HPE PRESENTS

The Gorilla Guide to...®

FOURTH EDITION

Key Hyperconverged Infrastructure Use Cases

Scott D. Lowe

INSIDE THE GUIDE:

- Learn about the key uses cases that are driving adoption of hyperconverged infrastructure
- Discover how hyperconvergence has become an ideal environment for edge computing needs
- Find out how hyperconvergence can help align IT to better meet the needs of the business

Helping You Navigate The Technology Jungle!

ActualTech Media
www.actualtechmedia.com

In Partnership With

Hewlett Packard Enterprise

THE GORILLA GUIDE TO...

Key Hyperconverged Infrastructure Use Cases

By Scott D. Lowe

ACTUALTECH MEDIA

6650 Rivers Ave Ste 105 #22489

North Charleston, SC 29406-4829

www.actualtechmedia.com

PUBLISHER'S ACKNOWLEDGEMENTS

EDITOR
Keith Ward, ActualTech Media

PROJECT MANAGER
Wendy Hernandez, ActualTech Media

EXECUTIVE EDITOR
James Green, ActualTech Media

LAYOUT AND DESIGN
Olivia Thomson, ActualTech Media

WITH SPECIAL CONTRIBUTIONS FROM
Charles Wood

ENTERING THE JUNGLE

CALLOUTS USED IN THIS BOOK

The Gorilla is the professorial sort that enjoys helping people learn. In the School House callout, you'll gain insight into topics that may be outside the main subject but are still important.

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

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

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

Discusses items of strategic interest to business leaders.

ICONS USED IN THIS BOOK

DEFINITION

Defines a word, phrase, or concept.

KNOWLEDGE CHECK

Tests your knowledge of what you've read.

PAY ATTENTION

We want to make sure you see this!

GPS

We'll help you navigate your knowledge to the right place.

WATCH OUT!

Make sure you read this so you don't make a critical error!

CHAPTER 1

Introduction to Hyperconverged Infrastructure

The world of technology is changing at both a broader and faster pace than ever before. In years past, IT practitioners and decision makers might have had to deal with a few new hardware options and some software upgrades. Today, changes come every day as companies break the data center walls in favor of hybrid cloud, as cloud providers continue to grow, as software updates deploy silently behind the scenes, and as entire new architectures come and go.

As employees devour technologies such as smartphones, tablets, wearables, and other devices, and as they become more comfortable with solutions such as Slack, Teams, Dropbox, and Zoom, their demands on enterprise IT intensify. On top of this, management and other decision makers are also increasing their demands on enterprise IT to provide application environments that have higher levels of availability and easier growth capability, but with the kinds of pay-as-you-grow economics that they see from the cloud. Unfortunately, enterprise IT organizations often don't see much, if any, associated increases in funding to accommodate these demands, particularly as spending increases on other business priorities, such as improved security, analytics, remote work, and digital transformation efforts.

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 its 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.

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 quickly 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. Moreover, organizations find themselves increasingly turning to technology-centric solutions to solve real-world crises, placing further strain on IT resources.

Even as the public cloud continues to gain favor for certain workloads, the real center of workload action remains the on-premises data center. Unfortunately, traditional data center architectures actively work against modern goals, because with increasing complexity comes increased costs—and things have definitely become more complex. Virtualization was a fantastic opportunity for companies, but with virtualization came new challenges, including major issues with storage. With virtualization, enterprise IT 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.

Further, legacy data center infrastructure was often not very intelligent. Once appropriately cobbled together, the IT department was able to provide a generally suitable workload operating environment,

but it was sometimes operated via the *brute force* administrative method of bashing on keyboards and dragging mice to perform a multitude of manual operations. These legacy environments didn't have the kinds of software-centric, data-driven intelligence that now enables autonomous operation.

Here are some other challenges legacy data center admins 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's increasingly complex, decreasingly flexible, 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 extremely 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.
- **A focus on security.** Security has always been important, but never more than it is today. Increasingly a boardroom issue, security spending is accelerating, sometimes at the expense of other areas of IT, a situation that data center architects need to accommodate.

- **Taking efficiency to the next level.** Businesses today want an IT function focused on business workloads, not one that has to constantly wrangle storage into submission with esoteric scripts. Next-level efficiency includes imbuing infrastructure with an AI-led, data-driven operational framework that bludgeons the complexity out of the environment to enable faster results to help accelerate business initiatives.

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

Hyperconverged Infrastructure from 30,000 Feet

A number of years ago, a new data center architectural option, dubbed *hyperconverged infrastructure*, came on the scene and exploded faster than anyone could have imagined. Hyperconvergence is a 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 with modern iterations of the technology. This modern technology includes intelligence features that streamline overall management and bring in aspects of AIOps.

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, while providing an infrastructure that's self-managing.

Why are storage and compute at the core of hyperconverged infrastructure? Simply put, storage had become an incredible challenge for many companies. It's traditionally been one of—if not the—most expensive resources in the data center, often requiring a highly skilled person or team to keep it running.

Moreover, for many companies, storage became a single point of failure. When it failed, swaths of services were negatively impacted.

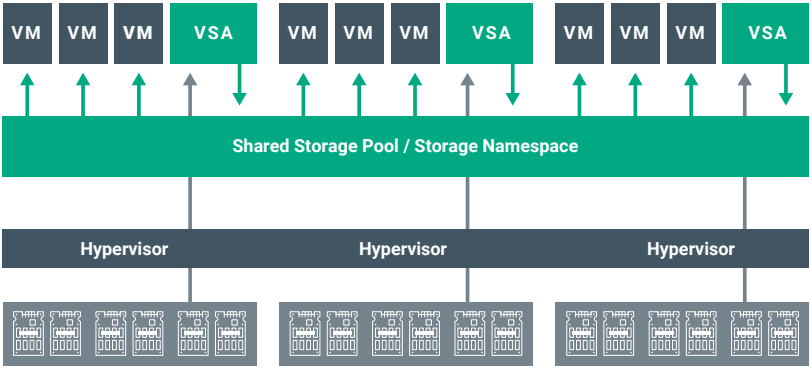

Figure 1-1: An overview of a Virtual Storage Appliance

Finally, with the increase in the growth of data volume, legacy storage architectures were beginning to crumble. This data explosion overwhelmed systems that simply weren't designed to handle the flood of information coming their way.

