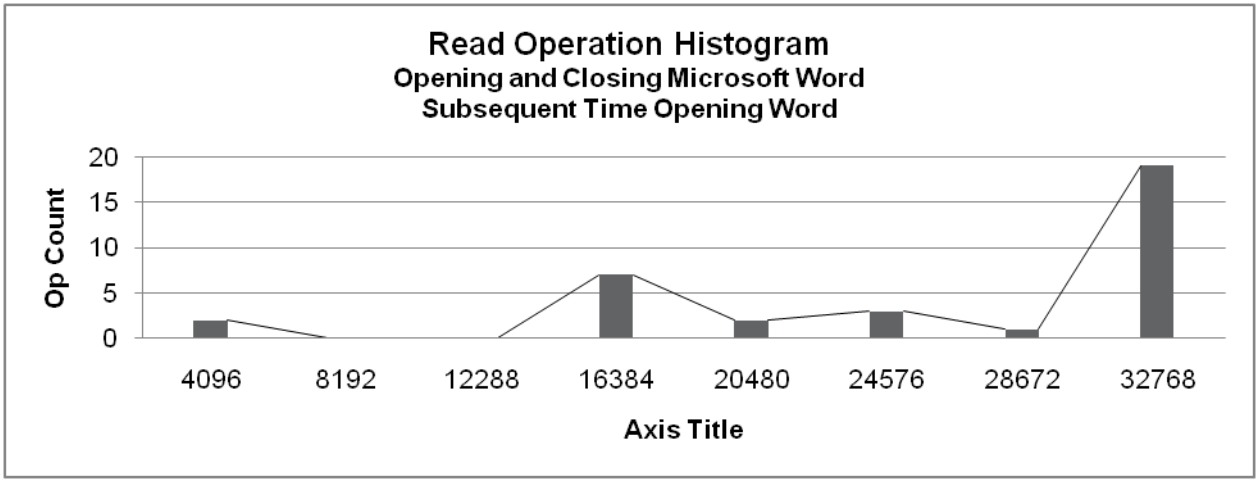
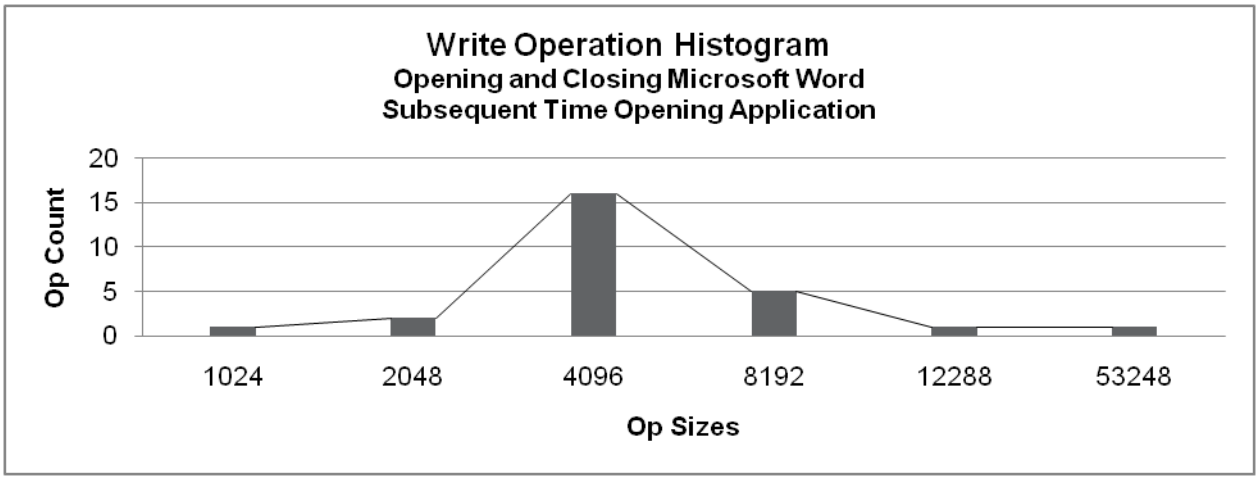


Figure 47) Write operations for subsequent opening and closing of Microsoft Word.



MICROSOFT WINDOWS MEDIA PLAYER

This set of charts documents the workload encountered when the user opened Windows Media Player, played a 134MB movie, and then closed the application. The first two charts show the workload generated the first time the media player was opened since reboot. The second set of charts shows what happens when the user repeats the steps subsequently.

