THE GORILLA GUIDE™ TO...

# Hyperconverged Infrastructure for Virtual Desktop Infrastructure

**Scott D. Lowe**
Consultant & Industry Veteran

Brought to you by
Hewlett Packard Enterprise

*Helping you navigate the technology jungle*

THE GORILLA GUIDE TO...

# Hyperconverged Infrastructure for Virtual Desktop Infrastructure

**Scott D. Lowe** • **David M. Davis**

Author: Scott D. Lowe, ActualTech Media
Editors: David M. Davis, ActualTech Media
Hilary Kerchner, Dream Write Creative

Book Design: Braeden Black, Avalon Media Productions
Geordie Carswell, ActualTech Media

Layout: Braeden Black, Avalon Media Productions



Printed in the United States of America

First Printing, 2015

ISBN 978-1-943952-02-1

ActualTech Media
Okatie Village Ste 103-157
Bluffton, SC 29909
*www.actualtechmedia.com*

# Table of Contents

# 1

# Introduction to Hyperconverged Infrastructure

In recent years, it seems like technology is changing faster than it used to in decades past. As employees devour newer technologies such as smartphones, tablets, wearables, and other devices, and as they become more comfortable with solutions such as Dropbox and Skype, their demands on enterprise IT intensify. Plus, management and other decision makers are also increasing their demands on enterprise IT to provide more infrastructure with less cost and time. Unfortunately, enterprise IT organizations often don't see much, if any, associated increases in funding to accomodate these demands.

These demands have resulted in the need for IT organizations to attempt to mimic NASA's much-heralded "Faster, Better, Cheaper" operational campaign. As the name suggests, NASA made great attempts to build new missions far more quickly than was possible in the past, with greater levels of success, and with costs that were dramatically lower than previous missions. NASA was largely successful in their efforts, but the new missions tended to look very different from the ones in the past. For example, the early missions were big and

complicated with a ton of moving parts, while modern missions have been much smaller in scale with far more focused mission deliverables.

### What is NASA?

NASA is the United States National Aeronautical and Space Administration and has been responsible for helping the U.S. achieve success in its space programs, from the moon landing to recent high quality photographs of Pluto. NASA has faced serious budget cuts in recent years, but has been able to retool itself around smaller, more focused missions that cost less and have achieved incredible results.

The same "faster, better, cheaper" challenge is hitting enterprise IT, although even the hardest working IT pros don't usually have to make robots rove the surface of an inhospitable planet! Today's IT departments must meet a growing list of business needs while, at the same time, appeasing the decision makers who demand far more positive economic outcomes (either by cutting costs overall or doing more work within the existing budget).

Unfortunately, most of today's data center architectures actively work against these goals, because with increasing complexity comes increased costs — and things have definitely become more complex. Virtualization has been a fantastic opportunity for companies, but with virtualization has come some new challenges, including major issues with storage. With virtualization, enterprise IT has moved from physical servers, where storage services could be configured on a per-server basis, to shared storage systems. These shared storage systems, while offering plenty of capacity, have often not been able to keep up in terms of performance, forcing IT departments to take corrective actions that don't always align with good economic practices.

For example, it's common for IT pros to add entire shelves of disks, not because they need the capacity, but because they need the spindles to increase overall storage performance. There are, of course, other ways to combat storage performance issues, such as through the use of solid state disk (SSD) caching systems, but these also add complexity to what is already a complex situation.

There are other challenges that administrators of legacy data centers need to consider as well:

- **Hardware sprawl**. Data centers are littered with separate infrastructure silos that are all painstakingly cobbled together to form a complete solution. This hardware sprawl results in a data center that is increasingly complex, decreasing flexibility, and expensive to maintain.

- **Policy sprawl**. The more variety of solutions in the data center, the more touch points that exist when it comes to applying consistent policies across all workloads.

- **Scaling challenges**. Predictability is becoming really important. That is, being able to predict ongoing budgetary costs and how well a solution will perform after purchase are important. Legacy infrastructure and its lack of inherent feature-like scaling capability make both predictability metrics very difficult to achieve.

