THE GORILLA GUIDE™ TO...

Hyperconverged Infrastructure for Remote Office / Branch Office

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 Remote Office / Branch Office

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.

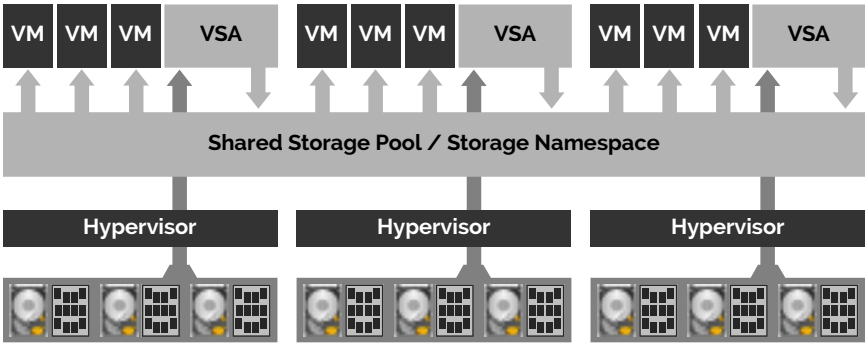

Figure 1-1: An overview of a Virtual Storage Appliance

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

Remote Office/ Branch Office

Remote Office and Branch Office (ROBO) IT can create some pretty tough situations for IT pros to conquer. Perhaps the most significant problem is one of scale. Many ROBO environments have the need to grow very large, but need to do so by remaining very small. Consider a fictional company that has 500 locations. This is a big company, so the overall aggregate technology needs of the organization will be significant. At the same time, though, each branch location in this company supports 20 employees and is also a sales outlet. The individual sizing needs of each branch are relatively modest.

Traditional ROBO Challenges

At first look, it might seem like a simple solution. Throw a couple of servers into each location and call it a day. Unfortunately, it's not that easy. There's a lot more to the scenario than meets the eye.

First and foremost, just a couple of servers may not meet the needs of the branch office. Each branch is probably sized a little differently, so some may be able to operate with just a couple of servers while others may need more substantial storage capabilities. You'll

probably want two or more servers just so you can have some level of availability. If one server fails, the other one can pick up the load. Getting high availability with just two servers, while solvable, isn't always straightforward.

At the same time, you have to keep an eye on performance to make sure that poorly performing local applications don't negatively impact the branch's business. You can't forget about data protection, either. If this was a single-site company, data protection would be relatively easy; you just back data up to a tape, disk, or a second location. But if you have many sites and some have slow Internet links, it can be tough to protect data in a way that makes sense. You don't want to have local IT staff that needs to change tapes, or watch backup appliances. You also don't want to have non-technical people trying to do this as a part of their jobs. It doesn't always work out well.

Plus, there's ongoing support. Stuff happens. You need to be able to keep every site operational. However, with each site you add (each with its own unique needs), the overall complexity level can become overwhelming. As complexity increases, efficiency decreases and it becomes more difficult to correct problems that might occur. **Figure 2-1** provides a demonstrative overview of today's data center. In many ROBOs, centralized IT delivers services to the remote sites from a centralized location over a WAN. By centralizing IT, the company eliminates the cost of skilled IT staff on site at remote sites and reduces the risk to business continuity since IT handles data protection. However, the major drawbacks are often poor application performance, scattered management, and difficulty correcting issues that arise in remote sites.

To summarize the challenges faced in ROBO environments:

- There is a need for a lot of decentralized systems to support individual branch offices, and there is often lack of a cohesive management platform.

- Bandwidth to branch offices can often be limited and may not be reliable. Most ROBO sites lack the full breadth of data center services (high-performance storage, WAN accelerators, etc.) enjoyed by headquarters and by single-site companies.

- Data generated at branch offices needs first-class citizen protection, but often can't get it using legacy tools.

- Hardware at branch offices might run the gamut from just a server or two to a full cluster with a SAN, but most companies want to be able to have minimal hardware at branch locations when possible and need to be able to centrally manage solutions.

- There is a lack of technical personnel at most ROBO locations and companies don't want to have to hire dedicated technical staff for each one.

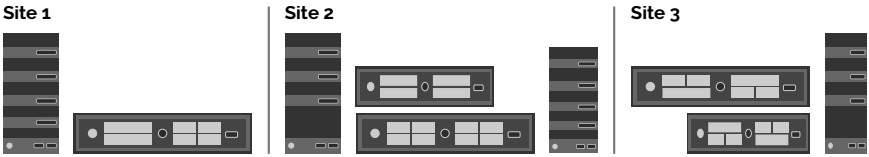

Figure 2-1: Chaos is the norm in many ROBO environments

Without some kind of change, the dystopian future for the ROBO will be so challenging that even Katniss Everdeen would call it quits and hang up her bow and arrow.

