THE GORILLA GUIDE TO...™

# Hyperconverged Infrastructure for Cloud

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

Brought to you by
Hewlett Packard Enterprise

*Helping you navigate the technology jungle*

# Hyperconverged Infrastructure for Cloud

**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

**Chapter 1:** Introduction to Hyperconverged Infrastructure **3**

Hyperconverged Infrastructure from 30,000 Feet 6
Resources to Consolidate 8

**Chapter 2:** Hyperconvergence & the Public Cloud **11**

Why Is Cloud so Desirable? 11
The Public Cloud 13
On-Premises Reality 16
Private Clouds 17
Hybrid Cloud 19
The Intersection of Cloud and Hyperconverged Infrastructure 19

**About the Author** **xxii**

**About the Editor** **xxii**

**About ActualTech Media** **xxiii**

**About Hewlett Packard Enterprise** **xxiv**

**Gorilla Guide Features** **xxv**

# 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.

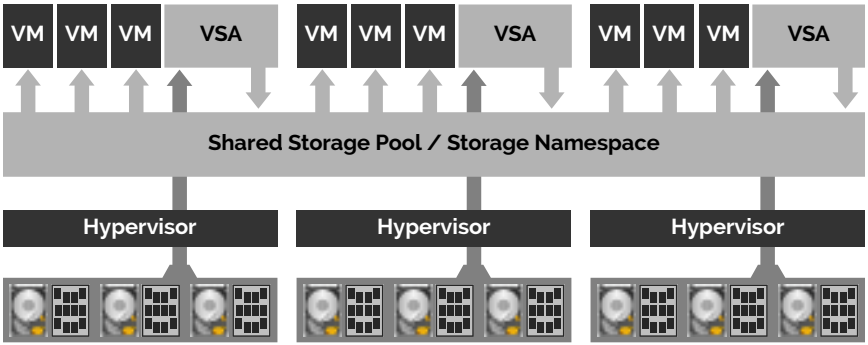


**Figure 1-1:** An overview of a Virtual Storage Appliance

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

# Hyperconvergence & the Public Cloud

This chapter will help you understand the ways by which you can leverage cloud services as a part of your hyperconverged infrastructure solutions. It will also help you better understand the concept of "private cloud" and how that fits with hyperconvergence. Can a hyperconverged solution deliver some of the things that cloud can give you?

## Why Is Cloud so Desirable?

You will learn more about what defines *cloud* a little later in this chapter. Before getting into the various definitions, though, let's discuss the traits inherent in cloud systems which make them a popular and desirable choice for service deployment.

### The Economic Model

Everything eventually comes down to money. Business decision makers are constantly on the lookout for ways to reduce costs while also boosting efficiency and outcomes. This is often a seemingly impossible task described as "doing more with less." IT was supposed to be an enabler, but for many companies, it has become a money

pit — an expense center to be minimized. Obviously, when leveraged properly, IT can be an incredible enabling function, but even in these cases, no one wants to spend more than they have to.

When you buy your own data center hardware and software, you incur pretty significant CapEx. This initial cash outlay necessary to procure a solution can be pretty high and can result in the need to cut corners or even delay upgrades if there is not enough cash available.

When you decide to start consuming resources from the public cloud, there is no initial cash outlay necessary. You don't incur capital expenses. Sure, you might have to pay a bit in the way of startup costs, but you don't have to buy hardware and software. You simply rent space on someone else's servers and storage.

Business decision makers love this shift. They don't need to worry about huge capital outlays, and they know that they're paying for what they use. They're not paying for extra hardware that may never end up actually being leveraged to help solve business needs.

### Scale

When you build your own data center, you have to scale it yourself. Sometimes, you can scale in increments that make financial sense, while other times you have to add more than you might like due to predefined requirements from your vendors.

When you use the public cloud, you don't have to worry about inherent scaling limits or increments. Remember, you pay for what you use. As your usage grows, so does your bill, but you don't generally need to manually add new resources to your account. It can happen automatically.

Scalability granularity often isn't a problem with the public cloud. You grow as you need to. There is no practical limit to how far you can grow as long as the cloud provider still has resources.

### Geographic Diversity and Disaster Recovery

Building multiple data centers can be an expensive undertaking, but it's one that is being executed more and more as companies seek ways to protect their data and ensure continuity of their business in the event of a disaster striking the primary data center. The separate data centers are generally geographically diverse so that a single natural disaster can't take out both sites at the same time.

Public cloud providers often already have systems that can quickly enable geographic diversity for applications that are already running on their systems. Enabling geographic diversity is often as simple as clicking a mouse button and, most likely, paying some additional money to the cloud provider.