Combining storage with compute is in many ways a return to the past, but this time serious brains have been wrapped around it. Before virtualization and before SANs, many companies ran physical servers with direct-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 without a centralized coordinating mechanism. Virtualization solved this resource sharing problem, but introduced its own problems, which were discussed earlier.

Hyperconverged infrastructure distributes the storage resource among the various nodes that comprise a cluster. Built using standard 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 communicates with all of the other VSAs in the cluster over an Ethernet link, thus forming a distributed file system across which virtual machines (VMs) are run.

The fact that hyperconverged systems leverage standard off-the-shelf hardware is critical. The power behind hyperconverged infrastructure lies in its ability to corral resources—RAM, compute, and data storage—from hardware that doesn't have to be custom-engineered. This is the basis for hyperconverged infrastructure's ability to scale granularly, work more efficiently and save money.

Resources to Consolidate

The basic combination of storage, servers, and a hypervisor is the undergirding of all hyperconverged infrastructure. And the more hardware devices and software systems that can be collapsed into a hyperconverged solution, the easier it becomes to manage and less expensive to operate.

Here are some data center elements that can be integrated in a hyperconverged infrastructure. Often, these simply become invisible features of the solution and don't even require separate management.

Deduplication Appliances

In order to achieve the most storage capacity, deduplication technologies are common in today's data center. Standalone storage solutions and dedicated appliances are available that handle complex and CPU-intensive deduplication tasks, ultimately reducing the amount of data that has to be housed on primary storage.

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.hpe.com/HCI/fordummies

But even in these cases, deduplication isn't always 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. In essence, you may end up in a scenario in which data is being constantly reduced and rehydrated as it meanders around the environment. It's a terribly inefficient use of resources that could be better targeted to support more workloads.

SSD Caches, Hybrid Storage, and All-Flash Arrays

To address storage performance issues, companies increasingly deploy either solid-state disk (SSD)-based caching systems, hybrid storage arrays, or full SSD/flash-based storage arrays. However, all of these 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's 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 service provided by 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, "The Ins and Outs of Backup and Recovery"—are both shrinking metrics upon which IT needs to improve.

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, some baseline solutions include comprehensive backup and recovery capability that can enable extremely short RTO windows, enabling very small RPO metrics—both *very* good characteristics!

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 replication 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 in traditional scenarios, companies implement a variety of other data center services. For example, to minimize replication impact on bandwidth, companies deploy WAN-acceleration 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 operation costs 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.

The Ins and Outs of Backup and 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 **The Gorilla Guide To...® Hyperconverged Infrastructure Implementation Strategies**.

Up Next

You now have an understanding of hyperconverged infrastructure and knowledge about many of the resources that can be consolidated into such solutions. In the following chapters, you'll learn about a number of ways that you can begin using this knowledge in order to solve some of your most challenging business problems.

CHAPTER 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 for hyperconverged infrastructure was virtual desktop infrastructure (VDI), a use case that's still strong today.

VDI is an interesting solution. Like many trends in IT, VDI has gone through the Gartner Inc. "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?

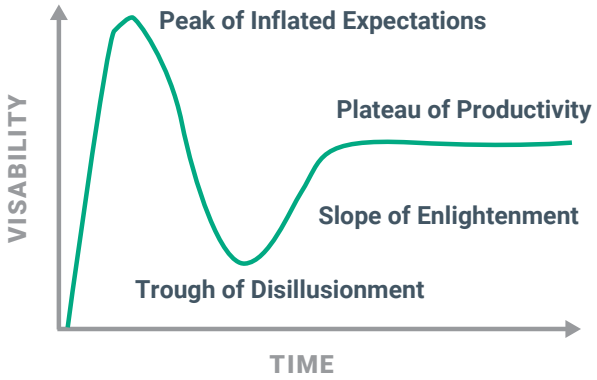

Figure 2-1: The Gartner Inc. Hype Cycle (courtesy of Wikipedia)

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. And, really, who wants an "adequate" experience when an awesome one is possible?

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 kinds 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-to-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 aren't maintained between sessions. Each time a user logs in, it's as if they have 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 log in. This is because a lot of information has to be read from disk and placed into memory on a traditional PC. There are also copious write operations taking

place, including writing all kinds of information to logs, including 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. This can quickly overwhelm storage.

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 catapulted VDI directly to the Gartner 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 its journey into the enterprise. With the ability to eat IOPS faster than anything previously on the market, flash has become a go-to technology for virtual desktops. For quite some time, 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, although, in recent years, the price of flash storage has plummeted as the technology gains traction and economies of scale come into play.

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 today's hyperconverged infrastructure systems, you're likely getting an all-flash system, although there are products on the market that combine flash and spinning disk in a hybrid storage configuration.

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're very easily reduced at the storage layer. With reduction, you're able to store more VMs on the storage that exists in your hyperconverged infrastructure, which saves you a lot on disk costs. Deduplication and compression are the key technologies that enable 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.

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.

Overcoming Challenges and Operationalizing VDI

VDI sounds like a great idea, but many that embark on a solo journey to deploy the technology quickly discover that it's not as easy as deploying a few VMs—a number of inherent challenges have to be overcome.

First is budget. For VDI to work, you need hardware and software. On the hardware front, you need server hosts on which virtual desktops will run. You may also need user endpoints, if you don't already have a suitable fleet of such devices. The server side is handled elegantly with a hyperconverged infrastructure solution, as discussed earlier. In terms of hardware, you may have high-end graphics requirements, which is historically one of the more vexing challenges associated with VDI.

For software, you need a hypervisor, brokers, gateways, and other constructs that provide a secure, scalable VDI deployment. You also need to make sure you can quickly recover your desktop environment if disaster strikes.

Beyond providing a robust, linearly scalable environment in which to run your virtual desktops, HPE SimpliVity provides complete data protection, including support for HPE SimpliVity RapidDR, to help automate recovery in the event of a failure. Further, your

infrastructure is managed under the watchful eye of HPE InfoSight, providing robust intelligence and operational benefits.

HPE provides an array of HPE SimpliVity offerings for VDI, each offering solutions to the challenges described. If you have heavy graphics needs, for example, HPE SimpliVity 380 or 2600 have you covered with GPU support. Even better, thanks to the right integration in the deployment, you get upgrade capabilities from a single location that includes firmware, hypervisor, and SimpliVity software updates.