The charts are histograms: Each vertical bar represents a different operation size, and the height of the bar represents how many operations of that type were performed. These graphs were created from the output of vscsiStats captured on the ESX server. Notice the large read operation sizes used when playing the movie.

The following breakdown is a high-level summary of the workloads documented in the following graphs:

Run 1

- Data read: 146MB (99% of throughput)
- Data written: 1MB (1% of throughput)
- Read operations: 3,568 (92% of ops)
- Write operations: 306 (8% of ops)
- Average read size: 42KB
- Average write size: 4KB
- Run time: 142 seconds
- Average ops/sec: 27 ops/sec
- Average throughput: 1,060 KB/sec

Run 2

- Data read: 122MB (99% of throughput)
- Data written: 1MB (1% of throughput)
- Read operations: 1,967 (89% of ops)
- Write operations: 226 (11% of ops)
- Average read size: 64KB
- Average write size: 4KB
- Run time: 141 seconds
- Average ops/sec: 15 ops/sec

- Average throughput: 893KB/sec

Notice that the operations decreased between the two runs by 43%, whereas the throughput decreased by 17%. The decrease in IOPS was due primarily to the application's having been loaded in the first run. The drop in throughput was due in part to partial caching of the movie. Notice that the average read size calculated in the first run was 22KB smaller than the second. This was due to the loading of the libraries.

The graphs in Figure 48 and Figure 49 describe the workload generated by the initial opening and use of Windows Media Player.

Figure 48) Read operations for first opening Windows Media Player and streaming a movie.

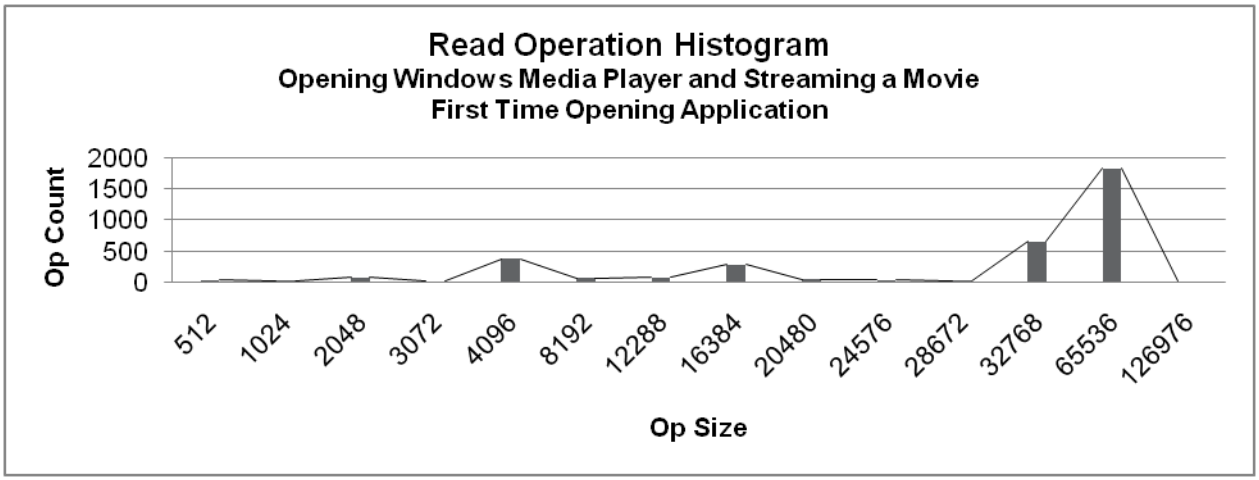
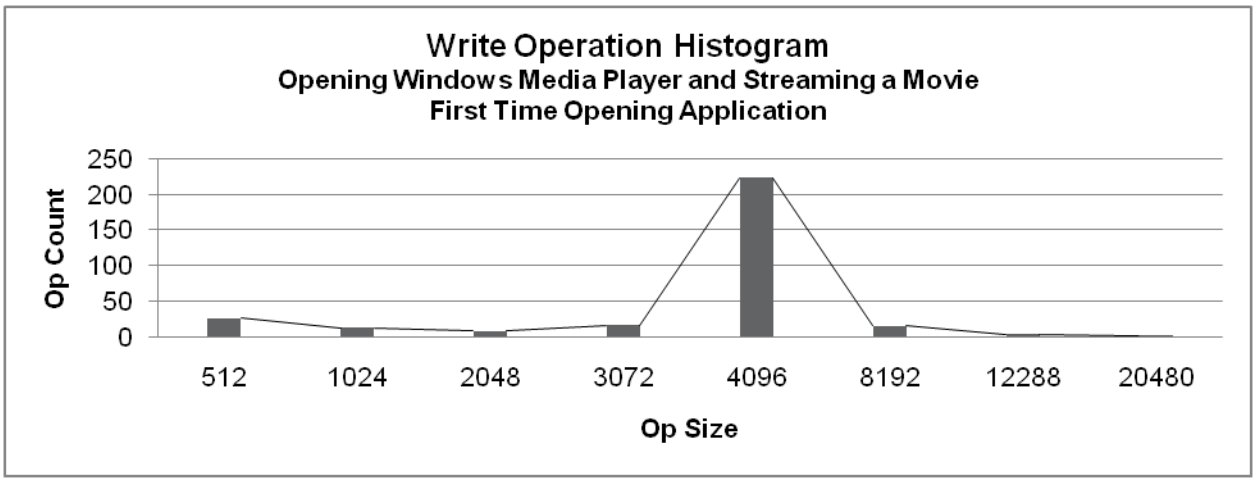


