Lab Report

Infinio with XenDesktop VDI Workload Analysis

By David M. Davis, Brian Suhr, and Scott D. Lowe
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Introduction</td>
<td>4</td>
</tr>
<tr>
<td>How Server-Side Caching Helps You</td>
<td>6</td>
</tr>
<tr>
<td>Considering Server-Side Caching for vSphere</td>
<td>8</td>
</tr>
<tr>
<td>Infinio with XenDesktop and VDI - Lab Setup</td>
<td>10</td>
</tr>
<tr>
<td>Infinio with XenDesktop and VDI - Lab Analysis</td>
<td>11</td>
</tr>
<tr>
<td>Infinio with XenDesktop and VDI - Lab Results</td>
<td>13</td>
</tr>
<tr>
<td>Infinio Accelerator Flattens VDI Latency Spike</td>
<td>19</td>
</tr>
<tr>
<td>Bottom Line</td>
<td>22</td>
</tr>
<tr>
<td>About the Authors</td>
<td>23</td>
</tr>
</tbody>
</table>
As a result, enterprises are going “all in” with virtualized EUC environments with 10,000+ devices, and savvy end users are starting to see desktop-as-a-service as their computing option of choice. However, for administrators who are deploying and managing desktop virtualization / virtual desktop infrastructure (VDI), these solutions can be fraught with risk and uncertainty. Not only can software complexities challenge administrators, but
performance-related challenges all too often derail such initiatives.

The reason that VDI environments commonly struggle is very easy to see in hindsight. For years, enterprises have virtualized server environments with ease and, with those successes in mind, jump eagerly into the VDI waters only to find that VDI workloads are very different from the traditional server workloads to which virtualization administrators and traditional storage hardware has become accustomed. Nowhere is this more evident than during early morning hours, for example, when VDI workloads spike as hundreds or thousands of virtual desktop machines boot all at once or when end users login to their desktops at the same time or when IT deploys refreshed desktop images. Most VDI workloads are almost entirely random when it comes to storage, but this randomness is characterized by periods of unexpected spikes that can bring storage hardware to a standstill, resulting in users that cannot do their jobs.

To combat this, VDI administrators and architects may design the storage infrastructure to support these demand spikes by overprovisioning storage or adding expensive flash storage, potentially overspending on infrastructure to meet peak demand scenarios. On the other hand, if they don't size for demand spikes, then they will undoubtedly deal with horrendous delay times when common VDI issues occur.

So what’s the answer? Does VDI mean turning to a storage vendor and asking for expensive upgrades to support these important workload needs?

"VDI WORKLOADS ARE VERY DIFFERENT FROM THE TRADITIONAL SERVER WORKLOADS TO WHICH VIRTUALIZATION ADMINISTRATORS AND TRADITIONAL STORAGE HARDWARE HAS BECOME ACCUSTOMED."
HOW SERVER SIDE CACHING HELPS YOU

When companies experience poor storage performance, they generally turn to the same hardware vendors that sold them their existing storage area network (SAN) or network attached storage (NAS). Those hardware vendors, naturally, recommend more hardware as the cure to what ails the virtual desktop infrastructure.

For example, in the past, the hardware-based answer to improving I/O throughput for I/O intensive applications – including VDI – has been to add more “spindles” to the existing SAN. As more spindles (hard disk drives) are added to the SAN, the performance ceiling, as measured in I/O Operations Per Second (IOPS) is raised. By spreading the VDI load across more hard disk drives, each individual drive is able to do less work, so the workload is managed better and storage latency may be reduced.

Today, storage hardware vendors would likely recommend options such as:

- **Buy new storage** – purchase a new SAN or NAS that offers greater throughput via beefier controller hardware and all-flash or hybrid-flash storage media; hybrid storage devices are those which combine both traditional hard drives as well as speedy (and expensive) flash-based storage.

- **Upgrade existing storage** – purchase additional disks, cache, flash storage, or implement storage tiering.

- **Buy I/O acceleration hardware** – purchase hardware cards to be installed in servers that accelerate server performance.

“BY SPREADING THE VDI LOAD ACROSS MORE HARD DISK DRIVES, EACH INDIVIDUAL DRIVE IS ABLE TO DO LESS WORK, SO THE WORKLOAD IS MANAGED BETTER.”
To be clear, these are all viable solutions and will help to address the storage challenge. However, many organizations are loathe to rip and replace their storage investments and even expanding existing systems can be an expensive proposition.

Perhaps there is a better option: software-led server-side caching.