The turnkey nature of the HPE SimpliVity platform means that any budget challenges that may exist aren't exacerbated by poor implementation. And the platform choice ensures that you can find a VDI solution that matches your budget.

If you prefer an OpEx financing model over a CapEx model, consider consuming HPE SimpliVity under HPE's GreenLake pay-as-you-go program. HPE GreenLake offers "hyperconvergence as a service" and can help you accelerate your deployment efforts while using a pay-as-you-go model on the financial side.

Up Next

With easy scalability, excellent performance capabilities, and great data protection features, hyperconverged infrastructure has become a natural choice for VDI environments. Next is a deeper look into how this architecture can address edge scenarios.

CHAPTER 3

Supporting Edge Computing

Edge IT demands can be difficult, and perhaps the most significant problem is one of scale. The technology requirements for many edge environments can grow very large over time, but the infrastructure footprint in each location must remain very small. (**Figure 3-1** shows how this growth can occur haphazardly and, therefore, inefficiently).

Consider a fictional company that has 500 locations. This is a big company, so the overall aggregate technology needs of the organization are significant. At the same time, each branch location supports 20 employees and is also a sales outlet. The individual sizing needs of each branch are relatively modest, but that doesn't mean there are insignificant needs at each location—in fact, just the opposite is true.

As organizations look for ways to streamline and enhance the customer experience, the technology requirements in each locale will increase.

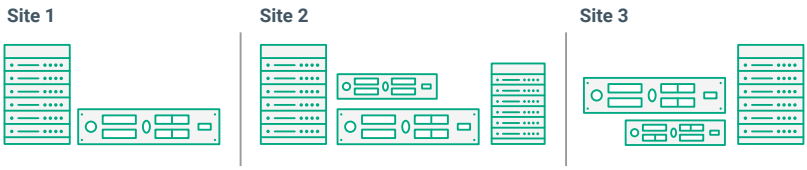

Figure 3-1: Chaos is the norm in many edge environments

Traditional Edge Challenges

At first look, the solution might seem simple—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.

Exploding Capacity Needs

For many organizations, edge environments are the front line in terms of data acquisition. Expanded product information and processes around point-of-sale systems are pretty obvious potential generators of data, but they pale in comparison to the vast quantity and variety of data being vacuumed up by Internet of Things (IoT) devices and other sensors. These devices can include cameras, both for security purposes and for customer experience improvements.

As you look at the number and variety of edge deployments, it becomes even clearer how the edge can dramatically impact capacity requirements. With manufacturing equipment today containing sensors for all aspects of operations, and beds being just one of dozens of devices in a hospital room generating a constant flow of data, the need for easily expanded capacity and a system that enables high levels of data efficiency becomes even more obvious.

It's more than just gathering that data, though. Edge computing environments are also about keeping that data closer to the source so that it can be acted on in some way. And that calls for increasing usage of compute and advanced applications capabilities, too. It's easy to see that there's a need for intelligence at the edge—and that intelligence needs to be just as capable as the intelligent systems deployed in central data center locations.

A Need for Differentiation

First and foremost, just a couple of servers may not meet the needs of the branch office. Each branch is probably sized differently, so some

may be able to operate with just a couple of servers while others may need more substantial capabilities. You'll probably want two or more servers for availability reasons. 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.

Constantly Advancing Needs

Further, you may have more robust needs in your edge locations and require more advanced functionality. As more retail environments work to reshape the customer experience, the market is seeing more technology adoption. Some stores are deploying futuristic systems based on myriad sensors and cameras placed strategically throughout the store and backed by powerful AI systems that help to reduce theft and streamline the checkout experience, sometimes not even requiring a visit to a checkout lane.

In these cases, just throwing regular servers into the fray won't solve the problem. Edge environments require the same suite of computing services as their data center–based cousins. This includes comprehensive data protection, an ability to run a variety of workloads, including virtual ones and those that run in containers, and the ability to leverage GPUs for modern, AI-driven applications.

Maintaining Performance

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, a second location, or the cloud. 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.

Lack of Onsite Technical Staff

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 3-1** provides a demonstrative overview of today's data center. In many edge sites, 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 edge environments:

- There's a need for a lot of decentralized systems to support individual branch offices and edge needs, and there's often lack of a cohesive management platform
- Bandwidth to edge locations can often be limited and may not be reliable. Most edge locations lack the full breadth of data center services (high-performance storage, WAN accelerators, and so on) enjoyed by headquarters and by single-site companies
- Data generated at edge locations needs first-class citizen protection, but often can't get it using legacy tools
- There is often a mashup of different hardware platforms at edge locations and, even in consistent environments, there may be more hardware than is necessary, just to maintain availability
- Hardware at edge locations might run the refresh-cycle gamut from aging devices to brand new components. Some sites have

just a server or two where others have a full cluster with a SAN. It can be a logistical nightmare to maintain minimal hardware at such locations and somehow centrally manage solutions

- Operating edge-based container-based workloads or applications that require GPUs isn't always feasible using traditional infrastructure
- There's a lack of technical personnel at most edge locations and companies don't want to have to hire dedicated technical staff for each one

Transforming Edge Operations with Hyperconverged Infrastructure

Edge operations are areas in which the right hyperconverged infrastructure solution has the potential to completely transform the environment and how those environments are managed. The overall results can be lower costs, improved efficiency, and better overall disaster recovery capabilities.

So what does it take to achieve this edge 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 edge situation much more discerningly and deploy solutions that overcome their unique challenges. They need to choose solutions that unify management across all edge sites in a way that makes them appear as if they're a single entity even while they support a dispersed organization.

With an HPE SimpliVity hyperconverged infrastructure solution powering edge environments, administrators gain one-click upgrade capabilities previously discussed. From firmware to VMware vSphere to the SimpliVity software, a single-pane update manager ensures that upgrades are handled quickly and without hassle, even if you're updating hundreds of sites.

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.

Virtual Machines and Containers in One

With HPE SimpliVity's support for both VM-based workloads and integration with containers via Kubernetes orchestration, you can run both your traditional applications and your modern applications, which helps accelerate your edge ambitions.