Figure 49) Write operations for first opening Windows Media Player and streaming a movie.



The charts in Figure 50 and Figure 51 show when the user repeats the steps of opening Windows Media Player, loading and playing the 134MB move, and then closing the application. Notice the marked decrease in read operations and the minor decrease in write operations.

Figure 50) Read operations for subsequent time opening Windows Media Player and streaming a movie.

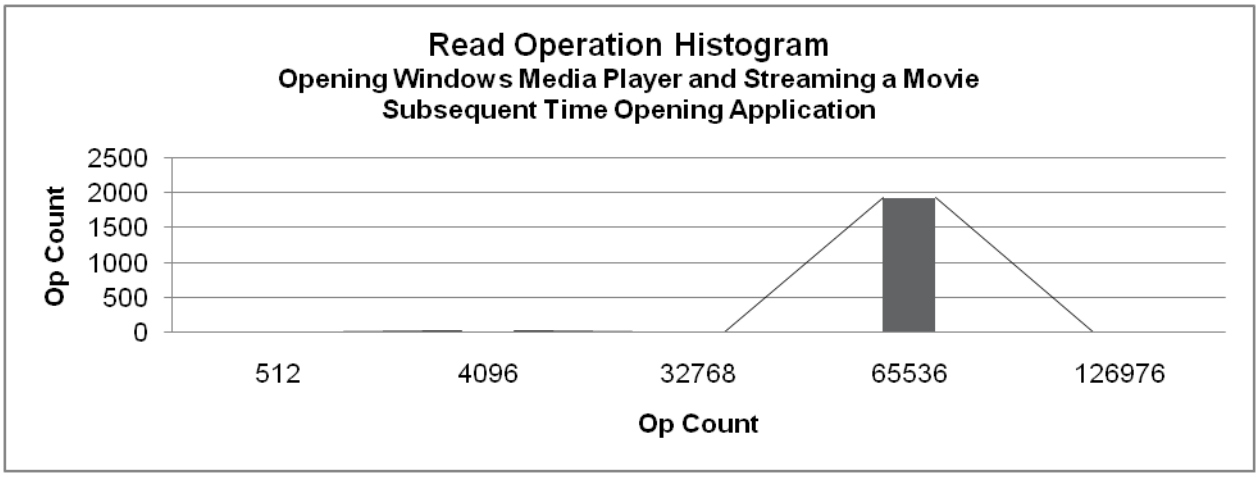
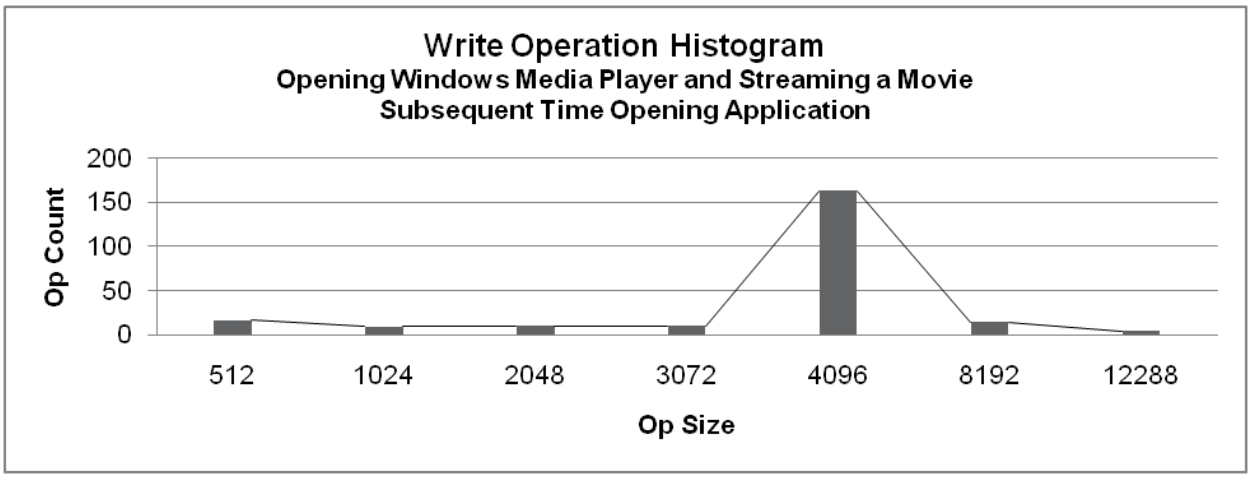


