The Gorilla Guide to...

Enabling the Composable Cloud with Cloudistics

Inside:

- Discover the key differences between on-premises and cloud-based architectures
- Find out how composable infrastructure is poised to displace both traditional and newcomer data center architectures
- See how the Cloudistics Composable Cloud works and how it can enable positive business outcomes

Scott D. Lowe
ActualTech Media

In Partnership With

cloudistics

Helping You Navigate The Technology Jungle!

ActualTech Media
www.actualtechmedia.com

The Gorilla Guide To...

Enabling The Composable Cloud with Cloudistics

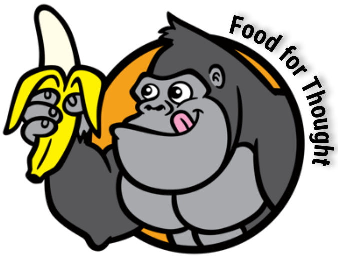
Author
Scott D. Lowe, ActualTech Media

Editors
Hilary Kirchner, Dream Write Creative, LLC
Christina Guthrie, Guthrie Writing & Editorial, LLC

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

Icons Used in This Book

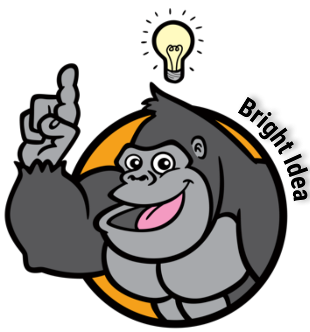
Discusses items of strategic interest to business leaders.



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

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

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



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

Chapter 1

Comparing Public Cloud and On-Premises Architecture

There is a fundamental promise involved in cloud. When an organization decides to make an investment in IT staffing, rarely do they think, “Man… I can’t wait to have someone manage the network!” Rather, such investments are generally made with the idea that the selected individuals will help to propel an organization forward. All too often, however, we see companies hire very smart, very capable people with great technology skills but lament the fact that so much of their time is spent on infrastructure and so little on the business. Of course, we all know that keeping the network operational is incredibly important, and, even if it doesn’t seem like it to decision-makers, infrastructure is the lifeblood of the business.

That said, for most companies, IT isn’t the business, so there’s not always the interest and attention paid that many think it deserves. There’s also the fact that IT infrastructure has become notoriously complex over the years, forcing organizations to hire hordes of expensive people to keep it all in check. To make matters worse, every few years, the business needs to make a massive capital outlay to replace the infrastructure they have with currently available gear. In general, this is a “grin-and-bear-it” scenario for business decision-makers.

What's an IT decision-maker to do? There is only so much money in the budget, and business expectations around how IT supports the business are continually shifting to an application-first, business-first mentality, which might not align well with the reality that we face when it comes to traditional IT infrastructure.

These are among the many reasons that the public cloud has become such a force, although, as you will learn in the book, it has its own set of challenges. Setting those challenges aside for now, the promise of cloud is that it frees IT from a great deal of the infrastructure, thereby allowing the department to focus more on applications and the business. This, after all, is exactly what business decision-makers are looking for. They want to be able to leverage IT to propel the business forward, not view IT as a necessary anchor that needs to be dragged along.

This freedom to focus on the business and the applications that drive it is the fundamental promise of cloud. As you'll learn in this book, that promise doesn't have to mean that the public cloud is your only option.

The Architecture of Cloud

"The cloud" is someone else's computer. When the phrase "the cloud" is invoked, it's generally referring to the public cloud. When you subscribe to cloud services, you're basically renting cores on someone else's servers, blocks on someone else's storage, and bandwidth on someone else's network. That's the simplest explanation, but there is far more to it than that.

With the cloud, the infrastructure becomes the problem of the provider. You just consume the resources, and they manage them. There's a whole lot under the hood, though, that makes this kind of rental opportunity possible.

First, there's automation. When you request a new virtual machine or other resource from a provider, it doesn't mean that provider personnel need to jump into action. Rather, the resources you request are automatically provisioned. This is what allows a provider to scale their service without having to linearly add staff as they add customers.

Second, there's orchestration. When you request a virtual machine, you may also be getting other services, such as storage, a network, or any of a multitude of other services. The automation engine that powers each individual component at the cloud provider participates in a larger automation cluster that can orchestrate requests across resources. This is why you get immediate provisioning from the cloud.

Finally, there's self-service. As mentioned, there is no need for provider personnel to spring into action when you request a new resource. The service request portal you use is integrated with the automation and orchestration engine so that your wishes are carried out with just a few taps of the mouse.

For business decision-makers, cloud is seen as a panacea of sorts, thanks to its cost model. No longer do companies need to expend massive mounds of money at routine intervals and hope that they ultimately grow into their purchases. Instead, cloud enables a consumption-based pay-as-you-go model that operationalizes expenses. You pay only for what you use. There are no massive capital investments to be made.

At inception, this is pretty compelling, but, over time, costs may not always work out the way you'd expect, a fact we'll discuss later in this book.

But, can't I finance or lease to get the same benefit?

Financing IT data center infrastructure has been going on for decades, so you might wonder why people are so captivated by cloud economics. After all, with financing and leasing options, aren't you also paying as you go?

Well, yes and no. Financing and leasing do enable you to lay out less capital, but you're still ultimately procuring a fixed infrastructure configuration. You're just paying for it a bit differently.

With cloud, companies are getting instant procurement. They can buy capacity in minutes instead of waiting weeks for vendor selection and procurement processes to take place.

Cloud vs. On-Premises: A Business Primer

For the reasons outlined previously, the public cloud has become a key factor in decisions related to how to operate business workloads. Between lower initial costs, a less burdensome buying process, and less complexity, there's a lot to love about the public cloud, but it's not perfect. We'll go more in-depth into this later in this book. For now, we'll look at things from a high level.

There are really good reasons that companies choose to keep workloads on premises. The first is a matter of performance, and this is a scenario that can't be overstated. Certain workloads, when moved to the cloud, will perform horribly, at least from the perspective of your users. And, let's face it; that's what really matters. If your users are unhappy with performance, that frustration will make its way up the chain of command. For this reason, it's rare to see bare metal workloads moving to the cloud.

The second reason workloads often need to remain on premises is all about regulation and security, which can include data locality. Although there are public clouds that can support the compliance and security needs of regulated verticals, there is often a lack of comfort associated with losing full control of these workloads. Moreover, there is often concern around where organizational data will ultimately reside. Different jurisdictions carry different laws and regulations, making it very important to be able to answer the questions, "Do you know where your data is?" or "Can you prove deletion?"

There are more reasons that I'll discuss later in this book. For now, let's turn our attention to the business front for a minute.

Busting Public Cloud Myths

So often in enterprise IT, as new technologies emerge they are riddled with problems. Over time, these problems are often addressed, but the perception of the original problem remains. This is where a lot of enduring enterprise IT myths originate, and they need to be debunked on a continual basis.

"Public Cloud Is Cheaper Than On-Premises"

You've heard the rallying cry of cloud providers insisting that jumping feet first into the cloud will save multiple truckloads full of cold, hard cash. And, frankly, they're right!

Wait, what?

Moving to the public cloud will save you money but with some really important caveats that depend on:

- How you look at your budget
- Your time horizon
- The size and scope of your deployment.

If your ultimate goal is to set up a single virtual machine in the cloud to run your entire business, and that's all you need, then cloud will be less expensive than buying a server and hiring a person to manage it. But most companies have far more needs than this. Instead, they need masses of servers all intertwined with one another to create a digital work of art that is functionally pleasing for the business, if not always aesthetically so. There is a point at which the public cloud becomes *more expensive* than simply keeping workloads on-premises. This line will be different for different organizations.

In fact, in a phenomenon that is becoming known as *cloud sobriety*, we see companies of all stripes reversing cloud deployments and returning some workloads to on-premises environments. There are a lot of reasons this is taking place, including emerging private cloud data center options that rival public cloud in terms of cost.

The Cloud Cliff

There is a hype cycle associated with many technologies, and "the cloud" is no exception. Over the years, we've seen a lot of companies jump into the public cloud with both feet.

Now, many have discovered that the public cloud is costing far more than they expected, and it has even surpassed what it would cost to run workloads on-premises. Although the choice of the public cloud over on-premises is not always just about money, cold, hard cash does tend to be a driving factor in decision-making.

This cloud cliff is a term that describes the point at which the public cloud becomes more expensive than simply going it alone. Reaching this summit is but one reason that we see a number of organizations rethinking their public cloud strategies and bringing some workloads back to local data centers.

"Public Cloud Is Easier Than On-Premises"

The difficult fact is that on-premises data center infrastructure has become incredibly complex. As mentioned before, companies were hiring entire teams of people to manage technology, while at the same time telling the business to be more efficient. It was a case of "do as we say, not as we do," and people noticed. As costs for infrastructure continued to rise, decision-makers realized that there had to be a better way.

Then the cloud emerged, and their hopes and dreams were answered!