Comprehensive Data Protection

Data generated or managed at edge locations needs to be treated just like data generated at HQ. In many cases, the data originating from such locations is even more important because it's the information that's created as the result of sales, sensors, or other production 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 edge-generated data to other edge sites or to HQ 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 onsite staff to perform tasks such as changing tapes and increases the potential for successful recovery in the event of a disaster. **Figure 3-2** illustrates a well-organized, efficient, streamlined infrastructure.

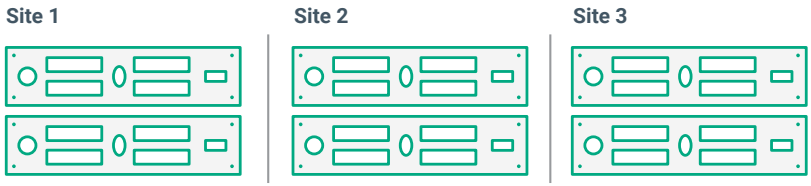

Figure 3-2: Hyperconverged infrastructure tames the ROBO chaos

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 3-3**). With this architecture, there's 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 to the hub.
- **Mesh** (**Figure 3-4**). 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 hyperconvergence 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

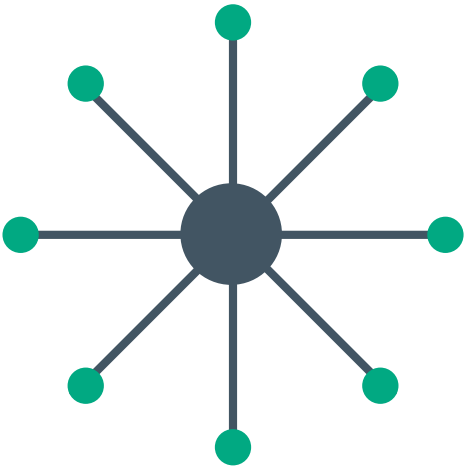

Figure 3-3: A look at a hub and spoke ROBO model

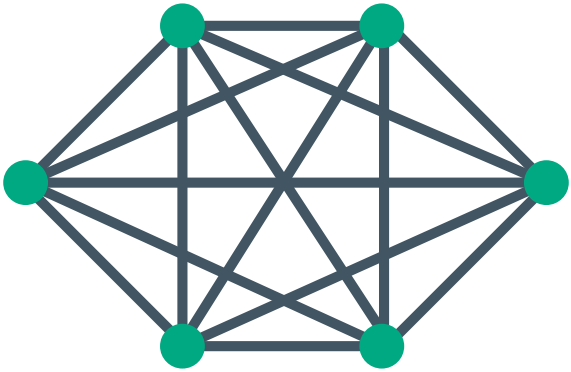

Figure 3-4: A look at a mesh-based ROBO model

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.

Up Next

Edge environments can be considered as "applications" that require some specialized attention. This is a perfect use case for hyperconvergence. However, there are some actual applications that have special resource needs as well. In the next chapter, you'll learn about these applications and how hyperconvergence helps.

CHAPTER 4

Tier 1/Dedicated Application Support

Not every company needs to tear down their entire data center and replace everything with shiny new hyperconverged infrastructure appliances. The chances are pretty good that you can't really do that even if you wanted to. However, you may have a single application that's challenging you and needs to be tamed. Or, perhaps you have a new application that you need to deploy, and you can't deploy it on your existing data center infrastructure.

For you, hyperconverged infrastructure still might be just the answer. In fact, even if you have only a single application, you might still be able to leverage hyperconvergence.

Enterprise Application Needs & Challenges

Not all enterprise applications are created equal. Every application has a unique performance profile, and each requires a varying amount of resources to be dedicated to that application. In Chapter 2, you learned about the popular enterprise application VDI and discovered that it has very different resource needs than general server virtualization.

Many traditional data center architectures aren't equipped to handle applications that don't fit a mainstream operational envelope. That is, these traditional data centers are equipped to operate a broad swath of mainstream applications, but don't always have the capability to support applications with very unique resource needs. The kinds of

applications that fit into this category will vary dramatically from company to company. For some, the entire centralized IT function consists of just a file server, so even something as common as an Exchange system would place undue stress on the traditional environment. For others, the traditional environment handily supports Exchange, but SQL Server or Splunk would be a step too far.

Every application has some kind of an I/O profile. This I/O profile dictates how the application will perform in various situations and under what kind of load. On top of that, every organization uses their systems a bit differently, so I/O profiles won't always match between organizations. As you deploy new applications, it might be time to leverage hyperconverged infrastructure.

A lot of people worry about virtualizing some of their resource-hungry applications for fear that they won't perform well. This is why, even to this day, many companies still deploy physical SQL Server, Exchange, and SharePoint clusters. While physical deployment isn't "wrong," the benefits of virtualization are well-known and include better overall hardware utilization and better data protection capabilities.

Hyperconvergence and Dedicated Applications

The right hyperconverged infrastructure solution can help you to virtualize even the largest of your Tier 1 mission-critical applications while also ensuring that you have sufficient resources to operate these workloads. Plus, don't forget the major role hardware acceleration plays in some hyperconverged systems.

Modern hyperconverged infrastructure solutions offload deduplication operations either to a dedicated accelerator card that sits right inside the server or to a powerful software stack that takes advantages of the fact that Intel keeps adding processing cores to its CPUs. Some

of the cores are carved out to enable deduplication without imposing a significant performance impact.

Elements of the Microsoft stack, including SQL Server and SharePoint, can be safely virtualized and significantly accelerated by moving to hyperconvergence. The same holds true for Oracle. Other I/O-hungry applications are growing in popularity, too. Splunk and Hadoop are two emerging applications that carry with them pretty significant I/O requirements. Splunk is a logging tool that subjects to abusive write-intensive workloads, while Hadoop is a big data analytics tool that requires a whole lot of both read and write I/O capability. Both need a lot of storage capacity, too, which is where deduplication features come into play.

Even better, as you need to grow, you just grow. Scalability is a core part of the infrastructure. When you grow, you can add more storage capacity, more storage performance, more CPU, and more RAM as needed, so you don't need to worry about encountering a resource constraint somewhere along the line. That said, one common misperception about hyperconverged infrastructure is that you are absolutely required to scale all resources at exactly the same rate. This is simply not true. For example, with HPE SimpliVity, you can add compute-only nodes that don't have any storage. It's not a one-size-fits-all conversation. Even in these cases, scaling is simple.