Figure 51) Write operations for subsequent time opening Windows Media Player and streaming a movie.



MICROSOFT EXCEL

The workload documented here focuses on the saving of a 125MB Excel workbook. The intent of this section is to demonstrate that virtual desktops are capable of generating large write operations. The previous workloads focused on read-centric operations, for example, opening an application as well as playing a movie. The 4KB write operations in these workloads were primarily the effect of background processes running on the guest, as mentioned in section 5.7, “Observations and Lessons Learned.”

The workload documented in the charts in Figure 52 and Figure 53 excludes the opening of the application and the loading of the file. This workload focuses exclusively on the operations of saving and closing the application.

The following breakdown is a high-level summary of the workloads documented in the graphs:

- Data read: 3MB (99% of throughput)
- Data written: 208MB (1% of throughput)
- Read operations: 142 (25% of ops)
- Write operations: 681 (82% of ops)
- Average read size: 24KB

- Average write size: 313KB
- Run time: 73 seconds
- Average ops/sec: 11 ops/sec
- Average throughput: 3000KB/sec

The most important point here is that write operations vary across many buckets, ranging from the traditionally seen 4KB all the way out to 1MB (1,048,576 bytes). Keep in mind that the NFSv3 protocol as implemented has a 64KB maximum read and write size. This means that every SCSI operation in excess of 64KB gets broken down into 64KB NFS operations. For example, this means that for every 1,048,576 bytes (1MB) SCSI write operation displayed in the write chart in Figure 53, 16 NFS write operations are sent to the storage controller.

Figure 52) Read operation for saving an Excel workbook.