- **Desire for less technical overhead**. Businesses want analysts and employees that can help drive top line revenue growth. Purely technical staff are often considered expenses that must be minimized. Businesses today are looking for ways to make the IT function easier to manage overall so that they can redeploy technical personnel to more business-facing needs. Legacy data centers are a major hurdle in this transition.

So, with all of this in mind, what are you to do?

## Hyperconverged Infrastructure from 30,000 Feet

An emerging data center architectural option, dubbed *hyperconverged infrastructure*, is a new way to reduce your costs and better align enterprise IT with business needs. At its most basic, hyperconverged infrastructure is the conglomeration of the servers and storage devices that comprise the data center. These systems are wrapped in comprehensive and easy-to-use management tools designed to help shield the administrator from much of the underlying architectural complexity.

Why are these two resources, storage and compute, at the core of hyperconverged infrastructure? Simply put, storage has become an incredible challenge for many companies. It's one of— if not *the* — most expensive resources in the data center and often requires a highly skilled person or team to keep it running. Moreover, for many companies, it's a single point of failure. When storage fails, swaths of services are negatively impacted.

Combining storage with compute is in many ways a return to the past, but this time many new technologies have been wrapped around it. Before virtualization and before SANs, many companies ran physical servers with directly attached storage systems, and they tailored these storage systems to meet the unique needs for whatever applications might have been running on the physical servers. The problem with this approach was it created numerous "islands" of storage and compute resources. Virtualization solved this resource-sharing problem but introduced its own problems previously described.

Hyperconverged infrastructure distributes the storage resource among the various nodes that comprise a cluster. Often built using commodity server chasses and hardware, hyperconverged infrastructure nodes and appliances are bound together via Ethernet and a powerful software

layer. The software layer often includes a *virtual storage appliance* (VSA) that runs on each cluster node. Each VSA then communicates with all of the other VSAs in the cluster over an Ethernet link, thus forming a distributed file system across which virtual machines are run.

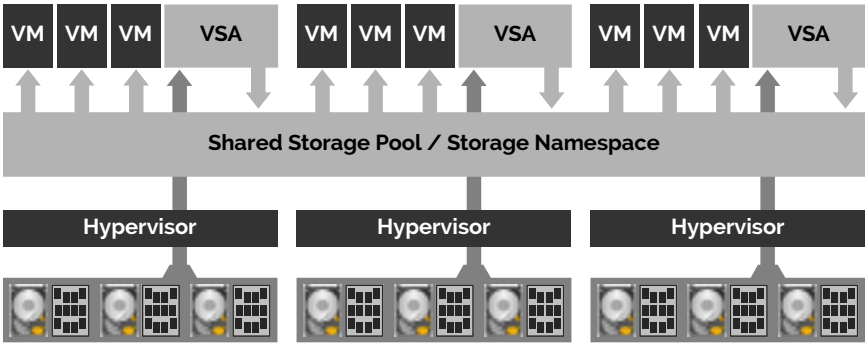


**Figure 1-1:** An overview of a Virtual Storage Appliance

The fact that these systems leverage commodity hardware is critical. The power behind hyperconverged infrastructure lies in its ability to coral resources – RAM, compute, and data storage – from hardware that doesn't all have to be custom-engineered. This is the basis for hyperconverged infrastructure's ability to scale granularly and the beginnings of cost reduction processes.

The basics behind hyperconverged infrastructure should be well understood before proceeding with the remainder of this book. If you're new to hyperconverged infrastructure or are unfamiliar with the basics, please read *Hyperconverged Infrastructure for Dummies*, available now for free from www.hyperconverged.org.

## Resources to Consolidate

The basic combination of storage and servers is a good start, but once one looks beyond the confines of this baseline definition, hyperconverged infrastructure begins to reveal its true power. The more hardware devices and software systems that can be collapsed into a hyperconverged solution, the easier it becomes to manage the solution and the less expensive it becomes to operate.