Well... maybe only to a point. One area in which the public cloud absolutely trounced on-premises IT was in the user experience. While expensive IT personnel had to cobble together a series of resources in order to deploy a new workload internally, with the cloud, anyone with a mouse, a keyboard, and a semi-functional left hemisphere could point and click their way to a new workload.

Instant workload! Or, in some cases, instant chaos if not carefully monitored.

This public cloud-centric exclusivity with regard to ease of use didn't last long, but it did last long enough to cement in people's minds the perception that public cloud is easier than keeping things local. However, over the years, a whole lot of really smart people have developed tools that bring the kind of ease-of-use just described to the on-premises data center.

Even better, more than just plain old software tools, truly forward-thinking companies have enabled their software tools to work with commodity x86-based hardware to create new platforms that can help companies ease the complexity pain on both the hardware and software fronts.

And, here's the thing: even with cloud, you need subject matter experts, just *different* ones. You need people to manage public cloud just like you need people to manage on-premise IT. There's a security expert, a capacity planning expert, project planners, migration experts, etc. Just because you don't need a networking guy with his CCIE doesn't mean you don't need a networking expert that can integrate your cloud services with your local data center. Public cloud doesn't mean that complexity goes away.

In short, this myth is long dead, but many people just don't know it.

"The Private Data Center Will Be Dead in Five Years"

We've heard variations of this phrase for decades. "Tape is dead." "Disk is dead." "The mainframe is dead." In some cases, the prognostications came to be, and in other cases, things didn't play out as predicted. One of today's favorite emerging forecasts indicates that the private data center will be dead in five years, ten years, or some other figure in the not-too-distant future.

Here's the skinny on that: It's not going to happen. That said, it's an absolute certainty that certain workloads will—and *should*—move to the public cloud. However, there are just too many challenges to simply throwing everything over that public cloud fence and turning off the lights in the local data center.

First and foremost, we have latency, the silent workload and business killer. The farther that some workloads are from users, the longer it takes network communications to take place. It's pretty simple physics. Too many of today's applications remain sensitive to latency issues on the communications fabric. For example, a whole lot of companies continue to run client/server applications. If you were to try to push the server side into the cloud and continue to run clients locally, you would end up wrecking the user experience.

Of course, you may say, "Well, then I can just push the clients into the cloud, too! We need VDI anyway." Now, I'm not here to tell you to avoid VDI, but I will tell you that doing it just so you can run a bunch of stuff in the cloud means you will increase complexity and probably bump up your costs in the process.

Speaking of costs, refer to the previous section in this chapter in which you learned that public cloud costs can often exceed private cloud costs. That's another reason that we won't see the death of the private data center anytime soon.

"Amazon and Microsoft Are My Only Options"

As people hear the siren call of the public cloud, it looks as if Microsoft and Amazon and other similarly situated cloud players are the only viable options forward for enterprise IT. Frankly, nothing could be further from the truth. Of course, these and other providers will play a major role in your workload operations plans, but they are hardly the entirety of the future of IT.

For all of the reasons that have been outlined thus far in this book, the on-premises data center will continue to thrive, although it will look vastly different than it does today. CIOs and IT gurus across the globe will need to learn about the new opportunities that are at hand and, more importantly, will need to continue to discover how best to determine where specific kinds of workloads will reside.

Application Location Qualification

In general, to figure out just where to run workloads, you can create a taxonomy of workload characteristics. It may sound like a lot of work, but it's not really all that difficult. You just need to consider each of your workloads and determine how each one operates, details of which are explained in the following sections. From there, you have the beginnings of a strategy for what to run where. The sections below are some of the high-level areas you need to consider as you try to decide whether to run your workloads in your on-premises environment or start the process of pushing them into the cloud.

Topic Area	App 1	App 2	App N
Governance and Regulatory			
Performance			
Customer Support			
Cost			
Data Locality			
Dynamic vs. Static Workloads			
Result (On-Premises or Cloud)			

Figure 1-1: Application Location Qualification Checklist

Never believe that you need to look at the items in the following section individually. You must consider all of your workload characteristics in aggregate so you can make the best possible decision.

The first thing you'll want to do is create a complete list of all of your applications and other workloads so that you can get started on some classification. Figure 1-1 above provides you with a sample checklist that you can use.

Governance / Regulatory

There are times that things are just out of your control. Someone else tells you what you have to do. In some cases, that someone may be a government or another regulatory agency that places right restrictions on how and where workloads can operate. As a result, some workloads just can't run in the cloud. For example, some financial systems can't easily move to the cloud due to regulatory requirements. This is starting to change as more secure cloud environments become available, although these types of environments may be more expensive than plain old cloud. Until this is more predominant, you'll be sticking with your on-premises data center.

Performance

So much effort is expended to eke out every ounce of performance because the business and certain applications demand it. With on-premises infrastructure, you can load a single virtual machine atop a massive host with SSD-infused storage to make the application scream. You can do the same thing with a bare metal workload. You own the hardware, so why not? In essence, you can decide how resources should be allocated to your internal workloads to maintain performance.

In the cloud, you have a lot of options with regard to workload sizing, but you may not always get the performance you expect, particularly if you are still running the server portion of a client/server workload in the cloud without accounting for the client part.

While there are ways that you can bring back that well-understood client/server experience and run in the cloud, doing so imposes all kinds of new potential performance problems and increases overall complexity.

Customer Support

Rightly or wrongly, IT generally wants to jump in when something goes awry with infrastructure so that it can correct the problem. You may be asking, "How can this be wrong?" It's only wrong if you're running workloads in the cloud and the provider has a really bad day or event that brings you down, and you feel helpless because you can't jump in.

That's part of the tradeoff. When you're running workloads in someone else's data center, support falls to them. It may take longer for them to restore your services, or it may not take as long as it would have taken you.

The reality is that, in the cloud, you are at someone else's mercy, and your support will only be as good as the people on the other end. If you've chosen poorly, that can and probably will impact service levels. Even if you've chosen well, if and when things do start to go sideways, you're something of a spectator rather than an active participant in restoration efforts.

If you have workloads that demand 24/7 attention and you can staff that way, run them locally so you can have control. Otherwise, decide what makes the most sense for each workload, whether that's on-premises or in the cloud.

Cost

I discussed this just a couple of short pages ago, but it's worth repeating. At first, with just a few workloads, cloud is most definitely less expensive than on-premises infrastructure. But, if you're reading this (which I assume you are), you probably have a lot of applications already. So, you may already be past the tipping point for bringing applications back in-house for cost reasons.

Let me be clear for a minute, though: if your *only* driver for technology is how cheap it is, you may be very disappointed at the eventual outcomes. That said, we all want systems that are cost-effective both at inception and on an ongoing basis.

As you undertake your workload analysis, you must decide where it makes the most sense—operationally and financially—to run each one. Email, for example, is almost a no-brainer for the cloud. Even as an Exchange Server-hugger myself, there's no way I'd stand up a new Exchange environment unless there was a truly compelling reason. The staffing cost can be significant when you consider deployment and ongoing administration. Moving to Office 365 takes that ongoing burden away, and you simply pay for what you use.

Similar decisions have to be made about *all* of your workloads. So sharpen that pencil and get to work!

Data Locality

Where your data resides is really, really important. With the cloud, you often hear people say things like, "It doesn't matter where your data lives. It's just available when you need it!" This is an unfortunately simplistic view of the situation, though. There are all kinds of things you need to consider. For example, for many reasons, people want to ensure that their data stays inside their countries. While that can be accomplished with region-based cloud service providers, as your company grows and crosses boundaries, there is a web of data sovereignty laws you need to adhere to.

For many, maintaining internal data centers helps in this effort, but certain workloads aren't really feasible for on-premises deployment. Consider, for example, services such as salesforce.com and Office 365. At some point, you'll need to develop a comprehensive plan around data in the cloud so that you remain compliant with your organization's policies and the laws of the countries in which your organization operates.

Data locality also has another side to it; this one is all about performance. When you split data from workloads, you run the risk of wrecking the good thing you may have. Many applications dislike—sometimes intensely—being separated from their data, and they may throw tantrums as a result. Such tantrums can include subjecting your users to less-than-desirable response times or may include not working at all. Bear this in mind as you journey ahead.

The Role of Data Locality in Workload Performance

There are a lot of different ways you can look at the issue of data locality. Some infrastructure vendors look at data locality very locally and track exactly which data blocks reside where so that data is close to running virtual machines.

But, there's a higher order version of data locality that comes into play as you consider whether to run server-side applications in the cloud or keep them on-premises. Although many applications and workloads will run just fine in the cloud, a great many will not.

Let's take a look at client-server applications, for example. These are applications for which there is a data center component—often a database server—and a client component that runs on a user's computer. Most of these kinds of applications (ones that have a dedicated client) generally expect the database to be reasonably local, not in a far-flung cloud somewhere. If you make the fateful decision to move such workloads to the cloud, you'll suffer application latency and performance issues unless you also take steps to move the client aide to the cloud.

While possible, it can often add undesirable complexity to make these kinds of applications run well in the cloud. In this context, you need to ensure that you fully understand, on an application-by-application basis, the potential locality issues that you might encounter.

Data locality is something you need to consider on a workload-by-workload basis, so make sure that's a part of your overall analysis.