Moreover, for whichever applications you choose to include in your hyperconverged infrastructure, depending on the hyperconverged infrastructure solution you select, you can gain comprehensive data protection capabilities that will help you more quickly recover in the event of a disaster or another incident. In addition, you can also inherit the ability to manage the hyperconverged environment from a single administrative console.

Finally, if you're thinking "private cloud" with regard to your data center, you have to virtualize your Tier 1 applications in order to bring them into the centralized, API-driven management fold. A private

cloud is a VM-centric construct that requires high levels of virtualization to imbue the environment with the agility and flexibility needed to get things done.

Up Next

Just when you thought that you had everything solved by virtualizing and moving all of your Tier 1 applications to hyperconvergence, now comes a directive to consolidate your disparate data centers.

That's the topic of Chapter 5.

CHAPTER 5

Data Center Consolidation

Mergers and acquisitions. Cost cutting. New business initiatives. "We're moving to the cloud!" There are all kinds of reasons why companies make the decision to consolidate data centers. Maybe your company undergoes explosive, barely controlled growth or they may decide to stop, take a pause, and reconsider how IT does business in their organization. Or maybe your company decides to buy out another company, and you suddenly inherit a whole series of data centers that you're not prepared to handle.

Want to know a secret? It will be up to you to figure it out. Furthermore, you'll probably be asked to do it with the same budget you already have.

Here's the thing, though. Data center consolidation isn't always just about reducing the number of data centers from a big number to a smaller number. Sometimes, it's about reducing the amount of stuff strewn about the data centers that you already have.

Today's IT organizations generally buy and integrate numerous point solutions from a plethora of vendors, each with its own training courses, licensing, refresh cycles, and mode of operation. These point products are the result of years of planning and investments to support business applications.

We hear the same story time and time again. Does this sound familiar? You virtualized a decade ago and naturally a data protection strategy project came directly after that. The SharePoint implementation project for your marketing organization took eons to complete and required purchasing a new SAN. It feels like you just bought that SAN yesterday, but you blinked, three years flew by, and it's time to

refresh … again. The decision to buy all of these products made sense at the time, but today data center complexity can feel overwhelming and discourage innovation.

All of this has really cooked up several challenges for IT organizations, including:

- Time overhead spent on operational tasks
- Mobility and management of VMs
- Budget constraints
- Breaching service-level agreements (SLAs)
- Operational efficiency
- Application performance

The list goes on and on. Complexity is one of the biggest issues holding back IT, and the business, today. It's also a key inhibitor to your company's digital transformation efforts, which rely on a simple approach to technology.

Ask yourself these questions to determine where you might have pain points:

- As data growth explodes, can you continue to operate the same way you always have?
- Are your legacy technologies designed for virtualized environments?
- Are you spending more time on maintenance, upgrades, deployments, provisioning, and management tasks instead of building more valuable innovation for the business?
- Do you have the necessary expertise to manage each of these products separately?
- Are your applications and programs delayed by siloed teams and backlogged requests?

- Are you struggling to meet your SLAs with the business?
- Are you missing data protection objectives like RPOs and RTOs?
- Can you point directly to technical complexity as a major factor holding back your digital transformation needs?

Review your answers carefully. You may find that you're the perfect candidate for hyperconverged infrastructure.

Consolidation with Hyperconvergence

In every chapter of this book so far, you've learned about how hyperconverged infrastructure solutions can reduce the variety of hardware and software you have to manage in the data center. Every time you eliminate a class of hardware or software in your data centers, you're on your way to answering "Yes" to all of the questions outlined in the previous section.

That's the ideal scenario.

At the most basic level, hyperconverged infrastructure consolidates storage and compute, enabling you to eliminate the monolithic SAN environment. From there, some hyperconverged infrastructure vendors make things pretty interesting. For example, Nutanix and HPE both provide something in the way of data reduction via deduplication and compression. HPE SimpliVity, however, takes this to the extreme through the use of their accelerator components, which form the basis for what they call their HPE SimpliVity Data Virtualization Platform.

By enabling global inline deduplication and compression with a solution like HPE SimpliVity, you suddenly need less overall capacity, which means you need less overall hardware. With constant data reduction, you no longer need:

- WAN accelerators to reduce data over the wide area network because data stays reduced

- Separate backup software to protect the data in your environment
- Separate deduplication appliances
- Separate SD arrays

Instead, you can massively reduce the amount of hardware and software that you're operating, maintaining, getting trained on, and, maybe even struggling with. With less stuff to manage and worry about, you can better focus on the business and on improving SLAs, RTOs, and RPOs. You get to focus on the business rather than on the technology.

If you're in a situation in which you need to reduce the number of data centers you're managing, hyperconverged infrastructure can help you there, too. How? For the same reasons that we just discussed. Rather than just taking all of the hardware from the various sites and combining it all into one supersite, you can rethink the whole model. In addition to cutting down physical locations, you can also minimize complexity.

Up Next

Data center consolidation is important, but you still need a place to run applications and perform testing and development. Testing and development environments are often short-changed in the world of IT. In the next chapter, you'll learn why that's not a good situation. Plus, you'll learn how hyperconverged infrastructure can help you to improve your operations—from testing to production.

CHAPTER 6

Test/Development Environments

It's pretty clear that production environments enjoy premier status in most data centers. Production gets the fastest storage, the biggest servers, and all of the supporting services that make the application magic happen. Meanwhile, the poor test and development (test/dev) environment doesn't get all that much attention.

Let's take a look at what the test/dev environment supports. Test/dev consists of important activities, which include:

- Testing new application versions as they're released in order to determine potential impact on production
- Creating new custom software to serve the needs of the business
- Having a place to perform unit testing and load testing for new software being created by developers

In fact, in some organizations, even the developers' development machines are virtualized, and they work against virtualized instances of production software to ensure that their efforts will translate well into production.

The State of Test/Dev Environments

In many companies, test/dev environments are often given leftovers and hand-me-downs. For example, production servers that have been decommissioned might be moved to the test lab or to a development lab. These servers are configured just like they were three to five

years ago when they were originally purchased, and they generally don't have warranty support. Further, they use hardware that's one or two generations removed from current products.