Here are some data center elements that can be integrated in a hyperconverged infrastructure.

### Deduplication Appliances

In order to achieve the most storage capacity, deduplication technologies are common in today's data center. Dedicated appliances are now available which handle complex and CPU-intensive deduplication tasks, ultimately reducing the amount of data that has to be housed on primary storage. Deduplication services are also included with storage arrays in many cases. However, deduplication in both cases is not as comprehensive as it could be. As data moves around the organization, data is rehydrated into its original form and may or may not be reduced via deduplication as it moves between services.

### SSD Caches/All-Flash Array

To address storage performance issues, companies sometimes deploy either solid state disk (SSD)-based caching systems or full SSD/flash-based storage arrays. However, both solutions have the potential to increase complexity as well as cost. When server-side PCI-e SSD cards are deployed, there also has to be a third-party software layer that allows them to act as a cache, if that is the desire. With all-flash arrays or flash-based stand-alone caching systems, administrators are asked to support new hardware in addition to everything else in the data center.

### Backup Software

Data protection in the form of backup and recovery remains a critical task for IT and is one that's often not meeting organizational needs. Recovery time objectives (RTO) and recovery point objectives (RPO) — both described in the deep dive section entitled "The Ins and Outs of Backup and Recovery" — are both shrinking metrics that IT needs to improve upon. Using traditional hardware and software solutions to meet this need has been increasingly challenging. As RPO and RTO needs get shorter, costs get higher with traditional solutions.

With the right hyperconverged infrastructure solution, the picture changes a bit. In fact, included in some baseline solutions is a comprehensive backup and recovery capability that can enable extremely short RTO windows while also featuring very small RPO metrics.

#### The Ins & Outs of Backup & Recovery

There are critical recovery metrics – known as Recovery Time Objective (RTO) and Recovery Point Objective (RTO) that must be considered in your data protection plans. You can learn a lot more about these two metrics in Chapter 4.

### Data Replication

Data protection is about far more than just backup and recovery. What happens if the primary data center is lost? This is where replicated data comes into play. By making copies of data and replicating that data to remote sites, companies can rest assured that critical data won't be lost.

To enable these data replication services, companies implement a variety of other data center services. For example, to minimize replication impact on bandwidth, companies deploy WAN accel-

eration devices intended to reduce the volume of data traversing the Internet to a secondary site. WAN accelerators are yet another device that needs to be managed, monitored, and maintained. There are acquisition costs to procure these devices; there are costs to operate these devices in the form of staff time and training; and there are annual maintenance costs to make sure that these devices remain supported by the vendor.

# 2

# Virtual Desktop Infrastructure

For years, IT pros have been trying their best to improve what has become a chaotic desktop management environment and to reduce costs for providing desktop computers. One of the original use cases around hyperconverged infrastructure was virtual desktop infrastructure (VDI).

VDI is an interesting solution. Like many trends in IT, VDI has gone through Gartner's "hype cycle" (**Figure 2-1**). It went through both a period of sky-high expectations and also hit rock bottom as people became increasingly disillusioned with the technology. Today, however, it's reaching the top end of the Slope of Enlightenment and entering the Plateau of Productivity.

How did we get to where we are?

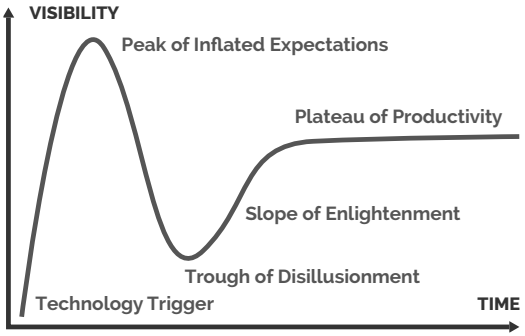


**Figure 2-1:** Gartner's hype cycle (courtesy of Wikipedia)

## VDI Through the Years

Long before x86-based virtualization became the norm, IT departments searched for ways to simplify and streamline desktop computing. Microsoft and Citrix led the way in this space and, for a time, their products were ubiquitous. People deployed thin clients based on specialized editions of Windows Server and had an adequate experience. Unfortunately, their experience was one that was mostly useful where terminals, not full desktop capabilities, were needed.