## The Public Cloud

It's hard to avoid the term *cloud* today. It's everywhere. For many, the term itself has become synonymous with "Internet" or is just another way to described what used to be called "hosted services." However, there are a number of traits that make a public cloud a public cloud.

First, in general, public cloud systems are comprised of multi-tenant environments operated by a service provider with the hardware and software located in the provider's data center. In these environments, the customer may not always even be aware in which provider data center the services reside, nor does the customer have to be aware. The beauty of these systems is that workloads can move around as necessary to maintain service level agreements.

Cloud service providers generally build their systems with the assumption that hardware will likely fail, which means the you, as the customer, can avoid the need to buy expensive failover and availability systems on your own.

For scale, the cloud provider can provide grid-like scalability to great levels so that you don't need to worry about how to grow when the time comes.

**For public cloud, there are a number of pros and cons to consider. On the plus side, cloud will:**

- Enable immediate implementation.
- Carry low to no initial deployment costs.
- Provide a consumption-based utility cost model.
- Provide more cost effective scale than would be feasible in a private data center.

**However, there are definitely some downsides to cloud as well, which include:**

- Potentially unpredictable ongoing usage charges
- Concerns around data location; many do not want data stored in US-based data centers due to concerns around the NSA and PATRIOT Act
- Charges across every aspect of the environment, from data storage to data transfer and more
- No control over underlying infrastructure
- Care needs to be taken to avoid lock-in

## The faces of the public cloud

**Here is a brief look at the different kinds of public cloud services that are available on the market.**

### Software-as-a-Service (SaaS)

From a customer perspective, software-as-a-service (SaaS) is the simplest kind of cloud service to consume as it is basically an application all wrapped up and ready to go. Common SaaS applications include Salesforce and Office 365.

With SaaS applications, the provider controls everything and provides to the customer an application layer interface that only controls very specific configuration items. Because all of the infrastructure and the fact that most of the software is hidden from the you as the customer, you don't need to worry about any underlying services except those which may extend the service, such as integrating Office 365 with your on premises Active Directory environment.

### Platform-as-a-Service (PaaS)

Sometimes, you don't need or want a complete application. In many cases, you just need a place to install your own applications but you don't want to have to worry at all about the underlying infrastructure or virtualization layers. This is where platform-as-a-service (PaaS) comes into play.

PaaS provides you with infrastructure and an application development platform that gives you the ability to automate and deploy applications including your own databases, tools, and services. As a customer, you simply manage the application and data layers.

### Infrastructure-as-a-Service (IaaS)

In other cases, you need a bit more control, but you still may not want to have to directly manage the virtualization,

storage, and networking layers. However, you need the ability to deploy your own operating systems inside vendor-provided virtual machines. Plus, you want to have the ability to manage operating systems, security, databases, and applications.

For some, infrastructure-as-a-service (IaaS) makes the most sense since the provider offers the network, storage, compute resources, and virtualization technology while you manage everything else.

## On-Premises Reality

Even though public cloud has a number of desirable traits, there are some harsh realities with which CIOs and IT pros need to contend:

- **Security** – For some, particularly those in highly-regulated or highly-secure environments, the idea of moving to a multi-tenant public cloud is simply not feasible.

- **Bandwidth** – Many areas of the world remain underserved when it comes to bandwidth, and companies can't get sufficient bandwidth with sufficiently low latency to make cloud a feasible option.

- **Cost** – There may come a point at which cloud may become more expensive than simply building your own environment.

These challenges are reasons that many organizations are turning to private cloud environments.

## Private Clouds

The term *private cloud* is often, well, clouded in confusion as people try to apply the term to a broad swath of data center architectures. So, let's try to clear up some of the confusion.

First and foremost, a private cloud environment generally resides in a single tenant environment that is built out in an on-premises data center, but it can sometimes consist of a single tenant environment in a public data center. For the purposes of this chapter, we'll focus on the on-premises use case.

Private cloud environments are characterized by heavy virtualization which fully abstracts the applications from underlying hardware components. Virtualization is absolutely key to these kinds of environments. Some companies go so far as to offer internal service level agreements to internal clients in a cloud-like manner. The key phrase there is "internal clients" — that is the customer in a private cloud environment. For such environments, being able to provide service level guarantees may mean that multiple geographically dispersed data centers need to be built in order to replicate this feature of public cloud providers.