The same goes for your storage systems that support the test/dev environment. Storage might consist of the old SAN that was removed from production. Or it might include a cheap array of disks, which provides reasonable capacity but is lacking performance.

At first, this may seem like a reasonable thing to do. After all, test/dev is a lower priority than production, right? Well, there are a few reasons why test/dev is more important than you might think:

- **Time is money.** That's the old adage. By using older, slower equipment in test/dev, you waste staff time that could be better spent doing other things.
- **Development efficiency.** Your developers are likely among your higher-paid staff. The more you short-change their work environment, the slower they work and the less efficient they become. This leads to slower overall development time and increases time to market for new features, products, and services.
- **Work stoppage.** Not having a warranty equates to having non-existent or slow service, if and when equipment fails. Failure in test/dev means that a critical piece of your environment is no longer available.

The Impact on Production

In most organizations, it's good to make sure that the test/dev environment resembles the production environment, especially when it comes to developing software and pushing it from test/dev to production.

When there's massive variance between test/dev or when test/dev isn't sufficient, bad things can happen, including:

- **Perplexing performance.** An inability to truly determine how well an application will perform in production means that you can't quickly resolve performance-related problems. When hardware between production and test/dev isn't close, applications will probably run very differently. This means you can't easily predict how well applications will operate.
- **Elevated expenses.** Some say that having underpowered hardware in test/dev actually makes sense since it means that, if an application performs well there, it's guaranteed to work well in the more robust production environment. In essence, they're saying that overbuilding production makes sense. That means that you're buying resources you may not need.
- **Insidious inefficiency.** The fact is that having two complete sets of hardware doesn't always make sense, even when it's necessary.
- **Dubious data defense.** Many people don't do data protection in test/dev since it's not as critical as production. For those that do a lot of internal development, they often do take steps to protect code, but not always to the level that they do in production and they may leave test/dev more vulnerable than they would like.

Hyperconverged Infrastructure in Test/Dev Environments

Once again, the right hyperconverged infrastructure solution has the potential to address all of the challenges identified in the previous section. Further, with the right solution, you can also add test/dev capabilities to companies that may not have had it in the past.

When There Is No Test/Dev

There are companies that don't have any test/dev environment. They don't have the budget, the personnel, or the space to stand up a complete test environment, so they operate by directly updating production before performing complete testing. This is a relatively high-risk activity that can be disastrous if a mistake is made. We recommend having at least some kind of testing capability to make sure that updates to production don't result in downtime.

There are a couple of ways you can stand up a test/dev environment using a hyperconverged infrastructure solution:

- Build a separate environment
- Add an additional node or two to production
- Use hyperconverged infrastructure for test/dev only

Each of these methods has its own benefits. Building a complete environment that mirrors production makes it possible to truly see how well applications will perform in the production environment and also provides plenty of capacity to allow development to take place.

Figure 6-1 gives you a look at how such an environment might be structured.

Further, this method makes it possible to use each hyperconverged environment as a replication target for the other. You can protect production by replicating it to test/dev and vice versa. When you stand up the environment like this, you can also take advantage of any global deduplication capabilities offered by your hyperconverged infrastructure platform.

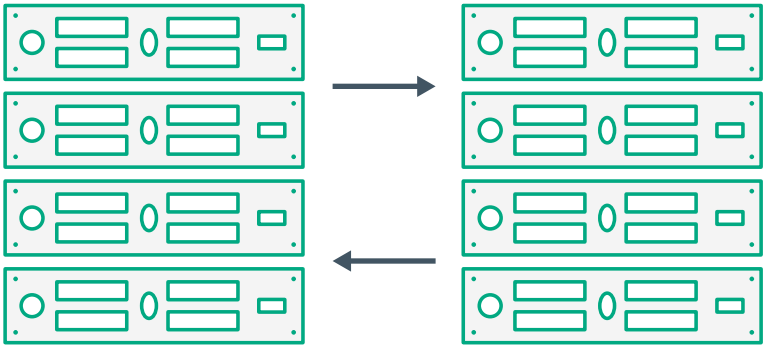

Figure 6-1: Building out two environments to support separate test/dev and production scenarios

This is a key factor in containing costs. In essence, you can deduplicate the entire environment and, since test/dev mimics production, the capacity savings can be huge.

If you don't need a complete replica of production, you can also opt to simply add an additional node to your existing production hyperconverged infrastructure environment. As is the case with building a completely separate test/dev environment, you'll still enjoy the incredible capacity savings that come with global deduplication. This benefit is also for the same reason—the test/dev workloads mimic production, so even though there are a lot of identical blocks floating around the workloads, each of those only has to be stored one time.

When it comes to disaster recovery, you have a few options as well. With a separate environment scenario, you already know that you can do disaster recovery between the two environments via replication.

With a hyperconverged infrastructure solution that deduplicates across the entire environment, you'll save a whole lot of storage capacity.

Also, if you choose to simply add nodes to production to handle test/dev needs, you can still replicate everything from production to a secondary site if you have one. Regardless, you'll still be able to

withstand the loss of a node in the cluster while maintaining operational production and test/dev capabilities.

There's also a third potential use case: using hyperconverged infrastructure for test/dev only. It's entirely possible that you already have a well-running production machine and you don't want to move it to hyperconverged infrastructure. There's no reason that you can't consider using the architecture for test/dev only. This will ease the administrative burden in test/dev and avoid the need to get too deep into the technical weeds for that environment. Things will just work. You won't need to buy a separate SAN and you'll get great performance for this critical infrastructure arena. Further, since you'll probably have a lot of different copies of the similar VMs in test/dev, you'll be able to get great benefit from any data reduction services that may exist in the hyperconverged infrastructure solution.

With hyperconverged infrastructure, you won't have to maintain the IT skills around the dev/test SAN and other needs and will be able to focus the development budget on application development. There are also some other benefits that can be had by using hyperconverged infrastructure in test/dev:

- Allows you to keep pace with business needs by quickly turning around incremental tasks in a production-like environment
- Ability to clone production environments and integration environments in minutes
- Having a well-defined process that includes ways to push changes to production, including creating backup of the original environment to have ability to roll back
- Developers and businesses would like to adopt a SaaS model for test/dev and are looking for cloud-like elasticity and ease of getting environments established

Adding Support for Container-Based Workloads

Containers and microservices are the new normal for application development, particularly for the emerging breed of cloud-native applications. Still, many organizations continue to struggle with adoption, as these new application development architectures need a modernized infrastructure approach to achieve the speed, simplicity, and insights necessary to drive transformation.