A server-side caching environment is created by utilizing high-speed flash or RAM-based storage on a virtual host to accelerate I/O going to and from the shared storage (SAN or NAS). Various vendors accomplish this acceleration goal in different ways. Some provide a read-only cache while others are read-write systems. Some require new hardware to be installed on the host (such as PCI-e-based flash storage) where others do everything in software. Some work for file-based storage where others work for block-based storage.

No matter the solution, server-side caching is becoming a welcomed way for enterprises to:

- Extend the life of their existing storage environment, preventing them from having to upgrade existing storage or from having to purchase a new SAN or NAS (which could cost hundreds of thousands of dollars)
- Accelerate storage performance across the virtual infrastructure for other IO-hungry workloads, such as big data analytics
- Eliminate latency spikes caused by VDI events such as boot storm or login storms
CONSIDERING SERVER-SIDE CACHING FOR vSPHERE

When shopping for server-side caching solutions, keep in mind that not all server-side caching solutions are identical.

Each solution varies greatly in how it is implemented, how it performs, and how it is priced.

When shopping for a server-side caching solution, consider the following factors:

- Return on investment (ROI)
- Hardware requirements and potential downtime for hardware installation (for those solutions that require hardware)
- Potential for downtime related to the installation of the server-side caching software
- Ease of implementation of the overall solution (time is money!)
- The overall impact of the solution; how much performance can be gained once implemented?
- Any cost avoidance opportunities, such as storage hardware that you don’t have to purchase, thanks to the caching solution
With numerous solutions available, you must consider all the aspects of implementing a server-side caching solution given that solutions work differently and have different requirements. Some may necessitate downtime for the infrastructure during implementation and may offer different performance benefits.

One of the leading server-side caching solutions available today is Infinio Accelerator but does Accelerator really provide all the benefits that it promises? These promises include:

“Get SSD-class performance without SSDs”

“No need to rip-and-replace storage hardware”

“Make your desktops run faster, your storage run smoother, and support more desktops”
INFINIO WITH XENDESKTOP AND VDI – LAB SETUP

To see if we could observe, first hand, how Infinio’s Accelerator could improve Xen Desktop VDI workloads, ActualTech Media put the product to a test using simulated – but real world – end user workloads in a lab environment. That lab environment consisted of the following:

**Hardware**
- 4 x Cisco UCS B200M3 blade servers, each with dual hex core processors and 192 GB of RAM
- An EMC VNX5500 storage array with 9 x 200 GB solid state disks plus 27 x 600 SAS disks with hosts connected via NFS
- Cisco 10 Gigabit Ethernet switching to connect the UCS fabric to the VNX storage

**Software**
- VMware vSphere 5.5, installed on each host, and VMware vCenter
- XenDesktop 6.0
- VMware View Planner 3.0
- Windows Server 2012 for Active Directory and DNS
- Windows 7 running on end user desktops

With the hardware and software in place, we began our analysis...
The other metric that is critically important in terms of performance is latency, or response time. When it comes to testing virtual end user computing solutions, latency is actually a far more critical performance metric than IOPS. This is because, of course, end users aren’t just running file transfers all day long (which would benefit from throughput), end users are running real applications like Microsoft Office applications, printing, file conversions, text editing, and more.

The types of activities that end users run typically don’t benefit from a solution that offers very high storage throughput (measured in IOPS). End user activities benefit from a storage solution that offers them the least amount of latency which translates to the best end user application response. Latency is a directly observable outcome of poor-performing storage as users are forced to sit and wait for the infrastructure to respond to their needs.

“END USER ACTIVITIES BENEFIT FROM A STORAGE SOLUTION THAT OFFERS THEM THE LEAST AMOUNT OF LATENCY WHICH TRANSLATES INTO THE BEST END USER APPLICATION RESPONSE.”
While end users spend most of their time working in their desktop applications, there are other activities related to VDI performance that are of importance when sizing storage performance for VDI environments. For example, it’s critical that enterprises also maintain solid performance during activities such as:

- **Boot storms** – when all VDI VMs boot at the same time, for example when infrastructure is restarted; this could happen every single morning as user arrive for work and start their machines

- **Login storms** – when users login at the same time, en masse, for example first thing in the morning

- **Anti-virus storms** – when client-based anti-virus agents simultaneously start scanning for malicious applications

These types of activities have much different performance demands than traditional end user activities. These types of “storms” are characterized by very heavy read-intensive I/O; traditional end-user activities outside these storms are still primarily read-oriented, but have a much more consistent and manageable I/O pattern.

