



## Alarming Consequences of Misleading NFS Myths in VMware Storage Environments

Marc Staimer, President & CDS of Dragon Slayer Consulting

### Introduction

There are numerous pervasive NFS storage myths that surround VMware® (ESX™ and vSphere™) infrastructure that have serious implementation, operational, and management implications to those who continue to believe them. What makes these myths so infuriating is that believing these myths end up adding significantly to the task load of already overworked VMware administrators. And like all IT professionals, they're human beings with inelastic task limits per hour, per day, per week, per year. More tasks only leads to more errors, problems, cost, and frustration.

Conventional wisdom is that the best storage for VMware infrastructure is Fibre Channel or iSCSI SAN storage. Like most conventional wisdom, it's incorrect. The best storage for VMware infrastructure, and by far and away the one requiring the least amount of management, is NFS NAS. But, there are myths surrounding NFS and VMware that inhibit its adoption. There are also some real problems with most NFS storage that must be addressed in order to get the promised benefits. To get a better handle on these issues requires first an understanding of the prevalent VMware storage ecosystem and its associated problems.

### VMware's Common VMFS Storage Ecosystem

The most common VMware datastore is VMFS (virtual machine file system) running on a Fibre Channel SAN. VMFS is VMware's own file system. It is natural for them to develop on VMFS first and then NFS. VMFS is a clustered file system that allows LUNs to be accessed simultaneously by multiple ESX or vSphere servers running multiple VMs.

The strength of VMFS is that it is robust, mature, and well understood. It provides the VMware administrator with a fair amount of independence from the storage administrator for shared infrastructure. The VMware administrator essentially takes the LUNs allocated by the storage administrator, slices, dices, and allocates them as required. Most data management operations are performed exclusively through VMware vCenter Server™. However, VMFS has a series of well known shortcomings.

Take something like [scaling](#), which based on how VM scales, should be simple. It's not. Scaling VMFS requires careful, manually intensive planning when deploying many virtual machines. This is because shared datastores do not differentiate VMs and pose a serious challenge for IO prioritization that can lead to VM storage time-outs. The SAN storage arrays cannot identify IO load by individual VMs. Monitoring IO load is best from within vCenter and is also a manually intensive process. For datasets of Tier 1 applications that have demanding IO requirements the VMFS best practice is to manually configure and deploy a virtual disk on a VMFS datastore that is not shared by other VMs.

[Troubleshooting IO performance](#) issues on VMFS is a time consuming and often wearisome exercise because of the inherent lack of automation. Troubleshooting intermittent performance problems is an order of magnitude more complex. Just capturing the data necessary to troubleshoot can take excessively long periods of time. And even then, there is no guarantee that the root cause can be found. This type of storage ecosystem seems completely opposite and out of place when compared to the simplicity of VMware's virtual server ecosystem.

As VMware virtualized [ecosystems get bigger](#), these issues intensify at a far greater pace. It becomes inherently apparent that the VMFS storage ecosystem is very inefficient for large numbers of VMs (hundreds to thousands). In a nutshell, VMFS is generally manually intensive to implement, operate, manage, and troubleshoot. For the VMware administrator the pedantic task load is becoming much too high.

### The Advantage of VMware NFS Datastores

The well documented advantages of NFS datastores show that an NFS based NAS storage solution is much simpler, more automated, and a superior storage ecosystem for VMware datastores. Some of the NFS datastore advantages include:

- Incredibly easy provisioning since all VMDKs are just files.



## Dragon Slayer Consulting

- VMware based VMDK thin provisioning (vSphere 4 default setting).
- NFS volume expansion or shrinkage on the fly with the changes being effective within the ESX or vSphere server by simply clicking the datastore “refresh” button.
- Zero management of, or need for VMFS or RDM (Raw Data Mapping).
- Elimination of constrained or single disk IO queues, making VM IO performance only limited by the server IO, NFS storage system IO, the end-to-end bandwidth, and/or latency.
- Removal of any management, operations, or change management for FC switches, zones, hard zones, HBAs, LUN masking, or identical LUN IDs across ESX/vSphere servers.
- Increased granularity of VM backups from the entire VM, down to specific files within VMs.
- VM restorability with a much finer granularity that can be multiple VMs, individual VMs, or files within VMs.
- Single or multiple VM cloning that’s as simple as point, click, copy.
- Optimized random IOPS response time, which is an NFS strong suite and ideal for ESX and vSphere since their IO is small block and extremely random (bandwidth is not as important as random IOPS response time).
- NFS provides the ability to allow a single mount point across multiple IP addresses.
- Performance scaling that’s comparatively easy with link aggregation IEEE 802.3ad.