In this area as well, SimpliVity's hyperconverged infrastructure platform has a solid solution. In addition to your VM-based workloads, you can run modernized applications on SimpliVity and have its differentiated and enterprise-class data management services applied to data stores created for the containerized applications in VMware vSphere environments.

This provides a fully integrated experience for Kubernetes as an application in a hyperconverged infrastructure environment. With its support for the Cloud Native Computing Foundation's (CNCF) Container Storage Interface (CSI) through a vSphere Cloud Native Services plugin, SimpliVity can integrate with any standard upstream Kubernetes distribution to create, manage, and orchestrate applications and associated persistent volumes (PV).

This feature provides a number of data services:

- Persistent volume (PV) create, delete
- PV copy, clone
- PV snapshot/backup, both local and remote (Kubernetes 1.17 or later)
- PV restore (Kubernetes 1.17 or later)
- Policy-based automation for PV data backup

- Always-on dedupe and compression across primary storage and backup
- Enterprise-class availability and resiliency

Having support for CSI allows both VMs and container-centric workloads to run on the same host or cluster, eliminating the need for dedicated infrastructure for container workloads. Also, the standard minimum two-host configuration inherent in the SimpliVity architecture is supported, although a three-host solution increases resiliency.

With CSI and Kubernetes support, you can build distributed data centers with SimpliVity federation and have the ability to move data (snapshot and restore) among them for better data protection, agile application development, DevOps, and other modernized application development and delivery pipelines for modernized workloads.

Up Next

It's clear that test and development environments can be significant assets, but they're only useful if they're leveraged in a way that supports the needs of the business. In the next chapter, we'll talk about what happens when IT goes rogue—or at least appears to. Alignment between IT and the business is the topic of Chapter 7.

CHAPTER 7

Aligning Architecture and Priorities

Perhaps one of the most enduring meta-conversations about IT in past decades has been focused on how well IT serves the needs of the business. Often referred to as "IT/business alignment" this conversation generally used to indicate when IT failed to meet the needs of the business. In a perfect world, there wouldn't have to be this conversation because IT would never be considered as off-track or "rogue." Unfortunately, that isn't reality. IT often struggles to maintain a focus on the business, a problem often exacerbated by the infrastructure solutions that have been adopted.

In fact, this whole idea of alignment is one that hyperconverged infrastructure has the potential to address head-on. No, it won't fix every alignment problem in every organization, but it can begin the process.

The State of Hyperconverged Infrastructure

ActualTech Media research has uncovered a pretty significant misalignment between IT priorities and potential hyperconverged infrastructure benefits.

Data centers are among the costliest physical assets owned and operated by organizations. The cost isn't just in the equipment that's deployed, but also in the sheer effort that it takes to manage that equipment, keep it running, and keep it maintained year after year. To

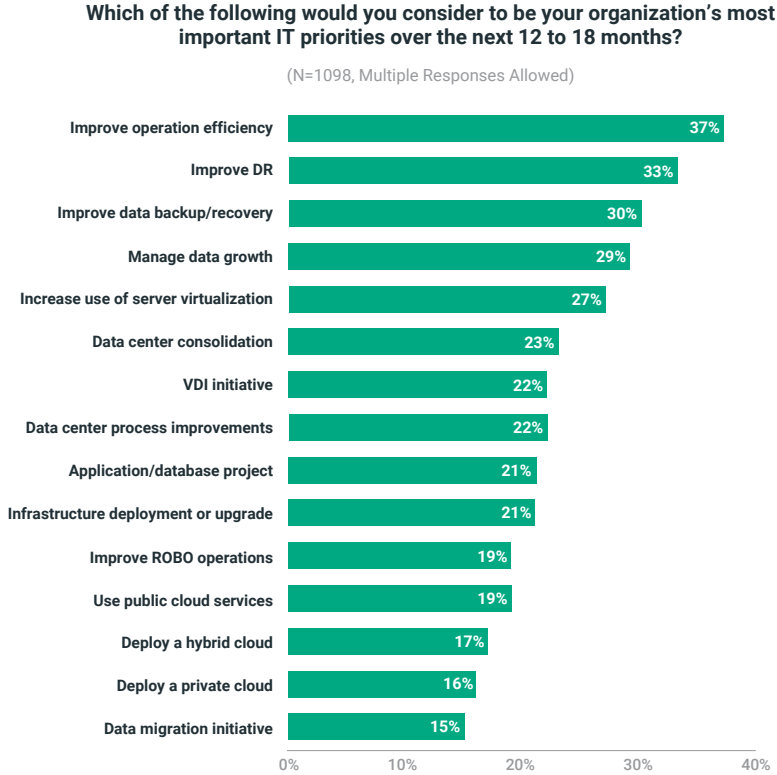

Figure 7-1: Primary drivers for interest in hyperconverged infrastructure

make matters worse, many companies have deployed Band-Aid-like solutions to patch over problems introduced as the data center grows more complex or is challenged to meet emerging business needs.

Let's start with the items considered priorities by respondents. In **Figure 7-1**, you'll see that improving data protection, improving operational efficiency, and implementing VDI are the top-three items on respondents' radars. Remember, these responses don't consider the role of hyperconverged infrastructure; these are simply overall IT priorities.