So how does Infinio’s Accelerator help to improve performance for both normal end user activities as well as common VDI storm events? We installed Infinio Accelerator into our lab and performed a battery of “before-and-after” tests to find out.

"While end users spend most of their time working in their desktop applications, there are other activities related to VDI performance that are of importance."
INFINIO WITH XENDESKTOP AND VDI – LAB RESULTS

Utilizing four vSphere hosts running Xen Desktop connected to our EMC storage, we performed a number of VDI analysis tests both before and after installing Infinio Accelerator.

Test #1 – Before Infinio Accelerator

In this test, we ran 300 Windows 7 virtual machines over a period of roughly 60 minutes. These virtual machines logged in over that timeframe (not all at once), as most end users would do. Once logged in, the simulated end users, performed a wide-variety of common activities such as:

- Opening, scrolling through, and editing documents
- Viewing image galleries
- ZIP'ing files
- Web browsing, scrolling through web pages
- Watching videos at a moderate resolution

As the number of desktops increased and the number of activities being performed across all desktop sessions increased, so did the storage utilization. Storage utilization increased in terms of the throughput demands but, more importantly in a VDI environment, the latency of the end user experience increased as the storage became more burdened. The latency of the storage was primarily viewed by observing the I/O wait statistics on the storage device.

These random user activities greatly increased the throughput on our storage array. As you see on the next page, we have a significant increase in demand for IO on our storage array during the test window.
Along with I/O utilization came an increase in CPU usage on the storage array. Our test array was dedicated to this performance test and it went from zero usage to nearly 20% CPU. CPU utilization spiked during the boot storm of 300 desktops starting up in a short window and then again during the simulated workload window.

Figure 2 – Storage IO Demands Increasing and Approaching 7,500

Per Controller I/O - Aggregate Approaching 7,500 Operations (compare to Figure 7)

Figure 3 – Storage Controller CPU Increasing and well above 15%

Per Controller CPU Utilization (compare to Figure 8)
Intermission

To determine if Infinio’s Accelerator could significantly reduce the I/O load, CPU utilization and end user latency, we installed the product. This was done by downloading a vSphere OVA file and using the Deploy OVF option in the vSphere client. We used the OVF method, but Infinio actually recommends using the Windows installer method since it is more robust and does additional housekeeping.

Within about five minutes, the Infinio console was up and running and we were able to access the GUI interface via our web browser. From there, all we had to do was to click Accelerate on the console to deploy the Infinio Accelerators on each ESXi host.

There was NO hardware to install; there was NO host downtime; there were NO vSphere configurations to be made. In fact, the entire installation took less than 30 minutes (including downloading the Infinio Accelerator software).

Figure 4 demonstrates what the Infinio Accelerator looks like, once deployed.
Test #2 – After Infinio
In the second test, with the Infinio Accelerator installed, we re-ran the same end user simulations using XenDesktop. With the same user count and workload, across the hosts and the storage, the load naturally began to build.

Using the Infinio console, we observed that the effective cache size used by the Infinio Accelerator was as much as 92GB+ of memory-based storage (which is even faster than SSD flash storage).

Even though no additional storage caching hardware was installed on the hosts or in the storage, the Infinio Accelerators were able to achieve a cache size of 92GB+ using only 32 GB of RAM (8GB of RAM was allocated to Infinio from each of the 4 hosts). This incredible cache size was thanks to the impressive deduplication rates that Infinio Accelerator is able to achieve.

**ACTUAL FINDING**
Infino was able to supply the storage I/O equivalent to 86 disk spindles. As you see below, at the busiest workload times (such as boot storms and heavy VDI user load), Accelerator was able to provide an “additional” 86 disk spindles through server-side caching.

![Effective Cache Size](image)

Figure 5 – Infinio Accelerator Effective Cache Size
With this added performance boost, the XenDesktop VDI workload that we placed on our hosts had less latency, better response and less impact on the storage array.

Figure 6 – Infinio Accelerator Storage Boost

Figure 7 – Disk I/O Reduced by 74% After Infinio (barely above 600, was almost 7,000 pre-Infinio)
ACTUAL FINDING
The first indicator was that the storage I/O dropped by roughly 74%, as show in Figure 7 below. As you can also tell, the volume of I/O that was sent to the disk was also reduced significantly from the before testing.

At the same time, the Disk I/O and CPU utilizations were significantly reduced, as you see in the graph above.

Across the numerous tests that we ran, the Infinio Accelerator was able to offload roughly 60-80% of I/O requests and improve I/O response times. As you see in the graph below, at this point in time 78% of storage requests were offloaded and 88% of bandwidth was saved. Typically desktop workloads are write heavy at times.
While testing Infinio Accelerator in our lab with XenDesktop workloads, we also noticed a number of other benefits around how Accelerator was providing a much greater quality of service for storage I/O.