Dynamic vs. Static Workloads

Some of your workloads are persistent and have pretty regular usage. Your ERP, for example, is relatively predictable and persistent and, based on that characteristic, is generally considered a very good candidate for on-premises operation.

For others, though, you may have a workload that doesn't have regular ongoing usage patterns and which may experience regular spikes and valleys. For these workloads, you may find it more cost-effective to operate in the public cloud where you can routinely turn up the resources and then scale them back as demand diminishes.

Each individual workload should be assessed to determine its ongoing performance patterns. From there, you can make the decision as to whether it should run in your local data center or be chucked over the wall into the cloud.

Chapter 2

Private Cloud Successes and Setbacks

There are some really key things to remember about the core capabilities of the private cloud. The most important is that most companies don't have one, even if they think they do. You see, virtualization alone does not a private cloud make. There are actually a number of characteristics that a data center environment must feature before it can really be considered any kind of cloud. If all you've done is virtualize your stuff, you have a highly virtualized data center and that's really it. It's admirable, and it brings you a lot of benefits, but it hasn't suddenly resulted in your company running a state-of-the-art private cloud.

Multitenancy

Amazon and Azure have more than one customer each. I know! I was surprised at first as well. Seriously, though, can you even imagine a reasonable scenario in which a cloud provider would custom build each and every environment as new customers come on board? Chaos would ensue, and the provider would quickly collapse under its own weight. And, economies of scale would be nonexistent.

Multitenancy is one of the cornerstones of the public cloud. Everyone shares the infrastructure, and the costs are spread among all of the users. The key here is that there is workload isolation. Each customer knows exactly what resources they're using and is charged for just those resources. In a private cloud environment, even though you may not want to implement internal chargeback mechanisms, you do need to be aware of who is using resources. Further, the environment should at least have the ability to isolate workloads between departments.

Microservices and Microsegmentation

Virtualization alone is like having a school bus. You can take a whole lot of people on a trip, but it's impractical transportation for your solitary commute. It's too big, takes up too much space on the road, and emits noxious carbon-containing particles into the atmosphere. What you really need for that lonely commute is a perfect vehicle designed for just you. You need one that you fit into well, whether you're short or tall, and that has the amenities you need to make your commute a bit more tolerable. Some of you will want satellite radio, and some will want a panoramic moon roof. And, your needs may change each day. As such, you need a vehicle that you can easily tear down and rebuild into different configurations.

These tailored needs are eerily similar to what happens in the world of microservices. Each individual component in a microservices architecture is responsible for carrying out a discrete task. Although the individual components include everything needed to execute on its own, including an operating system, they integrate with other microservices via application programming interfaces (API) to become a part of a larger assembly.

Containers, such as Docker containers, are probably the most well-known of the various services that help to support the emerging microservices market. As these kinds of constructs become more prevalent, you'll need a data center architecture that can support them. That makes support for microservices almost a requirement in the world of the private cloud.

Performance and Quality of Service

All too often, there is this idea that "performance" means that every workload must be able to run flat-out at all times. This is simply not the case. It's unlikely, for example, that a college's ERP system really needs sub-microsecond latency in the middle of the night, or that the storage needs to drive millions of IOPS at 3:00 A.M.

Instead, in a private cloud environment, you need to look at things in an aggregated way and ensure that performance is *consistent* among applications. Yes, you want them to run fast, too, but it's just as important that you maintain user expectations for how their applications actually operate.

A private cloud environment that provides a consistent level of performance needs to be able to continue to maintain that level of performance over time. This means that you have to deploy a solution to which you can add capacity and performance on a regular basis as aggregate workload needs demand.

User Self-Service

For those of you using Amazon, it's unlikely that you email Jeff Bezos and request that he spin you up a new virtual machine on demand. Cloud providers have a business model that is predicated on user self-service. That's one of the ways that cloud providers are able to reduce their costs.

You need to do the same thing. Do you have developers that have to work with IT operators to get their QA environments refreshed? Do you have savvy users that need to be able to deploy their own services, but are hindered by waiting for IT to get around to it? A key tenet of the private cloud is the ability for certain users to be able to administer their own resources.

This doesn't mean that you suddenly cede any and all control over the architecture; that would ensure a wave of chaos. You can still place reasonable limits on what users are allowed to do, but they need to be able to carry out ongoing operational tasks on their own.

Support

Can you support your internal environment as well as a public cloud provider can support theirs? Remember, public cloud providers have legions of experts. You have Chuck. Chuck is a great guy, but he can't possibly be an expert in everything, particularly when you have created your private cloud using legacy approaches to infrastructure that require highly specialized skills across a bunch of areas.

Of course, most enterprises have more than one person to maintain their data centers, but the point here is that to get to a place where you can support your infrastructure to the same level that a public cloud provider can, you have to do one of two things:

- Hire more people and ensure that, combined, they have more than sufficient skills in each of the individual resource areas, including servers, storage, virtualization, and networking.
- Deploy a far simpler data center environment that doesn't require hordes of specialists to support and maintain it.

In order to be a real private cloud, you need to go down one of these paths.

Hardware Lifecycle Management

In a cloud environment, you don't get notices from the provider telling you that they're taking down your workloads while they replace hardware. They may shift you to different systems as a part of a hardware change, but you don't end up being taken completely down.

Your private cloud needs to look like that. If you've virtualized most of your workloads, you're already down this road, since you can just migrate workloads around to different hosts. But, to be a private cloud, you need to be able to do this with *all* of your workloads so that you can manage your hardware lifecycle without impacting your workloads.

Successes and Setbacks

As you look at the current data center architecture landscape, it's clear that there are a number of trends at play. Let's explore some of the big ones.

VMware and Virtualization: A Success Story

Have you heard of VMware? They're kind of a thing in virtualization circles. Seriously though, can you imagine where we would be today had VMware not started the world down the x86 virtualization path? Some other company may have picked it up, but VMware managed to reshape the entire IT industry in under a decade, kicking off one of the biggest transformations in computing history.

VMware's vSphere became the must-have tool for organizations large and small. It shifted workload deployment from a weeks-long ordeal to a minutes-long process. It created new opportunities for data protection and disaster recovery.

In short, VMware's virtualization victory became a victory for all of us, and it continues to be a force to be reckoned with in the industry.

VMware: A Challenging Future

No matter the level of past success, though, technology companies often face existential threats on a regular basis. Failure to reinvent themselves or failure to innovate against upstarts can result in being relegated to the footnotes of computing history. Remember Novell? Back in the 1990s, Novell's NetWare was *the* network operating system of choice and Novell Directory Service (NDS) was emerging in similar fashion. And then three things happened. First, Microsoft Windows NT came along and began to steal shares from Novell. Second, in an effort to take on Microsoft, Novell bought WordPerfect and, for just a second, took their eye off the data center ball. Finally, Microsoft released Active Directory. The combination of those three events ultimately doomed Novell, a company that is now a division of Micro Focus.

VMware has relied on vSphere as a core part of its business, but up-and-comers have arisen to steal their thunder. On the hypervisor front, Microsoft's Hyper-V is a major threat, and, as it turns out, so is the open source KVM, which has been adopted and adapted by a number of infrastructure providers. The risk for VMware is that the "free" (open source) KVM has become a mature enough solution that it can supplant vSphere. And many are, in fact, choosing this option. The risk is further compounded as people seek ways to reduce ongoing licensing costs, of which VMware is often a big part.

For their part, VMware continues to innovate by creating new products that reinforce their hypervisor lead. The company has released a number of products around storage and networking to ensure continued success, but the number of competitors that have emerged in the market will continue to challenge VMware's dominance for the long term.

Containers: Up-and-Comer or a Solution in Search of a Problem?

As people adopted virtualization, it became clear that this hardware abstraction technology was not the end when it came to new ways to run workloads. Virtualization alone has a lot of overhead. You have to emulate a bunch of hardware, install a complete operating system, and then install the libraries and prerequisites demands by individual workloads. It's great for monolithic applications, but it's not so great for an emerging class of microservices-based workloads that are comprised of a multitude of tiny instances. Can you imagine the issues, for example, that Apple would have if, every time you made a request to Siri, you had to wait two minutes for an instance to boot so your voice request could be carried out? I bet you'd hate it.

What if you could cut out the middleman and just go straight to the libraries part? What if you could run lots of isolated workloads all on one operating system instance? That's where containers come into play. Containers are fully standalone packages that have everything needed to run the workload inside the container. Containers are isolated from everything around them, too.

Many people confuse containers and virtualization, incorrectly thinking that containers are the next wave of virtualization. In reality, containers run on top of an operating system. There is no hypervisor, although some people equate the container engine to one. The operating system-centric nature of containers means that they can run inside a virtual machine or on bare metal. It's up to you where and how you want to run them.

Figure 2-1 gives you a look at how containers compare to bare metal and virtualization.