The question naturally arises then that if NFS datastores is so much simpler, more automated, has more advantages, and requires fewer tasks than VMFS datastores, why is it not deployed more often? The answer comes down to the persistent, false mythology that surrounds virtualization environments and NFS datastores. In some cases the myth is based on past facts that have not caught up with the current state of VMware. In other cases the myth is based on a kernel of reality that doesn’t apply to NFS storage solutions from BlueArc®.

Examining these myths makes it clearly evident as to why they’re false, and the negative upshot for the many VMware administrators who believe these false assertions.

### False Myths of NFS Storage in VMware Environments

#### ○ *VMware does not fully support or support all advanced functions on NFS*

This is categorically [false](#). All of VMware vSphere and ESX features are supported with NFS datastores. VMware has in fact fully supported NFS as a standard storage option since ESX 3.0. The myth originates from VMware’s development cycle. They first develop and support new advanced features on VMFS datastores with SAN storage. The NFS support typically follows usually within a quarter or two; although, depending on circumstances, it can be longer. This is what happened with Storage Vmotion™ and Site Recovery Manager™ (SRM). The support for NFS protocol was delayed by the incredibly important VMware release of vSphere 4.

SRM (SRM 4) NFS support for vSphere and ESX did not occur until October 2009 approximately one year after VMFS datastore SAN support. Storage Vmotion NFS support also trailed the VMFS SAN storage support by about a year because of vSphere4 prioritization<sup>1</sup>.

A quick look at VMware’s website demonstrates this complete NFS support which is summarized in the table below.

---

<sup>1</sup> Storage Vmotion is fully supported on vSphere, but is considered experimental on ESX.



Protocol	Vmotion, DRS, HA	VCB	Storage Vmotion	RDMs	ESX boot from disk	Fault Tolerance
FC SAN	Yes	Yes	Full Support	Yes	Yes	Yes
iSCSI SAN HW init	Yes	Yes	Full Support	Yes	Yes	Yes
iSCSI SAN SW init	Yes	Yes	Full Support	Yes	No	Yes
NFS	Yes	Yes	Full Support w/vSphere Experimental w/ ESX	NA	No (Yes w/ PXE)	Yes

Table 1: VMware Advanced Function Storage Protocol Support (Feb 2010)<sup>2</sup>

BlueArc NFS works with all VMware advanced functionality.

o *VMware performance is much slower with NFS datastores*

This myth is perhaps the most pervasive, persistent, damaging, and also [false](#). Some SAN vendors and experts alike consistently repeat this myth. However the evidence does support the assertion. VMware ran performance tests that demonstrated NFS running on 1Gbps Ethernet performance provided less than a 9 to 10% decrease from that of 4Gbps Fibre Channel. This test did not evaluate performance when 1Gbps Ethernet NICs are trunked on both the VMware server and the NFS storage system.

Tests by BlueArc as well as Dell® have demonstrated that the performance of VMware using NFS running on 10Gbps Ethernet versus VMFS utilizing 8Gbps Fibre Channel is equivalent or even higher<sup>3</sup>.

BlueArc NFS on 1Gbps Ethernet ports, 10Gbps Ethernet ports, or trunked Ethernet ports has proven to be the fastest in the industry as a direct result of BlueArc's SiliconFS™. SiliconFS puts NFS into FPGA silicon, radically decreasing latency while increasing processing and performance.

o *VMware CPU load is significantly higher with NFS*

This myth comes from an “apples versus oranges” test. It compared the VMware CPU load when utilizing NFS datastores and TCP/IP protocol versus the VMware CPU load of the Fibre Channel and iSCSI (software and hardware based) protocols. The tests showed NFS TCP/IP overhead ranged from 15 to 40% more CPU overhead than the Fibre Channel protocol/drivers. A key problem with this test, and what makes it invalid, is that it did not measure the VMware CPU load for VMFS datastores when utilizing Fibre Channel or iSCSI protocols/drivers. It only measured the protocol/driver overhead. The overhead of VMFS was not measured.