Transforming ROBO Operations with Hyperconverged Infrastructure

ROBO operations is an area in which the right hyperconverged infrastructure solution has the potential to completely transform the environment and how that environment is managed. The overall results can be lower costs, improved efficiency, and better overall disaster recovery capabilities.

So what does it take to achieve this ROBO utopia and how does hyperconverged infrastructure fit into the equation?

Keeping IT Simple

Hyperconverged infrastructure brings simplicity to chaotic IT organizations and nothing says "chaotic" like dozens of different sites running disparate hardware managed as individual entities. By moving to a common hyperconverged infrastructure platform, you instantly gain centralized administrative capabilities that encompass every site. Moreover, when it comes to hardware support, every site becomes a mirror of the others, thereby streamlining your support efforts. Such an architecture eliminates the need for dedicated technical staff at each branch.

The need to keep management simple cannot be overstated. Companies are no longer willing to scale IT staff at the same rate that they add sites and services, but they expect consistent performance. To solve this seeming paradox of intentions, IT has to look at the ROBO environment much more discerningly and deploy solutions. They need to choose solutions that unify management across all ROBO sites in a way that makes them appear as if they're a single entity even while they support a dispersed organization.

Less Hardware

Some sites need very little hardware while others need more. Some sites traditionally needed dedicated storage while others didn't. It's chaos. With the right hyperconverged infrastructure solution, you can have sites that operate on just one or two appliance-based nodes without having to compromise on storage capacity and performance. You simply deploy the two nodes, and they form their own cluster that imbues the branch with highly efficient storage capabilities that include comprehensive data deduplication and reduction. For larger sites, you simply add more nodes. No SAN is needed and all of the hardware across all of the sites is common, easy to support, and provides enterprise-level capabilities, even in a single-node or two-node cluster.

The data reduction features available in some hyperconverged infrastructure solutions mean that you don't need to constantly add storage. With reduction, you get to cram more data into the same amount of overall capacity at the branch site. Reduction also has other benefits. Read on.

Comprehensive Data Protection

Data generated or managed at branch sites needs to be treated just like data generated at HQ. In many cases, the data at branch sites is even more important because it's the information that's created as the result of sales or other customer-facing efforts. With a hyperconverged infrastructure solution that has the ability to fully compress and deduplicate data and that can work with data in its reduced form, you can get data protection capabilities that allow you to replicate branch office data to other branches or to headquarters even over slow WAN connections. Better yet, you don't need WAN accelerators to accomplish this feat. With the right solution, reduced data is transferred over the wire and, even then, only the blocks that don't already exist at the target site are transferred, resulting in an incredibly efficient process. This kind of data protection infrastructure also eliminates the need for on-site staff to

perform tasks such as changing tapes and increases the potential for successful recovery in the event of a disaster. In **Figure 2-2**, you see a nice, neat, and streamlined infrastructure.

Figure 2-2: Hyperconverged infrastructure tames the ROBO chaos

Deployment Options

As you're deploying ROBO solutions using hyperconverged infrastructure, you need to take a look at how you want your ROBO sites configured. There are two typical models available:

- **Hub and spoke** (**Figure 2-3**). With this architecture, there is a centralized hub in the center and each remote site is at the end of a spoke. With this model, the various remote sites will generally talk to the hub spoke, but not often with each other. Backups and other data transfer operations will generally flow from the end of one of the spokes back the hub.

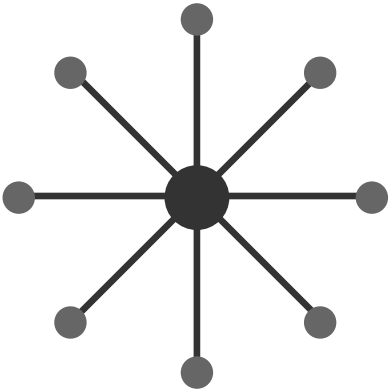

Figure 2-3: A look at a hub and spoke ROBO model

- **Mesh** (**Figure 2-4**). In a mesh environment, all of the sites can talk directly to the other sites in the mesh. Under this model, it's possible to have individual sites back up to each other and the organization can, theoretically, operate without a centralized hub, although one of the nodes often acts in this capacity.

As you're deploying hyper-convergence throughout your organization, it's important to ensure that the intended solution can easily support whichever deployment model you use, even if it happens to be a combination of the two. Most importantly, regardless of which model you use, you should be able to centrally manage everything and have the ability to implement data protection in whatever way makes the most sense for you. Finally, adding new sites – scaling the environment – should be a basic feature of the solution and not a complicated afterthought.

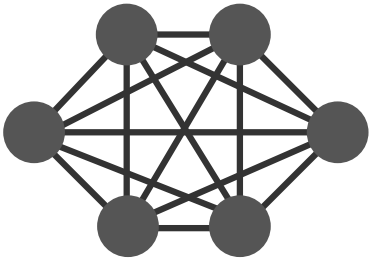

Figure 2-4: A look at a mesh-based ROBO model

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.