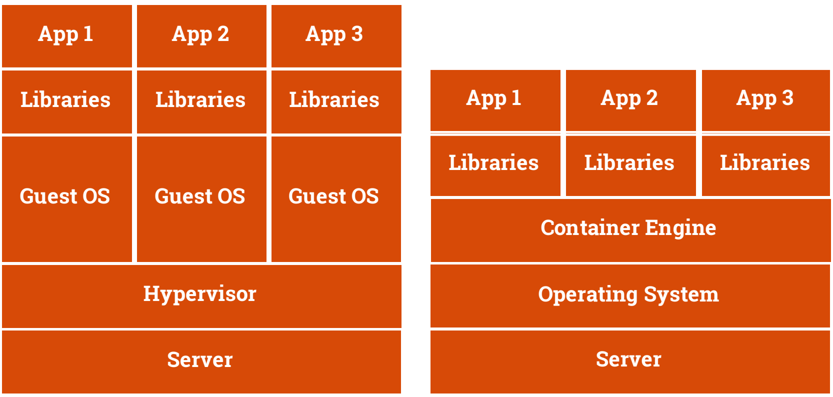


Figure 2-1. Virtual Machines vs. Containers

There has been a lot written about containers in recent years, and their popularity continues to increase. However, they have had—and will continue to have—a far different trajectory than virtualization did. Containers have proven to be very capable for microservices-based activities, but many of those kinds of applications are still in their infancy. If you recall the early days of virtualization, the challenge was getting vendors to support their workloads in a virtualized environment. Technically, most applications ran just fine in a virtualized world. After all, from the perspective of that application, nothing really changed. There was still a full operating system, and all of the resources that the application expected to see were present.

That is not so with containers. Containers have just enough parts to run whatever workload is inside, and that's it. No other overhead is included. That makes them ideally suited to a microservices-based mentality. As applications that leverage this architecture continue to be developed, expect the use of containers to go on the upswing but not necessarily for current mainstream workloads. Of course, this also means that you'll need an infrastructure that can support containers.

OpenStack: Bright Past, Bleak Future?

OpenStack began life in 2010 as a joint venture between NASA and Rackspace. At first, OpenStack appeared to be the perfect private cloud tool to counter Amazon, and many other enterprise IT vendors jumped aboard to help with development. OpenStack aimed to be an infrastructure-as-a-service (IaaS) offering for private cloud. OpenStack integrates with a lot of solutions available on the market. For example, you can choose from among a plethora of hypervisors to use in your OpenStack deployment. You don't have to settle for just one. The same goes for a number of other OpenStack components.

However, confusion abounds with OpenStack outside a core group of adopters. People don't know what it is, and they don't know how to set it up. Entire companies have launched in an effort to ease the deployment difficulties around OpenStack. It's also not proven to be all that efficient in many ways, requiring a lot of staff time and a lot of overhead to operate. Of course, over time, it's gotten better, but it's still more complex than many people want, particularly in an era in which simplicity reigns supreme.

Within the past year, some core support for OpenStack has gone away as HPE and Intel have either concluded or reduced their participation. However, these defections aside, OpenStack continues to do well in very large enterprises and service provider arenas.

Chapter 3

Data Center Architecture & Cloudistics Composable Cloud

Way back in the dinosaur age of computing—the first generation of IT after mainframes—infrastructures were siloed architectures in which compute, storage, and networking were selected separately. But, the selection process was just the beginning. After the selection process and the delivery of equipment to the loading dock, IT then had to deploy, configure, and manage each of these silos. And even that wasn't the beginning. The most frustrating part came when attempting to integrate all of the pieces. The unfortunate reality is that things from different vendors don't always play nice together.

In 2007, though, a new architecture came into being. This second generation of infrastructure—dubbed *converged* infrastructure—emerged as bundles that combined existing server, storage, and networking products together with management software. These solutions effectively eliminated the integration problem since everything arrived already tested and ready to go. IT just had to plug it in and turn it on.

But even that wasn't enough. Although converged infrastructure made it easier to buy and install new data center gear, it didn't really address the ongoing operational complexity inherent in data center equipment. Simply put, this stuff takes a lot of time, talent, and treasure to keep running in tip-top shape. Over the years, decision-makers have started to take a far more critical look at expanding IT staffing and, often thanks to what they see in the cloud, want more efficiency and less overhead.

And that's partially what brought us the third generation of infrastructure. The term *hyperconverged infrastructure* was coined in 2012 to describe an emerging technology that had recently hit the data center scene. Hyperconverged infrastructure offers tightly coupled compute and storage hardware that effectively eradicates the need for a dedicated storage layer in the data center, thereby eliminating what is arguably the most complex resource that IT departments have to contend with. And this is pretty much where hyperconvergence stops. It does a great job at servers and storage, but networking hardware is not typically included as part of the overall system and must be procured separately. So, hyperconvergence was a start, but there was a long way to go.

But…all is not well.

Customers are *still* looking for infrastructures that meet or beat even hyperconverged infrastructure systems in terms of cost, ease of deployment, and ease of use, while overcoming the limitations inherent in legacy approaches to data center infrastructure. Table 3-1 gives you a look at how the various technologies stack up against one another. In this table, you can see, in general terms, how traditional legacy infrastructure options compare to both converged and hyperconverged infrastructure solutions.

Infrastructure Comparison	Legacy	CI	HCI
Workload Diversity			
Independent scaling	✓	✓	X
Predictable performance	✓	✓	X
Efficiency & Cost			
Low cost	X	=	✓
Full virtualization	X	X	=
Efficient redundancy	✓	✓	=
No need for 3rd party SW	X	X	X
Performance			
High performance	✓	✓	✓
Avoid VM storage xlation	X	X	X
Network microservices	X	X	X
Predictable performance	✓	✓	X
Simplicity			
Single silo management	X	✓	=
Software defined	X	X	✓
Eliminates need to choose	X	X	X
VM-centric	X	X	✓
Composable	X	X	X
Resiliency			
Nondisruptive upgrades	✓	✓	✓
Site failover	✓	✓	✓
Low RTO/RPO	✓	✓	✓
Efficient storage failure	✓	✓	X
Open/Future-proof			
Standard hardware	X	X	✓
Open standards	✓	✓	✓
Native Docker support	X	X	=
Key	✓ - strong	X - weak	= - middle

Table 3-1: Comparing IT Infrastructure Options

None of these first three generations of IT architecture, which includes hyperconverged infrastructure, met all the needs of all customers. The fourth generation of infrastructure, however, solves the problems of the prior three generations and enables companies to surge forward with IT without the baggage of legacy infrastructure.

This is thanks to what is known as *composable infrastructure*. Composable infrastructure makes it extremely easy for customers to procure, deploy, manage, and consume IT while meeting all of their performance, resiliency, and cost objectives. Further, composable infrastructure solutions align well with Gartner's view of a third phase of integrated systems (http://www.gartner.com/newsroom/id/3308017). We'll be going into far more detail on this throughout this chapter.

Let's go a bit more in-depth into each model so you can better understand where things stand. By the time you're done with the chapter, I think you'll agree that composable infrastructure is where IT's at!

Comparing Data Center Architectural Models

Today, there are quite a few different data center architectures you can consider running in your data center. You've probably heard of most of these, but you may not know the pros and cons of each. So, let's make sure we're all on the same page here. (Get it? It's a book, same page…oh, never mind.)

Traditional Infrastructure

When you think of traditional infrastructure, you should be thinking of siloed servers, storage systems, storage communications fabric, network gear, hypervisor, and management tools. Read on to discover the challenges and opportunities inherent in these kinds of systems.

Workload Diversity

Traditional infrastructure can run just about anything. It's really quite flexible. You can deploy any combination of hardware and software that makes sense. You can run bare metal, virtualized, and containerized workloads.

Frankly, traditional infrastructure provides more workload diversity opportunity than many other options on the market.

Efficiency and Scaling

But, while traditional infrastructure is super flexible, it's also super inefficient in many ways. If you're running bare metal workloads, you're not generally making the most efficient use of the hardware, for example. Beyond that, the need to manage individual resource silos all the time just makes it even less palatable over time.

Moreover, as you find yourself needing to grow the environment, you have to do so on a resource level, which is a double-edged sword. On one hand, it's great that scaling the environment can be done at the resource level, allowing you to try to target these resource expansions independently. On the other hand, scaling a traditional environment can sometimes be difficult, particularly when it comes to storage. There is the storage itself to think about, but also the fabric. If you're using something like Fibre Channel, there is some additional complexity and cost to consider.

Cost

Traditional infrastructure is pretty expensive, especially when you consider everything that you need to put into the equation, including:

- Acquisition/purchase cost.
- Deployment time (staff salaries or consultant bill), which can also include dealing with interoperability issues between components.

- Ongoing administration for each resource, which may include hiring specialists for each one.
- Eventual replacing, which requires a migration to a new environment every few years.
- A hypervisor and all of the management tools necessary to administer the environment.

These are just the big costs. There are a number of areas where there is the opportunity for your company to pay even more to maintain a traditional environment.

Ease of Use

If you're looking for where things *really* break down, it's here. While many people have become proficient in managing and using traditional infrastructure, it has become increasingly complex to manage as it grows beyond its original plan. Everywhere you turn, there are new places from which to manage certain aspects of the environment. There are consoles for servers, for the hypervisor, for storage, for data protection, for networking, for the storage fabric, and the list goes on.