At numerous points while building and testing the lab environment where Infinio was tested, we noticed that Infinio was providing incredible response time improvements and storage offload.

We discovered that, when all Xen VMs were powered on at once (through Infinio’s caching) an 89% storage bandwidth decrease was achieved.

By serving these disk I/O’s from the Accelerator’s RAM-based cache rather than from the storage array itself, the desktops experience a far better performance experience and the stress on the storage array is reduced. By removing 60-80% of read I/O from the storage array it is able to provide better performance for the write I/O that is not accelerated.

As mentioned before, another aspect that admins must consider when it comes to VDI are the common I/O spikes that are caused by boot storms, login storms, AV storms, and the recomposition of VDI virtual desktops.

“AT NUMEROUS POINTS TESTING THE LAB ENVIRONMENT, WE NOTICED THAT INFINIO WAS PROVIDING INCREDIBLE RESPONSE TIME IMPROVEMENTS AND STORAGE OFFLOAD.”
ACTUAL FINDING
VDI boot storms are common and occur when VDI hosts are restarted or when VDI pools that use power consumption are powered on, for example, first thing in the morning. For end users who are waiting to access their desktop that might be powering on at that point, Infinio’s server-side caching solution would allow admins to bring up desktops faster and end users to get to work quicker (and without as much frustration).

Figure 10 – Boot Storm Latency Decrease
**ACTUAL FINDING**

At another point during testing, we observed that Infinio’s Accelerator was offloading 83% of all storage I/O. In other tests we have seen results above 90% for offloads. Again, while having long periods above 90% isn’t what one should expect to be the norm, it’s still impressive how much faster applications might be if a high percentage of storage I/O were served out of local server RAM vs. traversing the network to access a magnetic disk.

Infinio Accelerator should be seen not just as a solution to improve performance across the board but also as a powerful solution for smoothing out the latency spikes caused by the common VDI events we discussed here.

![Graph showing 83% of requests offloaded](image)

*Figure 11 – Infinio Accelerator drastically reduces requests that affect primary storage*
Enterprises running VDI solutions such as Citrix XenDesktop and XenDesktop View are ideal candidates for server-side caching because it provides their users faster application response without requiring the company to purchase expensive new storage hardware. Particularly pleasing to end users and admins alike will be when common VDI events like boot or login storms are no longer an issue, or even noticed, thanks to the improved quality of service offered by server-side caching.

Server-side caching is an effective new way to boost storage performance and eliminate latency bottlenecks.

“Particularly pleasing to end users and admins alike will be when common VDI events like boot or login storms are no longer an issue, or even noticed.”
ABOUT THE AUTHORS

Brian Suhr is Senior Solutions Architect for a VMware partner. He spends his days working with customers to Design and Architect solutions around virtualization technologies, VDI, Cloud computing, storage and compute infrastructure and related technologies. Brian has helped with leading the Chicago VMUG for five years and has presented multiple sessions at local and regional VMUG meetings. Brian writes for his virtualizatips.com and datacenterzombie.com sites and several community sites that include TechTarget, Petri.co and VMware SMB blog.

David Davis is a well-known virtualization and cloud computing expert, author, speaker, and evangelist, as well as a co-founder of ActualTech Media. David’s library of popular virtualization video training courses can be found at Pluralsight.com. He holds several certifications including VCP5, VCAP, CCIE, and has been awarded the VMware vExpert award 6 years running. David is also the Senior Editor for VirtualizationSoftware.com.

Scott D. Lowe is co-founder of ActualTech Media and the Senior Editor of EnterpriseStorageGuide.com. Scott has been in the IT field for close to twenty years and spent ten of those years in filling the CIO role for various organizations. Scott is also a micro-analyst for Wikibon and an InformationWeek Analytics contributor. In addition, Scott has also written thousands of articles and blog postings and regularly contributes to such sites as TechRepublic, Wikibon, and virtualizationadmin.com. Because of his unique blend of skills (CIO/strategic & Engineer/tactical) Scott is also a sought-after resource for marketing collateral, including white papers and e-books, for a variety of technology firms.

ABOUT INFINIO SYSTEMS

Infinio offers downloadable storage performance for virtualized environments. The software-only solution requires no new hardware, no reboots and no service disruption to deliver the IOPS your virtualized desktops and servers need. The company is backed by top-tier investors, and is based in Cambridge, Mass. Learn more about us at www.infinio.com.