Heavy use of virtualization coupled with comprehensive automation tools reveals an additional benefit of private cloud: self-service. Moving to more of a self-service model has two primary benefits:

- Users get their needs serviced faster

- IT is forced to build or deploy automation tools to enable self-service functionality, thereby streamlining the administrative experience

As mentioned before, many companies want to keep their data center assets close at hand and in their full control, but they want to be able to gain some cloud-like attributes, hence the overall interest in private cloud. As is the case with public cloud, there are a number of pros and cons that need to be considered when building a private cloud.

**In the pros column, private cloud:**

- Provides an opportunity to shift workloads between servers to best manage spikes in utilization in a more automated fashion.

- Enables ability to deploy new workloads on a common infrastructure. Again, this comes courtesy of the virtualization layer.

- Provides full control of the entire environment, from hardware to storage to software in a way that enables operational efficiency. In other words, routine tasks are automated and repeatable.

- Allows customers to customize the environment since they own everything.

- Provides additional levels of security and compliance due to the single tenant nature of the infrastructure. Private cloud-type environments are often the default due to security concerns.

As with everything, not all is a perfect picture. Private clouds do have a number of drawbacks, including:

- Requiring customers to build, buy, and manage hardware. This is often something that many companies want to reduce or eliminate.

- Not always resulting in operational efficiency gains.

- Not really providing what is considered a cloud computing economic model. You still have to buy and maintain everything.

- Potentially carrying very high acquisition costs.

In short, private clouds are intended to have some of the architectural characteristics of public clouds while offering internal clients cloud-like economic outcomes when chargeback processes are implemented. Even if the central IT department providing the service doesn't really use "the cloud," as internal clients are able to provision and consume resources on demand — at least to a reasonable point — there is the beginning of a private cloud.

## Hybrid Cloud

Increasingly, people are choosing both cloud options – public and private – to meet their needs. In a hybrid cloud scenario, the company builds its own on-premises private cloud infrastructure to meet local applications needs and also leverages public cloud where reasonable and possible. In this way, the company gets to pick and choose which services run where and can also move between them at will.

## The Intersection of Cloud and Hyperconverged Infrastructure

If you're wondering what all of this talk about cloud has to do with hyperconverged infrastructure, wonder no more. Depending on the hyperconverged infrastructure solution you're considering, there are varying degrees of association between the hyperconverged infrastructure product and both public and private clouds.

### Economics

Everything you've read so far leads to money. The potential to completely transform the data center funding model is one of the key outcomes when you consider hyperconverged infrastructure. With easier administration comes lower staffing costs. With the use of commodity hardware comes lower acquisition costs. With the ability to scale linearly in bite-size chunks, companies can get the beginnings of a consumption-based data center equipment acquisition model that enables closer to pay-as-you-go growth than traditional data center architectural models. As your environment needs to grow and as users demand new services, you can easily grow by adding new hyperconverged systems.

### Scale

Agility implies some level of predictability in how workloads will function. Public cloud provides this capability. For those wishing to deploy a private cloud environment, these needs can be met by leveraging hyperconvergence's inherent ability to scale linearly (meaning, by scaling all resources including compute, storage, and networking simultaneously). In this way, you avoid potential resource constraint issues that can come from trying to manually adjust individual resources and you begin to achieve some of the economic benefits that have made public cloud a desirable option.

Scaling the data center should not result in scaling the complexity. In order to attain the full breadth of economic benefits that go with cloud, you have to make sure that the environment is very easy to manage or, at the very least, that management is efficient. This means that you need to automate what can be automated and try to reduce the number of consoles that it takes to get things done.

With hyperconverged infrastructure, management efficiency – even at scale – is a core feature of the solution. You are able to manage

all of the elements included in the product from a single console and you are also able to apply a breadth of consolidated policies to virtual machines.

### Geographic Diversity and Disaster Recovery

Also on the economics front, the value of disaster recovery cannot be overstated. One of the benefits of the cloud is the geographic diversity that can be achieved to protect against natural disasters. With a hyperconverged infrastructure solution that has data replication as a part of the core offering, multisite redundancy capability is baked in as part of the solution.

For those that have opted to build hybrid clouds, some hyperconverged infrastructure solutions can leverage that public cloud deployment as a replication target. In other words, rather than going to the expense of building out a second physical site, the public cloud can be used to achieve data protection goals.

### Hyperconvergence and the Private Cloud

Building a traditional private cloud is hard. It takes a lot of work to get all the pieces aligned. Hyperconverged infrastructure can allow you to deploy private clouds in a fraction of the time it would normally take. Everything is built into the individual appliances, including centralized management, data efficiency, replication, and the ability to scale in incremental units. These are core needs in building an agile private cloud environment.

## 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.