The complexity problem is a serious one, particularly when comparing your traditional environment to other options on the market. Traditional approaches carry the most administrative overhead and are often the most expensive to operate.

Resiliency and Business Continuity

This category consists of the ability to perform nondisruptive upgrades, maintain low RPO and RTO, provide for automated site failover, and support efficient storage failover. In other words, what happens when an incident takes place? How difficult is it to construct an environment that can fail with minimal data loss? Also, how difficult is it to maintain operations when you have to perform routine maintenance?

With a traditional environment, it can be difficult to get all of this. Yes, it's possible, but since there is no homogeneity in how workloads operate—they can be bare metal, virtual, or container-based—you have to look at every method and make plans around each one. That adds cost and complexity to what can already be an expensive need to support.

Integrated Systems – Converged and Hyperconverged Infrastructure

Among all of the crazy downsides of traditional infrastructure, a few really stand out. First, is the buying process. You have to buy individual components and build it all yourself, which can take an inordinate amount of time. CFOs *love* being told that their multimillion-dollar infrastructure purchase will take a year to deploy. The previous section discussed all of the other potential downsides to the traditional route.

To address these and other concerns, converged and hyperconverged infrastructure hit the scene, creating a category that is now often referred to as *integrated systems*. These systems, as the category name implies, are integrated to different extents. With converged infrastructure, the integration is simply repackaging existing components. Rather than forcing customers to individually spec, buy, deploy, and configure everything, with a converged solution the vendor does that for you. You just buy a rack of pre-configured equipment, roll it into the data center, and turn it on. Figure 3-1 gives you a look at what a converged infrastructure system looks likes.

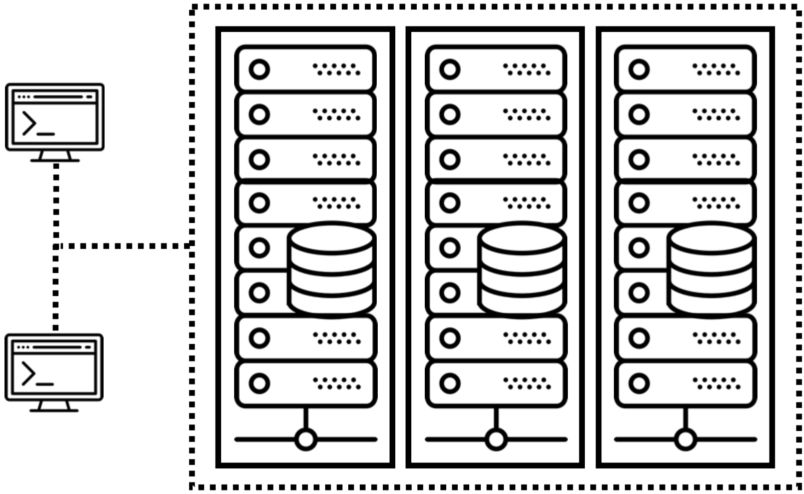


Figure 3-1. Three racks of converged infrastructure managed by one or more consoles, depending on solution

With hyperconverged infrastructure, you get to take things a bit further. With these solutions, there is no more storage. Well, there's storage, but it's not a standalone component anymore. With these solutions, storage as a standalone component has been ripped out and returned to the servers where a powerful storage management layer awaits. This management layer runs as either a virtual machine or as a module in the hypervisor kernel, where it manages all of the local storage. In Figure 3-2, you see a depiction of a hyperconverged infrastructure cluster.

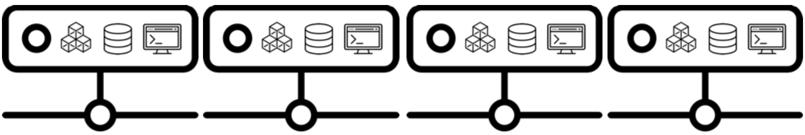


Figure 3-2. Four hyperconverged infrastructure nodes, each containing compute, a hypervisor, storage, network connectivity, and management capabilities

As you may have guessed, these integrated systems come with their very own pros and cons.

Workload Diversity

Let's start with workload diversity. With a converged solution, you're free to run anything you like, just as you can with traditional approaches. With hyperconvergence, though, you're going to be running everything virtualized. That's a prerequisite for the technology. The individual nodes that comprise the solution are all running a hypervisor, and all of the workloads run virtualized. That said, you can still run container-based workloads in hyperconvergence, but they generally have to run on top of a virtual machine, which is a fully supported container deployment model.

So, you're not going to get bare-metal workload support in hyperconvergence, but you will with converged infrastructure. If you're highly virtualized, the inability to run bare-metal workloads in hyperconverged infrastructure systems won't matter much, but if you're fully virtualized it could be a problem.

Efficiency and Scaling

Integrated systems are often less costly and more efficient than traditional solutions, but in different ways. Converged solutions, for example, may cost a bit more upfront than traditional solutions, but you get to eject the purchasing and deployment complexity, which can be an invaluable investment that can bring down the total cost of ownership (TCO). But, as is the case with traditional infrastructure, neither converged nor hyperconverged infrastructure address any network complexity via network virtualization. Further, only hyperconverged infrastructure provides any semblance of native storage virtualization. Although some converged infrastructure solutions may rely on virtualized storage, it's a hit-or-miss proposition.

Cost

If you look at the total cost of ownership for hyperconverged infrastructure, it can be far lower than what you get with traditional and even converged infrastructure. This is due to the removal of storage as a standalone resource, simplified administration, often lower upfront costs, and easier, more granular scalability. As mentioned in the previous section, converged solutions often carry a smaller total cost of ownership than traditional ones.

Ease of Use

Beyond deployment, converged infrastructure and traditional deployment look pretty much the same in the ease-of-use department. Hyperconverged, though, has a bit more going for it. By eliminating the storage silo, a whole lot of complexity is gone, as are some administrative interfaces. The virtual machine-centric nature of hyperconverged infrastructure also generally makes it easier to manage than other solutions

Composable Infrastructure

So, at this point, you're probably all dejected because you realize that there's nothing out there that's ideal for the data center. And then… you remember what this book is about and you start thinking, "Well, wait…what about composable infrastructure? How does that compare?"

I'm glad you asked! But, before we jump into comparisons, let's chat about what *composable* is. At its core, you can think about composable infrastructure as the next wave of an integrated systems approach. It solves a number of the issues inherent in other models.

The Cloudistics View of Composable Infrastructure

Cloudistics' definition of composable systems aligns well with Gartner's definition of the third phase of integrated systems from 2016–2025. In Gartner's view:

"The third phase of integrated systems will deliver dynamic, composable, and fabric-based infrastructures by also offering modular and disaggregated hardware building blocks, driving continuous application delivery and continuous economic optimization."

The remainder of this book will explain how composable infrastructure–specifically, Cloudistics' implementation of the technology–serves to help customers overcome the limitations inherent in other solutions.

Composable infrastructure:

- Tightly integrates server, storage, networking, virtualization, virtual machine, and container-centric management.
- Includes a comprehensive application marketplace, making enterprise app deployment as simple as installing Words with Friends on your iPhone.
- Includes native virtualization of compute, storage, and *networking* resources (which is often a forgotten resource in other architectures) without need for additional third-party software.
- Allows every resource to scale easily and independently of each other.
- Uses commodity off-the-shelf x86 server nodes, storage drives, and network switches—*composable infrastructure is entirely software-defined.*

- Employs a single unified management portal to manage the entire infrastructure using an API.

Cloudistics Composable Cloud Components

Cloudistics delivers Composable Systems, an integrated fully software-defined infrastructure platform that includes native virtualization of compute, storage, and networking resources able to run applications out of the box. Easy!

The Cloudistics platform ships with all of the software and hardware you need to run your applications. In fact, the platform itself is capable of running all of your enterprise applications. Perhaps one of the most challenging aspects of traditional infrastructure is what has become known as *time-to-value*. When CFOs approve massive data center purchases, rarely are they thrilled when they discover that the time of purchase order to full deployment is measured in terms of *months*. Like any investment, the business wants to begin to realize a return relatively quickly.

And no one wants a solution that is hard to use. Simplicity is in! Even the geekiest geek these days often throws their hands up and says, “Just make it work!”

To support fast time-to-value, reasonable cost, and ease-of-use, the Cloudistics platform doesn’t require any additional third-party software or hardware to run your workloads. Heck, you don’t even need to buy a hypervisor since that’s built in, too.

Remember that whole thing about ease of use? I'm willing to bet that you have an iPhone or Android device sitting near you. And, you probably have a lot of apps loaded. Was it painful to get those apps installed on your device? Probably not, since you more than likely used the app store made available by your device's manufacturer.

Cloudistics has you covered there, too. Included in the platform is an application marketplace for enterprise apps, which allows simple point-and-click deployment of common applications. The resources needed to run an application are composed—that is, appropriate resources are carved out of the central pool and made available to the new workload—on the fly, and intelligent workload placement with service-level agreement enforcement ensures guaranteed performance.