Since NFS datastores eliminate VMFS datastore overhead, the overall CPU load should be far more equivalent. Even the apples versus oranges comparison CPU load differential results was not statistically remarkable when considering Fibre Channel protocol is all silicon based, whereas the NFS and TCP/IP protocols were all software based. BlueArc's Network Storage rectifies this discrepancy by putting both NFS and TCP/IP protocols in silicon and eliminates any, and all, CPU load issues. This assertion is generally true for most NAS but [false](#) for BlueArc.

o *VMware is limited to only 8 NFS datastores*

This myth is based on ESX and vSphere NFS datastore setup defaults. Again it's [false](#). ESX supports up to 32 NFS datastores and vSphere supports up to 64.

BlueArc NFS supports VMware's ESX and vSphere maximum number of datastores.

<sup>2</sup> From VMware's website

<sup>3</sup> Fibre Channel can demonstrate higher aggregate performance when costly third party software such as EMC's PowerPath is used to load balance IO across multiple HBAs and Fibre Channel paths.



- *VMware NFS datastores only scale to 16TB*

The 16TB NFS datastore limitation is [false](#). 16TB NFS datastore is not a VMware limitation; it's the typical file system size limit for most NFS storage systems. When the NFS storage system has this limit, then so too do the NFS datastores. BlueArc has a 256TB file system and a global name space that provide up to a 4PB name space. This means a datastore can grow up to 4PB or 256 times larger than a 16TB datastore.

- *NFS thin provisioned VMDKs automatically rehydrate when moved or cloned*

This is true with ESX and [false](#) with vSphere. Thin provisioning is the default setting for NFS VMDK datastores with vSphere 4 and no longer require rehydration to be moved by Vmotion™ to a different NFS datastore or to be cloned.

- *Microsoft Windows® VMs can't boot or utilize NFS datastores*

This comes from the fact that the Microsoft Windows server doesn't really support NFS and can't boot from NFS. Therefore, many believe that when it runs as a VM in VMware that VMware cannot utilize NFS datastores. That is [false](#). VMware built the NFS protocol into the ESX and vSphere virtualization layer. Microsoft Windows VMs never see the NFS protocol and are unaware that they are utilizing NFS datastores.

## Four Key Problems with the Vast Majority of NFS Storage Systems

As good as NFS datastores are for VMware, there are nontrivial issues with how most vendors implement NFS in their storage systems. These include:

1. ***A 16TB file system that limits the size of NFS datastores to 16TB.*** As VMs and data scale beyond 16TB, this small file system limitation means more file systems, nodes, or NAS systems have to be implemented to meet that growth. That results in more system ports, cables, switch ports, switches, switch management, storage system management, data protection management, rack space, floor space, power, cooling and a heck of a lot more admin tasks and time they don't have. More NAS storage systems or nodes usually require more data migration as well. When redundant, duplicated, pedantic operational tasks are increased, so too are the probability of human errors. More errors, more troubleshooting, and more frustration. This is not what most VMware admins are signing up for when they select NAS datastores to simplify their lives.
2. ***Software based NFS that severely limits the IO performance of individual ports.*** In order to meet many IO performance requirements, the admin must implement additional server and storage Ethernet ports, bonding via link aggregation of IEEE 802.3ad link aggregation, resulting in increased task management of the servers (more NICs), Ethernet networks (more switch ports and switches), and NAS storage system (more NICs and potentially more systems or nodes). All of which leads to more power and cooling as well.
3. ***File system file object management limitations.*** This is a little known, though often experienced, NFS issue with so many NAS systems. Each system can only manage a limited number of file objects. That number varies by the size and characteristics of the files. When the limit is being approached, the NAS system provides no warning. Typically the first sign that the limitation has been reached is when the NAS system stops responding. All of the connected applications come to a halt. Then the storage admin must go into the system and manually remove enough files to restart the system and migrate more of the files and data to another system. Again more work for the admin and ultimately more NAS systems to balance the load and hedge against the file object limitations being met. More systems, networks, ports, cables, switch ports, switches, rack space, floor space, power, cooling, management, tasks, errors, troubleshooting, and all their associated costs.
4. ***Manual storage tiering.*** The value of files and data declines as it ages, which is why storage tiering conceptually makes sense (different price performance disk tiers such as Fibre Channel or SAS drives, and SATA drives). Unfortunately, the vast majority of NAS NFS data migration processes are manually intensive and often application disruptive. Even VMware's own Storage Vmotion is a manual process, albeit application non-disruptive. So even though storage tiering is