Then came along server virtualization. Server virtualization resulted in the ability to transform the business and IT — lowering costs while increasing productivity and efficiency along the way. With server virtualization, data center administrators could almost completely replicate their physical servers inside software with little to no loss of functionality.

At some point, someone somewhere had the bright idea to attempt to apply the same thinking to desktops in order to close the user-experience gap and make terminal-based desktops more like their PC brethren. Things didn't work out quite so well. IT pros quickly discovered that their path to VDI success would be littered with very different challenges than those faced on the road to server virtualization.

## VDI Workload Differentiators

Although servers and desktops are both computers, how they're used is very different. These differences have driven many of the challenges that doomed early VDI projects. Just because virtual desktops look like virtual servers, it doesn't mean they act like them. Whereas server-based workloads will have their own performance peaks and valleys, they're nothing compared to what happens in the world of the virtual desktop.

## Types of Virtual Desktops

There are two different kind of virtual desktops that you can use in a VDI environment: Persistent and Non-Persistent.

### *Persistent Desktops*

Persistent desktops are the type that closely resemble desktop computers in the physical world. There is a 1:1 relationship between a virtual desktop and a user. In other words, a user has his or her own virtual desktop that no one else uses. This model is the most seamless from a user perspective since users have become used to having their own space. Persistent desktops require you to have sufficient storage for desktop customizations.

### *Non-Persistent Desktops*

Think of a college computer lab: rows of computers available for any student, and students can go to different computers every day. The students really don't care which computer they use each day as long as one is available and they don't need to maintain user-specific settings. This is known as a non-persistent desktop. User settings are not maintained between sessions. Each time a user logs in, it's as if he or she has logged in for the first time.

## Linear Usage Patterns

In VDI environments, usage patterns directly follow user actions. When users log in or boot their virtual desktops in the morning, each virtual desktop undergoes significant storage I/O operations. Contrast this to a traditional PC, where you've probably seen it take minutes for computers to fully boot and login. This is because a lot of information has to be read from disk and placed into memory on a traditional PC. There are also write operations taking place, such as when Windows logs any exceptions that may take place at boot time.

Now, multiply all of this I/O by the number of users logging into their virtual desktops at the same time. In the world of the traditional desktop, each user has his or her own storage device (the local hard drive) to handle these I/O operations. In a VDI environment, the virtual desktops all share common storage systems, often a SAN or NAS device shared among the various hosts that house the virtual desktops. The amount of I/O that starts to hit storage can be in the hundreds, thousands, or even tens of thousands of IOPS.

## The Failure and Resurgence of Storage

This was the problem in the early days of VDI. Then-current disk-based storage systems simply could not keep up with demands and quickly succumbed under the IOPS-based assault that came their way. This led directly to the hype cycle's Trough of Disillusionment as people quickly discovered that there would be no return on their VDI investment because they had to buy shelves and shelves of disks to keep up with I/O demands. In technical terms, getting appropriate performance characteristics wasn't cheap at all.

Shortly thereafter, flash storage started on its road into the enterprise. With the ability to eat IOPS faster than anything previously on the market, flash became a go-to technology for virtual desktops. It was used in many different ways, but flash carried its own baggage on the VDI journey. First, some of the flash-based solutions added complexity to storage, and second, all flash systems tended to be expensive.

## Second-Class Citizenship for Data Protection

Protecting VDI environments was also a challenge. The nature of VDI didn't always mean that it would enjoy the same kinds of data protection services as server workloads, even though desktop computing really is a critical service. Between WAN bandwidth and backup storage needs, fully protecting the desktop environment wasn't always feasible.