As a hyper-scale architecture, the Cloudistics platform allows each discrete resource—servers, storage, and network—to scale independently without many of the traditional clustering limits inherent in other architectures, such as hyperconverged infrastructure, in which compute and storage are coupled and must scale together.

Rounding out this high-level overview is the fact that the platform supports full multi-tenancy and provides self-service capabilities for different organizations, customers, tenants, or workloads, making it attractive for both traditional data center customers—both large and small—and for Managed Service Providers (MSPs).

Pretty cool, eh?

Hardware Overview

Rather than spending time and money building unnecessarily complex hardware components, Cloudistics uses off-the-shelf hardware. The company's initial implementation leverages the latest Dell PowerEdge FX hardware, although support for other hardware platforms is expected to debut over time.

Although it's software that makes the Cloudistics platform tick, the underlying hardware is important. It's still critical that the hardware has the capability to get the job done in a way that overcomes the challenges found in traditional and even hyperconverged infrastructure approaches. To that end, the Cloudistics composable platform has separate compute, storage, and network blocks that are *independently scalable*.

That, folks, is the real power of composable infrastructure. You get hyperconvergence-like benefits, but you're not locked into having to scale resources that you don't need. If you need more storage, you just add more storage, and you do so without having to resort to expending unnecessary resources like you do with hyperconverged infrastructure. You don't have to add a bunch more compute that will go to waste and for which you sometimes incur additional hypervisor licensing costs. But, at the same time, you still get the centralized, streamlined management experience. One of the great things about hyperconverged infrastructure is the fact that you don't have to manage storage and compute as separate silos. The same is true for composable infrastructure, but you get the added benefit of being able to scale resources independently of one another.

Even better, you have a lot of choice when it comes to expanding. As growth occurs, some companies will want to add a lot of resources, and some will want to add just a few. With the composable platform at its heart, Cloudistics provides a great range of scale options:

- A 2U compute block has anywhere between 1 and 8 Intel Broadwell blades, with up to 576 hyper threaded cores and 4 TB of RAM.

- A storage block has from 6 to 192 usable terabytes of all-flash storage and 2 disk controllers. The result is an enterprise-grade feature set with the performance of proprietary RAID arrays, but without the cost.

- The network block has redundant SDN-enabled ONIE (Open Network Install Environment) switches, each with 48x10 GbE and 6x40 GbE ports, running the Pica8 network operating system (NOS). A 32x100 GbE switch is also available for even higher performance or for use as a spine switch in large deployments of the composable platform. The open switches used in the Cloudistics composable platform are less expensive than proprietary switches from traditional switch vendors, but they have all the enterprise functionality needed for use within the composable platform.

Software Overview

You know that the hardware is important, but it's only important insofar as its ability to provide resources to higher-level workloads. And, those workloads need constructs in which to operate. These days, the most common construct is, wait for it…a virtual machine! As you may be aware, virtualization is a thing. However, even as virtualization swept across the data center landscape, it left some negative things in its wake, including, in many cases, a lot of cost in the form of hypervisor licensing and new needs around managing workloads.

Hyperconverged infrastructure made a valiant first attempt to rein in some of the management challenges, but it didn't go far enough in many cases. It's in the software where the real power of the Cloudistics composable platform becomes very apparent. The solution consists of four separate software components, each described in the following sections.

Adaptive Operating System (AOS)

Not to be confused with the decades-old Data General operating system with the same initials, the Cloudistics Adaptive Operating System (AOS) is a KVM-based hypervisor. AOS runs atop all the compute nodes in your shiny new composable system, thereby eliminating the need to separately purchase virtualization licenses, modules, or costly license renewals. In other words, with Cloudistics, you don't need to buy VMware or Hyper-V. AOS' KVM-based nature takes care of the virtualization piece for you.

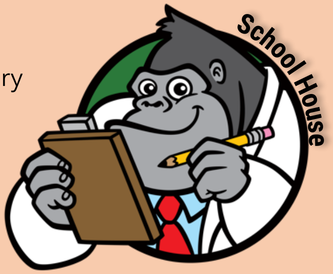
And, don't think that avoiding the high price tag of other hypervisors means that you have to skimp on features. AOS supports a number of critical features that can help you make the data center great again:

- High availability
- Full and para-virtualization
- Live migration
- CPU/memory/storage oversubscription
- CPU and memory hot-plug. An application running inside the composable platform can be imbued with more memory or processing cores without having to stop and restart the application.
- Built-in security

But, the platform doesn't stop at virtualization. Today, there is far more to the story. Right out of the box, AOS also supports Docker containers, both stateless and stateful. To support stateful applications, Cloudistics provides a Docker Volume Driver that leverages the platform's Elastic Block Flash (EBF) layer, which is comprised of efficient, economic, and scalable commodity hardware. We'll talk more about this later on.

To support containers, Cloudistics uses the concept of a Zone Virtual Machine. A Zone VM includes the Cloudistics Docker Volume Driver as well as the Docker Engine. You can cram lots of containers inside a single Zone VM. Once a Zone VM is started, the individual containers running inside it can be created and removed very quickly—much faster than is possible with traditional virtual machines.

Migration Zones



A Cloudistics Migration Zone acts as a boundary around a set of compute nodes. It provides a way to organize a collection of resources virtually and consists of a defined set of compute nodes (with any categories/tags) and connectivity to one or more storage pools. One compute node can belong to one migration zone.

Application instances can migrate among nodes within a migration zone but cannot migrate to nodes outside of a migration zone. Migrating an instance from one compute node to another within a migration zone is equivalent to VMware's vMotion: there is zero downtime, continuous service availability, and complete transaction integrity. When an instance starts in a particular migration zone, it will only utilize the compute nodes in that zone.

The approach used by the Zone VM is superior to the alternative approach sometimes used in the industry in which each container runs in its own virtual machine. Further, although you can run containers without any virtualization layer at all, by running them on the Cloudistics platform’s Zone VM, you are able to preserve two key advantages:

- Isolating the container from compute node failures
- Retaining the ability to live migrate containers

It’s important to note that you don’t need to choose between containers and virtual machines. Regular virtual machines and Zone VMs with containers can be run simultaneously on the platform, providing a great deal of flexibility and agility.

Adaptive Overlay Networking (AON)

One thing you don't hear much about in the world of hyperconvergence, but that plays a central role in the composable world, is networking. Like all resources, the network is essential, but prior to composable, it was simply a resource to plug into.

Those days are over!

The Cloudistics Adaptive Overlay Networking (AON) software component provides standards-based network virtualization operating at line speed. While many infrastructure solutions support virtualization of compute and storage, the Cloudistics platform is the first to throw in native support for network virtualization. By including this critical component of the software-defined data center (SDDC), the solution gets you closer to operating in a cloud-like environment.

How is this feat accomplished, you may ask?

Network virtualization allows virtual networks to be dynamically created and destroyed in response to application needs, without requiring the manual reconfiguration of the underlying physical network. When people hear the term *virtual network*, they often wonder how these differ from the virtual local area networks (VLANs) that have been a part of the networking landscape for decades. First, these modern virtual networks are much easier to use and manage than VLANs. Second, unlike VLANs, they support the ability to preserve network addressing during workload migrations, and they are not limited to 4096 sub-networks.

Network virtualization includes support for two key technologies:

- **Overlay networking.** Overlay networking allows VMs to communicate using virtual network addressing. The overlay network provides the agility needed to support modern applications; the underlying network provides the robustness needed in enterprise data centers.

- **Network Function Virtualization (NFV).** Network Function Virtualization implements networking functions, such as switching, routing, and firewalls, in software rather than using custom chips to handle these activities.

When it comes to overlay networking, Cloudistics is based on completely open standards, including standard IPv4 and standard switches and routers. Unlike other approaches, the composable platform doesn't require switches to be specially configured. Instead of compute-intensive encapsulation, the composable platform uses address substitution allowing the virtual networks to operate at line speeds without special hardware support.

Through network virtualization, Cloudistics supports microsegmentation to help you improve your organization's security. By implementing microsegmentation at the network layer, the composable platform secures individual applications into secure islands with additional security achieved through the platform's use of application-level firewalls. Application-level firewalls operate, as the name suggests, at the application level, where traffic into and out of an application is analyzed for validity.

Elastic Block Flash (EBF)

Next up on the software front is something that, on the surface, sounds eerily like hardware: Elastic Block Flash (EBF). EBF is the software application that runs on each of the storage controllers in your composable environment. The storage controllers are standard x86 servers with shared access to a series of hot-swappable enterprise flash disks. That simplicity makes ongoing management a breeze and makes it really easy to scale the environment.

EBF is a federated, all-flash storage subsystem that is designed for performance, redundancy, scalability, and manageability of virtualized and containerized workloads. That's a long list of goodness! Storage controllers come in pairs, and, as you add additional controller pairs,

Microsegmentation

Over the years, organizations have taken great pains to secure their network perimeter, but, past the firewall, there is still a soft, gooey interior just waiting to be exploited. The unfortunate reality is that once an attacker gets past the firewall, getting to everything internally is a breeze. Remember, getting past the firewall isn't as hard as it's often made out to be. A simple email message with a link to a ransomware-laden download is all it takes for your internal systems to be at high risk.