# Dragon Slayer Consulting

conceptually a good idea, most admins keep those files and that data on their primary Tier 1 storage. They do this because of a lack of automation.

There has to be a better, smarter way to provide the value of VMware and NFS while eliminating these severe operational limitations. There is. It's BlueArc's Network Storage systems and intelligent data management software.

## The BlueArc Resolution to These Very Real Problems

BlueArc's Network Storage systems deliver on the incredible simplicity of NFS with VMware while uniquely eliminating all of the issues that limit most NAS systems. It does it by providing:

1. **256TB file systems that scale to 4PB with built-in Global Namespace.** 256TB file system is 16X greater than 16TB and four times greater than the largest LUN volume that is available with VMFS. These BlueArc advantages grow another 16X with the built-in GNS. File system size is not a BlueArc limitation.
2. **Silicon based File System a.k.a. SiliconFS™.** BlueArc's SiliconFS has proven to be the fastest NFS in the industry on both a system and per Ethernet port basis<sup>4</sup>. In other words, BlueArc requires a heck of a lot fewer 1Gbps and/or 10Gbps Ethernet ports, cables, switch ports, switches, rack space, floor space, power, and cooling to achieve the same or better performance than anyone else.
3. **Non-stop file system.** Two other key advantages of the BlueArc SiliconFS are:
  - I. It's inherent ability to keep running (non-stop) even when the maximum number of file objects management limit is reached. Even though new file objects cannot be added, system access and performance does stop or even decline.
  - II. The number of managed file objects it can manage is larger than any other NAS system.
4. **Policy based automated storage tiering and data migration.** Storage tiering becomes not just valuable, but essential when it's simple, policy based, and automated. BlueArc's Data Migrator software automates that storage tiering and data migration. It's based on a rich set of predefined rules, templates, and policies that's set by the administrator. Data migration between tiers occurs whenever a rule is triggered. Data Migrator provides a 'pointer' to the original file location after files are moved to a different tier. This pointer allows the file to be accessed transparently by the VM, application, or user as if the file had not been moved at all. And unlike HSM solutions, it does not require that the file be rewritten (pulled back) to the tier it was originally written to in order to be accessed.

No other VMware certified NAS system provides the BlueArc combination of NFS benefits while eliminating common, manually intensive traditional NAS problems. BlueArc simplifies and reduces VMware admin tasks while doing the same for storage admins and providing unmatched per port performance.

## Summary and Conclusion

The simplest VMware networked storage and the one that quickly becomes the favorite of most VMware administrators, is NFS based Networked Storage. It dramatically simplifies VM storage provisioning, data protection, management, task management, and troubleshooting, especially the intermittent performance issues of VMFS datastores on SAN storage.

But since most NAS systems have issues with NFS datastore size limits of 16TB, per port IO performance limitations, annoying file object limitations, plus manually intensive, application disruptive storage tiering and data migration, it puts a damper on that VMware simplification.

BlueArc Network Storage meets or exceeds VMware storage requirements, simplifies them, and eliminates the most common NAS system issues. The choice is clear. You owe it to yourself to take a hard look at BlueArc Network Storage.

---

<sup>4</sup> <http://www.spec.org/sfs2008/results/sfs2008nfs.html>