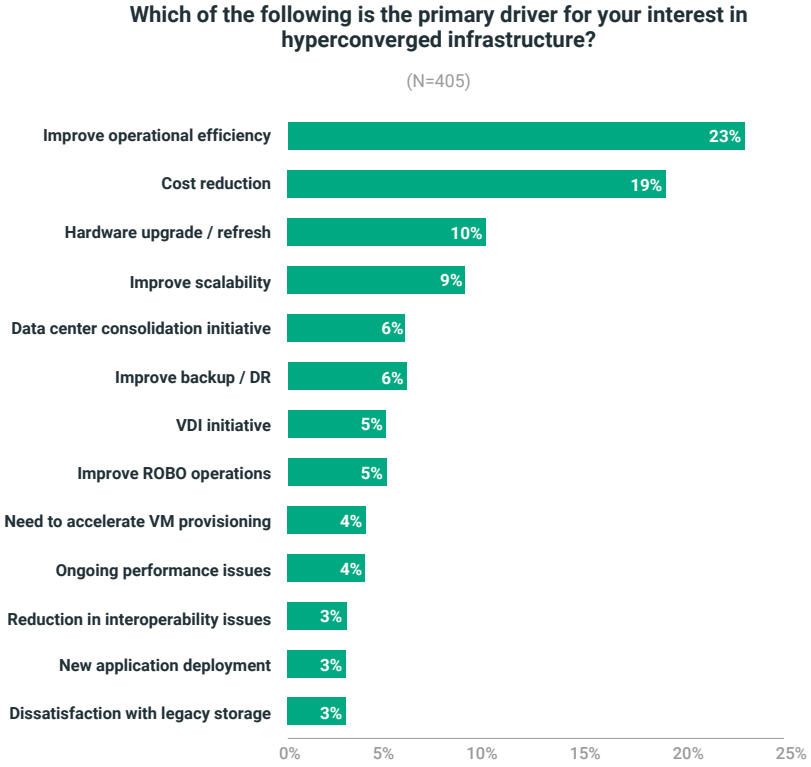

Figure 7-2: Primary driver for interest in hyperconverged infrastructure

Now, let's look at respondents' primary driver for considering hyperconverged infrastructure, the results of which are shown in **Figure 7-2**. See if you can tell exactly where the results of each question diverge from one another. Notice anything interesting?

Improving operational efficiency is near the top of both lists, as we'll discuss later in this chapter. What's a bit more interesting is where we see divergence, particularly as it pertains to data protection. There's a vast gulf between the importance of data protection on the overall IT priorities list and what people look for in hyperconverged infrastructure.

Data Protection

Improving data backup and disaster recovery emerged as the single most important overall need for the IT organization from this research. In comparing key drivers for hyperconverged infrastructure against larger IT initiatives, it was surprising to see that data protection ranked seventh on the list despite the fact that it was identified as the highest IT priority to address. This may be due to the fact that enterprises aren't equating modernizing the architecture with hyperconverged infrastructure with modernizing data protection; they continue to view hyperconverged solutions as simple conglomerations of servers and storage. Because, to many people, "hyperconverged" simply means exactly that, it may not be so far-fetched that they don't consider data protection a key part of the hyperconverged package. Many hyperconverged infrastructure solutions include backup, recovery, disaster recovery, and business continuity capabilities.

From work performed by IDC, it's clear that data protection continues to rise in importance. Learn more about this, including why IDC finds traditional backup/recovery software to be insufficient for today's workloads, by reviewing the paper that HPE has made available on the topic: https://www.hpe.com/us/en/resources/integrated-systems/hyperconverged-data-protection.html.

This book devotes an entire chapter to this topic, so we won't reiterate all of that here, except to say that those who have significant backup, recovery, and disaster-recovery needs would do well to carefully study the hyperconverged infrastructure market and understand what's possible in this realm. With the right solution, there are some impressive data protection capabilities available.

Operational Efficiency

The VM is the center of the universe when it comes to applications in most modern data centers. Most new workloads are deployed in VMs. However, consider the state of centralized policy in the data center. For data centers that have equipment from a wide variety of vendors, or that have a lot of "point solutions" (such as WAN accelerators and replication tools), there could be a number of touch points when it comes to policies.

These various touch points don't always align very well with one another, particularly when there are different vendors in the mix. For example, while it may be possible to define some policies at the hypervisor layer, it's often difficult to apply storage policies that have any awareness of VM boundaries. There are myriad other devices in the data center that can suffer from the same problem.

Since the VM is the center of the data center universe, why not implement a system that focuses directly on these constructs? Hyperconverged infrastructure solutions provide this opportunity to varying degrees, depending on the vendor. Rather than go to three different places to define storage, backup, and replication policies, some hyperconverged infrastructure systems enable these policies to be attached to the VM.

Policy application is just one aspect of operational efficiency. There are many more, including:

- **Shielding complexity from the administrator.** Even IT pros shouldn't be subjected to complexity in the infrastructure when it can be avoided. Hyperconverged infrastructure helps make this happen. Availability mechanisms, such as RAID configurations and management, are often hidden from view and are simply a part of the environment. Software-defined networking takes this infrastructure simplification one step further, further automating and masking complexity from the VM admin with intuitive management GUIs.

- **Use-case improvements.** Ensuring that new applications and use cases, such as edge and VDI deployments, can be supported without adding complexity and introducing inefficiency into operations, is critically important to help IT maintain alignment with business needs. When deploying these kinds of applications introduces inefficiency, IT and business alignment will suffer.
- **Overall alignment enhancement.** As has been mentioned, efficiency and simplicity can help IT better achieve alignment with the business.

You'll have noticed that cost reduction is also very high on the list for survey respondents. We believe that cost reduction and operational efficiency go hand in hand with one another for many people. However, we also understand that hyperconvergence has the potential to dramatically improve how the IT budget is constructed. You'll learn much more about the economics behind hyperconvergence in **The Gorilla Guide To...® Hyperconverged Infrastructure Economics and Impact on the Budget**.

That's a Wrap!

You've been introduced to hyperconverged infrastructure and its use cases. You're now versed in using the technology to solve some of your most challenging business problems. Now that you've made it through the jungle, you can move on to more harmonious locales. We hope your journey here has helped you better understand this important topic.

ABOUT HPE

Hewlett Packard Enterprise

Hewlett Packard Enterprise is an industry-leading technology company that enables customers to go further, faster. With the industry's most comprehensive portfolio, spanning the cloud to the data center to workplace applications, our technology and services help customers around the world make IT more efficient, more productive, and more secure.

For more information, visit www.hpe.com

ABOUT ACTUALTECH MEDIA

ActualTech Media is a B2B tech marketing company that connects enterprise IT vendors with IT buyers through innovative lead generation programs and compelling custom content services.

ActualTech Media's team speaks to the enterprise IT audience because we've been the enterprise IT audience.

Our leadership team is stacked with former CIOs, IT managers, architects, subject matter experts and marketing professionals that help our clients spend less time explaining what their technology does and more time creating strategies that drive results.

For more information, visit www.actualtechmedia.com