Microsegmentation is a method by which your general security posture is increased by dividing your data center into multiple smaller segments and applying traffic rules that dictate what is and is not allowed to talk with systems on each segment. This segmentation technique effectively limits an attacker's lateral movement between systems; a compromise on one system doesn't give the attacker carte blanche access to every other system in your data center.

EBF performance scales linearly. A pair of controllers coupled with related storage is called a *storage block*, which , when grouped, are referred to as *storage pools*.

It's important to keep that data safe! To that end, storage data protection includes RAID, hot sparing, and in-drive erasure coding. In addition, the controllers are fully redundant with automatic failover in the event that one of them runs into an issue and decides to take a break. If the primary controller fails, the secondary takes over immediately with zero downtime, no data loss, and no rebuilds.

As you might have guessed by now, there is a theme across the entire platform: virtualization. The storage layer is no exception. The Cloudistics EBF uses a unique approach to storage virtualization. Clustering and cross-node locking protocols are avoided in the

Cloudistics platform, allowing storage to scale without traditional limits. This means you get to grow storage more easily. Unlike other approaches, there is no file system or any other translation layer—virtual machines and containers have direct access to storage, resulting in all the raw performance of flash being directly available to these upper-level constructs. As a result, EBF delivers the highest I/O performance possible while maintaining all of the features you'd expect to find in an enterprise storage system, including:

- Thin provisioning.
- Single-instance storage.
- Live snapshots, cloning.
- On-the-fly storage expansion.
- On-the-fly VM disk expansion.
- Asynchronous disaster recovery.
- Incremental backups.
- Self-encrypting drives.

Cloudistics Intelligent Cloud Composer and Management Portal

The fourth and final component in our overview of the Cloudistics Composable Cloud is around how this whole thing is managed. As you may know, one of the biggest trends in the data center market today is simplicity. No longer are CIOs willing to deploy stuff that takes hordes of people to manage. Doing business like that is expensive, and it results in other bad things for the business, including increasing the time it takes to get new services up and running.

The Cloudistics platform is managed by the Intelligent Cloud Composer and Management Portal, which enables complete control of the platform from a single pane of glass and provides composability and usage metering. Usage metering is sometimes overlooked as an optional component of the private cloud. However, in order to be considered a real cloud-like environment, it's really important to understand who is using what—a need that is even more critical in environments that use chargeback mechanisms.

The Cloudistics Cloud Composer supports intelligent workload placement and service-level agreement (SLA) enforcement. Allowing the infrastructure to automatically decide where to place workloads frees up administrators to do better things. After all, the infrastructure probably knows more about what's currently happening than any human could. With the intelligent workload placement algorithms, workloads are placed onto compute nodes and storage pools only after the algorithm identifies throughput, bandwidth, and storage capacity needs of the workload. In addition, space-saving opportunities are considered. The composable platform also permits even more sophisticated placement algorithms, including those that factor in ownership of resources and availability of particular features on the nodes.

The portal gives visibility into everything from low-level hardware status—right down to the fans—all the way up to the application level, and it can support everything from an entry-level system to a large multi-data center implementation.

The health monitoring support system automatically alerts the ever-present Cloudistics support team about potential performance issues. This means that Cloudistics can see failures and initiate replacement service requests before the customer even knows there's a problem.

Chapter 4

Cloudistics Composable Cloud Outcomes

Meeting Customer Requirements

The goal with any data center architecture is to ensure that customer requirements are being met in every way possible. As you've learned in previous chapters, there are some fatal flaws inherent in some of the options that have been available to customers up to this point. With the advent of composable infrastructure, these flaws are addressed, enabling a series of outcomes, each of which is explained in the sections that follow.

Elevating Administrators from Specialists to Generalists

There was a day, not that long ago, when being an "IT generalist" was sometimes looked down on. Super-smart specialists would walk by those folks and snort in derision as generalists did a bunch of things while the specialists got to dive deep into a single resource area and become the alpha guru on that resource.

The tables have turned! Today, from a business perspective, specialists are considered too expensive and too hard to find, and organizations are seeking ways to reduce their reliance on them. As a result, we've officially entered the era of the generalist, and it's a beautiful thing.

Rather than needing people who have PhDs in things like storage, companies are now looking for people that have a series of Bachelor's degrees in a wide variety of disciplines (figuratively, of course). They need people that have a reasonable understanding of *all* of the elements of the data center and how the various puzzle pieces actually fit together. The thinking today is of a bigger picture than it was in the past.

These kinds of changes don't just happen in a vacuum. First of all, humans can't exist in a vacuum, but that's a different problem altogether. There need to be fundamental changes in the data center environment in order for generalists to have a shot at survival. We need simple systems that, frankly, just work and that don't require constant gymnastics to deploy, manage, and expand. We need systems that provide great application diversity while not breaking the budget.

Workload Diversity

No two workloads look the same. Different workloads carry different compute, storage, and networking requirements. This fact alone makes it really important to be able to scale these resources independently. Coincidentally, the Cloudistics platform has separate compute, storage, and network blocks, with any-to-any connectivity between the blocks. Within the integrated system, any one of these three blocks can be expanded without impacting the other blocks.

This makes it incredibly simple to scale compute, storage, or networking as needed to support a given workload or set of workloads with no specialized expertise required. In the Cloudistics Composable Platform, the architecture allows compute to be scaled to almost 700 nodes (50,400 hyperthreaded cores), storage to be scaled to tens of petabytes, and networking to be scaled to tens of high-performance leaf and spine switches. Architectural extensions that will support another factor of ten in scaling are in development.

This is a lot, in case you were wondering, and it's all managed with generalists. How?

Let's start with storage. The platform's storage blocks are built using dedicated commodity x86-based servers. Storage services, such as thin provisioning and data deduplication, are provided on a virtual machine or container basis, not on a LUN or volume basis. All the complexities of SAN management are eliminated and are hidden away from the administrator or user of the system.

In addition, thanks to the fact that the platform uses all flash storage in EBF, it's much easier to support guaranteed performance for applications. No nerd-knobs to tweak performance are required.

Efficiency and Cost

A good platform doesn't have to cost an arm and seven legs. The cloud has the cost issue nailed for many, and CFOs and other financial decision-makers look at the cloud and think, "I gotta get me some of that." Cloudistics provides complete virtualization of compute, storage and networking, making it possible to allocate exactly the resources needed—no more, no less—to each application. A low-overhead hypervisor integrates compute and network virtualization and runs on the compute nodes, while storage is implemented on a separate and dedicated set of nodes. The whole system is, at the same time, tightly integrated, but also loosely coupled.

But all the cost savings in the world won't make a difference if, for example, your storage is terrible. In the Cloudistics world, even storage resiliency is provided in a space-efficient—and therefore cos- efficient—manner. It might sound old-school, but Cloudistics goes retro (in a good way!) by using RAID instead of creating multiple copies of data. Each drive uses erasure coding to handle uncorrectable errors. Each group of seven drives is organized in a 6+P RAID-5 configuration with two global spares. RAID-0 is implemented across a group of thirty drives. This brings the power of all of those drives together to provide maximum performance while also being incredibly capacity-efficient.

But there's more to the story:

- The platform uses a single network for everything, including storage.
- Data reduction techniques, such as deduplication, thin provisioning, and snapshots are a base part of the platform.
- Data services, such as snapshots and replication, are space- and bandwidth-efficient.
- Since the platform includes the hypervisor and network virtualization software, no additional licensing costs are incurred, unlike most other systems.

6+P RAID-5. What the What?

If you're reading this book and wondering what I mean by "6+P RAID-5", never fear! Here's an explanation and it's pretty easy once you see it.

You probably know what RAID-5 is. It's a data protection scheme that combines data disks with a disk of capacity dedicated to parity, which is used to rebuild data in the event that a disk fails. 6+P simply refers to a RAID set that includes six disks plus a parity disk.

Price comparisons in Figure 4-1 show the Cloudistics platform to be at least 30% cheaper than best-of-breed hyperconverged infrastructure systems and at least two times cheaper than best-of-breed public cloud systems.