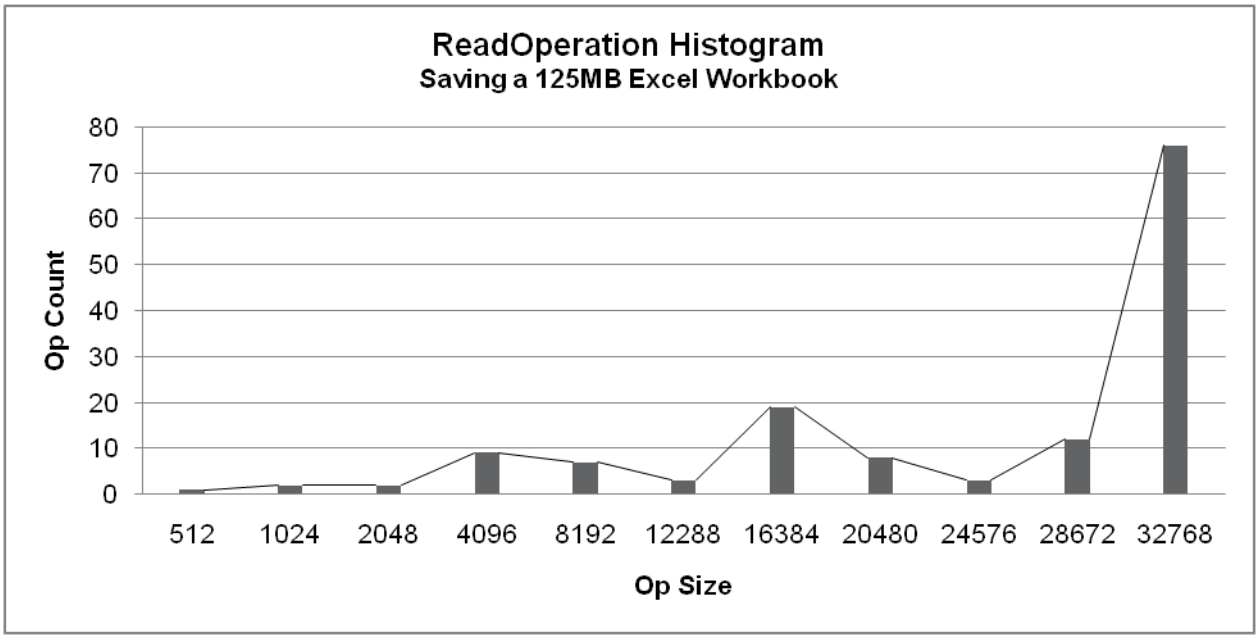
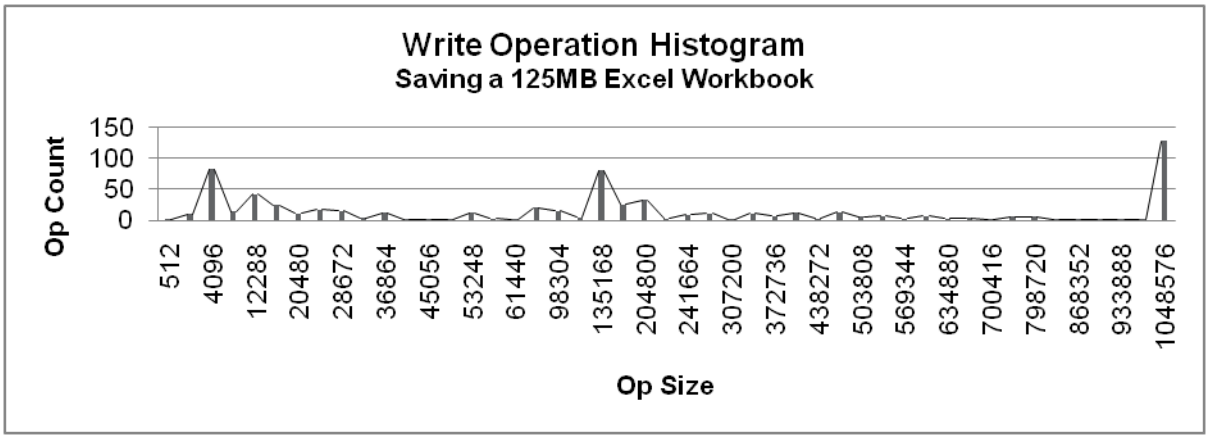


Figure 53) Write operation for saving an Excel workbook.



7 REFERENCES

Liquidware Labs Stratusphere UX
www.liquidwarelabs.com/docs/stratUXfrAndBk.pdf

Performance and Statistics Collector (Perfstat) – Overview
http://now.netapp.com/NOW/download/tools/perfstat/

VMware View Optimization Guide for Windows 7
www.vmware.com/files/pdf/VMware-View-OptimizationGuideWindows7-EN.pdf

VMware View Administrator's Guide
www.vmware.com/pdf/view45_admin_guide.pdf

50,000-Seat VMware View Deployment
http://media.netapp.com/documents/wp-7108.pdf

Understanding Memory Resource Management in VMware ESX 4.1
www.vmware.com/files/pdf/techpaper/vsp_41_perf_memory_mgmt.pdf

8 ACKNOWLEDGEMENTS

The authors of this technical report would like to thank Bhavik Desai and Chris Lemmons of NetApp, Mac Binesh and Ivan Weiss of VMware, and Dnyanesh Khare of Liquidware Labs for all of their contributions to the document.

Special thanks are due to Gary Little of NetApp and Fred Schimscheimer of VMware, without whom the work could not have been done.

www.netapp.com