### It's All About That Scale

Scaling VDI was, again, a far different chore than scaling server-centric workloads. Whereas server workloads were scaled based on individual resource need, VDI-based workloads scaled far more linearly, requiring RAM, compute, and storage to scale simultaneously.

### The User Experience Trumps All

Finally, let's talk about the user experience. In a perfect VDI world, you have persistent virtual desktops in which users' settings and experience are maintained between sessions. This is the scenario that most closely mimics the real desktop experience, and people like it. With legacy infrastructure, getting the performance and capacity needed to support persistent desktops can be a real challenge.

Many gave up on VDI, thinking that they would never be able to enjoy their dreams of an efficient desktop environment. But then something interesting happened. Hyperconverged infrastructure hit the market.

## Hyperconvergence and VDI Scaling and Performance

As mentioned earlier in this chapter, VDI became one of the original primary use cases for the introduction of hyperconverged infrastructure into a company. It's not hard to see where hyperconvergence solved just about all of the challenges — real and perceived — around VDI.

First, let's talk about the ability for hyperconverged infrastructure to scale. You learned earlier that hyperconvergence natively enables linear resource scalability, which is also necessary for VDI environments to be able to keep pace with growth. As you add virtual desktops, you need to assign both CPU cores and RAM to those systems along with sufficient storage for the operating system, applications, and user files.

Performance is one of the big challenges in VDI, particularly as it relates to storage. With most hyperconverged infrastructure systems, you're getting a combination of flash storage and spinning disk. The flash layer is used to make everything faster while the spinning disk allows you to store user files on media designed for capacity. You get the best of both worlds with hyperconverged storage systems based on hybrid storage.

Further, with hyperconverged systems that have deduplication and compression features at the storage layer, you get even more benefits. Virtual desktops are all very similar, so they are very easily reduced at the storage layer. With reduction, you're able to store more virtual machines on the storage that exists in your hyperconverged infrastructure, which saves you a lot on disk costs. Deduplication and compression is the key technology that enables the use of persistent desktops in a VDI environment. Deduplication also massively reduced the I/O footprint for VDI systems. Being able to efficiently cache deduplicated desktop systems can virtually eliminate the various storms – boot storms and login storms – that can negatively impact performance otherwise.

### Cache Is More Efficient When Deduplicated

*By Brian Knudtson*

In hyperconverged systems with full inline deduplication across all workloads, the data management layer tracks all of the individual references to the unique blocks that have been written to the hard disk drives (HDDs) in metadata.

This deduplication extends to data that is stored in cache. When placed into cache, a block is read from the hard disk drive (HDD), incurring a Read IO, and a copy is placed on the SSD drives. Although the block could be retrieved for a

single request from a single virtual machine (VM), it could be requested again by the same VM (e.g., a block used by multiple files) or even requested by a different VM on the same host (e.g., a core Windows file). When the second request comes looking for the same block, it has already been placed in cache. Now you have two VMs benefiting from a single cached block. Extend this example to 10 VMs, and we have a single block in cache that could be worth 10 or more blocks in a non-deduplicated environment.

Continuing this example, the nine additional blocks that were accessed directly from cache generated no I/O to the back-end disks. That's nine less HDD IOPS consumed that are now available for another read or write operation that hasn't been cached already.

*Deduplication in cache*

Imagine the benefits that a VDI environment could realize during a boot storm. All the VMs are based on the same template, and therefore they all have the same set of files during initial boot. Normally, 100 VMs all booting at the same time would require a significant number of HDDs, but with this hyperconverged infrastructure platform, the first VM to boot reads the block off the HDD, which promotes that block into cache. Now the next 99 VMs can all access that same block from cache. That's a 100:1 IOPS reduction on the IOPS-bound disks.

Some hyperconverged systems have caching solutions that utilize RAM as a tier. In such cases, you can now apply all the same benefits just discussed and use a speed of memory-based cache. Blocks used by 100 VMs brought into memory cache at the cost of only a single VM can bring a significant performance increase to a VDI environment.

Another huge advantage for cache that comes from tracking all data as metadata is more intelligent cache-warming algorithms. Instead of simply grabbing the next block on the disk and betting on previous writes being sequential (see the IO Blender effect), the hyperconverged infrastructure nodes will calculate predictive caching based on the metadata. This leads to a much more intelligent and successful predictive cache approach.

This is just one of many advantages hyperconverged infrastructure can provide. This data efficiency, when applied to cache, not only improves the space utilization of cache, logically providing a larger cache, but also prevents read operations from going to the HDDs. It helps you improve your application performance and realize, on average, 40:1
data efficiency.

Let's not forget about data protection and availability. In a traditional desktop environment, fully protecting workstations can be a tough task and, in the event that a workstation happens to fail, a user could be without a computer for an extended period of time. In a VDI environment, if a user's endpoint fails, it can be very quickly replaced with another endpoint —the user simply reestablishes a connection to the persistent desktop.

But data protection in VDI goes way beyond just making it easy to get users back up and running. In fact, it comes down to being able to fully recover the desktop computing environment just like any other mission-critical enterprise application. In a hyperconverged infrastructure environment with comprehensive data protection capabilities, even VDI-based desktop systems enjoy backup and replication for users' persistent desktops. In other words, even if you suffer a complete loss of your primary data center, your users can pick right up where they left off thanks to the fact that their desktops were replicated to a secondary site. Everything will be there — their customizations, email, and all of their documents.

## About the Author

### Scott D. Lowe, vExpert

Scott Lowe is a vExpert and partner and Co-Founder of ActualTech Media. 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 has written thousands of articles and blog postings and regularly contributes to *www.EnterpriseStorageGuide.com* & www.*ActualTech.io*.

## About the Editor

### David M. Davis, vExpert

David Davis is a partner and co-founder of ActualTech Media. With over 20 years in enterprise technology, he has served as an IT Manager and has authored hundreds of papers, ebooks, and video training courses. He's a 6 x vExpert, VCP, VCAP, & CCIE# 9369. You'll find his vSphere video training at *www.Pluralsight.com* and he blogs at *www.VirtualizationSoftware.com* and *www.ActualTech.io*.

# About ActualTech Media

ActualTech Media provides enterprise IT decision makers with the information they need to make informed, strategic decisions as they modernize and optimize their IT operations.

Leading 3rd party IT industry influencers Scott D. Lowe, David M. Davis and special technical partners cover hot topics from the software-defined data center to hyperconvergence and virtualization.

Cutting through the hype, noise and claims around new data center technologies isn't easy, but ActualTech Media helps find the signal in the noise. Analysis, authorship and events produced by ActualTech Media provide an essential piece of the technology evaluation puzzle.

More information available at **www.actualtechmedia.com**

Hewlett Packard
Enterprise

## About Hewlett Packard Enterprise

Hewlett Packard Enterprise (HPE) is an industry leading technology company with the industry's most comprehensive portfolio, spanning the cloud to the data center to workplace applications. HPE technology and services help customers around the world make IT more efficient, more productive and more secure.

Early in 2017, the company acquired SimpliVity and now offers HPE SimpliVity hyperconverged systems, complete hardware-software solutions that are designed, built, and supported by HPE.

Visit *www.hpe.com*

# Gorilla Guide Features

### In the Book

These help point readers to other places in the book where a concept is explored in more depth.

### The How-To Corner

These will help you master the specific tasks that it takes to be proficient in the technology jungle.

### Food For Thought

In the these sections, readers are served tasty morsels of important information to help you expand your thinking .

### School House

This is a special place where readers can learn a bit more about ancillary topics presented in the book.

### Bright Idea

When we have a great thought, we express them through a series of grunts in the Bright Idea section.

### Dive Deep

Takes readers into the deep, dark depths of a particular topic.

### Executive Corner

Discusses items of strategic interest to business leaders.