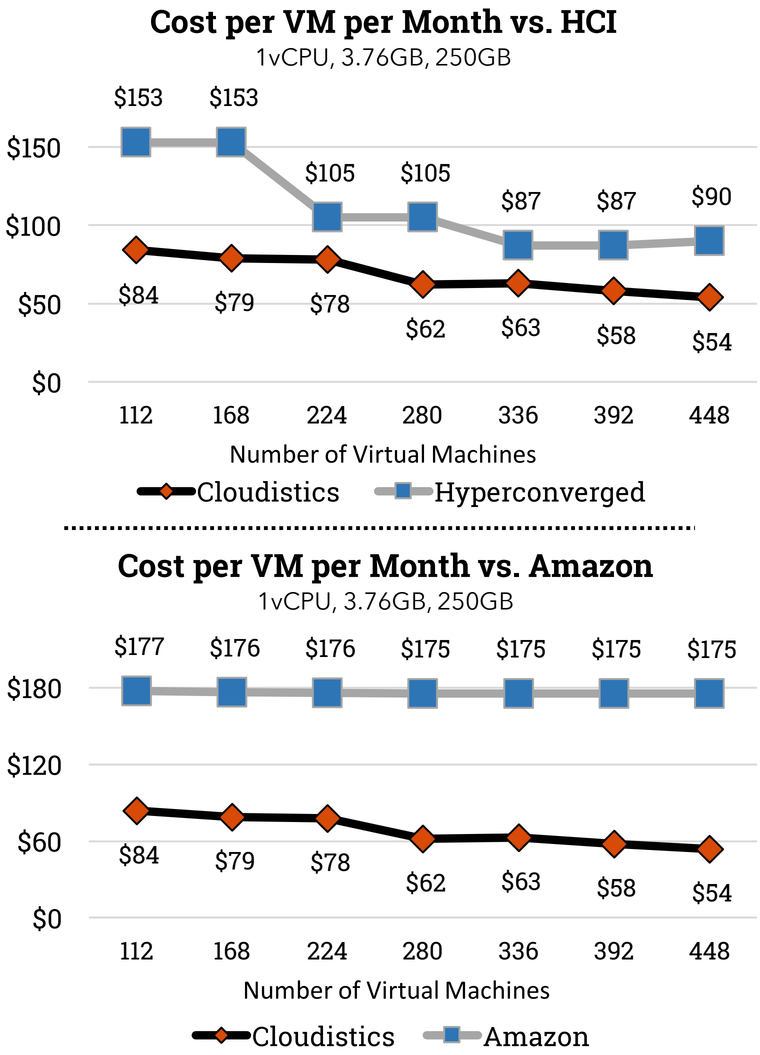


Figure 4-1. Cost comparisons of hyperconverged infrastructure and Amazon vs. Cloudistics composable infrastructure

High Performance

Getting good workload performance is generally considered a positive trait in any infrastructure environment. You need hardware and software components that work together to maximize performance. Too often, people think that “hardware doesn’t matter” in the world of software-defined storage, but nothing could be further from the truth.

You still need decent hardware. Even the best software can’t make a turtle run like a cheetah. The products that Cloudistics has initially chosen for its platform are very intentional and consist of some of the best hardware available in the market today.

The integrated Elastic Block Flash (EBF) storage has the performance of a best-of-breed modern all-flash array and utilizes modern 4 TB Samsung Enterprise SSDs with 3D V-NAND technology and drive-level erasure coding. Furthermore, the blocks from the EBF are made directly available to the applications inside the virtual machine, without the traditional block to file to block translations of hypervisors (for example, LUN to VMFS to VMDK in VMware). This improves overall performance. The dedicated x86 server nodes implementing storage function are powerful enough that they can support the full read/write performance capability of the flash drives attached to them. Each EBF storage block can support 200,000 IOPS at 120 microseconds of latency and can support up to 40 Gb/s of bandwidth to a single application, which matches or exceeds the capability of competing modern all-flash arrays.

On the networking side, the system includes high performance top-of-rack switches, each with 48 x 10Gb ports and 6 x 40 Gb ports, and spine switches with 32 x 100Gb ports. Cloudistics implements network virtualization in Layer 3 using an innovative new technique that allows for line rate performance. Figure 4-2 provides you with an overview of the technique.

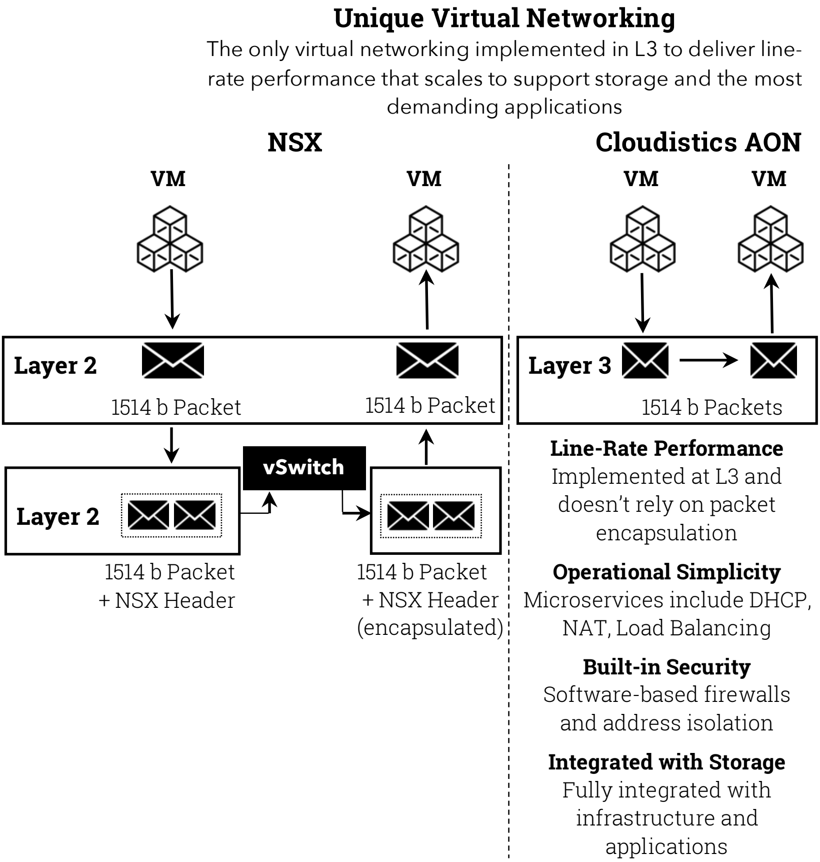


Figure 4-2. The Cloudistics AON platform carries with it a number of benefits over other virtual networking approaches

Virtual machine-level microservices, such as routing and firewalls, are all implemented with low overhead and minimal consumption of compute cores. By integrating such capabilities, the platform reduces the need for east-west traffic in the data center (Figure 4-3).

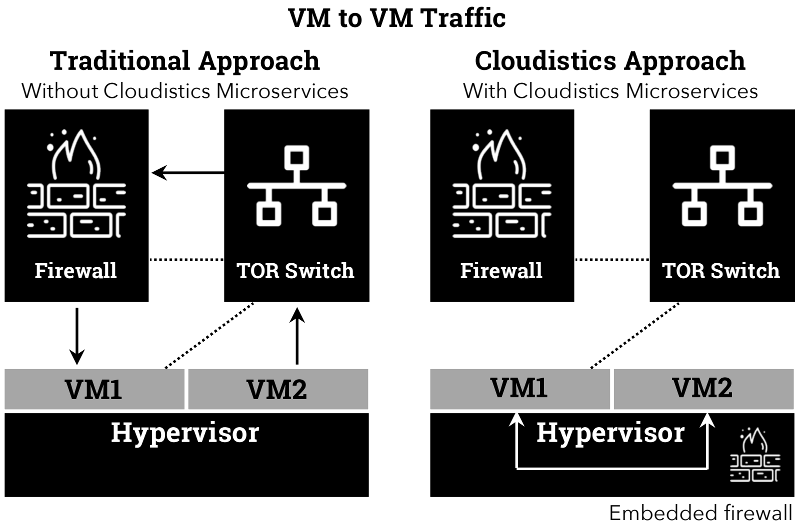


Figure 4-3. The Cloudistics approach to inter-VM networking as opposed to a traditional approach

Easy Administration

For most system admins, simplicity starts and ends with the console they use to interact with the data center environment. For these folks, Cloudistics includes a single software-as-a-service (SaaS)-based portal to manage all aspects of the infrastructure. The portal includes real-time monitoring and alerts so admins can jump right in when something gets wonky in the data center. For its operation, the portal uses APIs that are provided by the underlying hardware.

Since everything in the platform is virtualized, it's a breeze to create processes that enable automated provisioning of exactly what an application needs to operate. You can do this without worrying about over- or under-provisioning resources. There is no need to resort to scripting, CLIs, terminals, and keyboards.

It is really quite easy to *compose,* or carve out, a Virtual Data Center (VDC) that meets the exact and specific compute, storage, and networking needs of an individual tenant or infrastructure user.

But that's just the beginning. The hard part in infrastructure management is often application deployment and configuration. But all is not lost here! With the Cloudistics platform, you're also provided a handy application marketplace that allows you to download popular enterprise applications for free. This is accompanied by a portal, which allows one-click importing of existing virtual machines from VMware or Hyper-V straight into to the Cloudistics environment.

There are a number of other ways that the platform brings a refreshing ease of use to your life, too:

- All storage capabilities, such as replication, snapshots, and thin provisioning are delivered by a virtual machine or container.
- Leverages all-flash—EBF storage—eliminating the complexity of managing multiple storage types and simplifying SLA enforcement.
- Simplifies customer choice by eliminating the need to select between different RAID levels.
- Uses a single network for all needs, including storage.
- Runs side-by-side with your current infrastructure without disruption.

Cloudistics Ignite

The Cloudistics cloud controller, Ignite, is a free, cloud-based management controller that helps you build, deploy, and manage your cloud across multiple locations, all from a single pane of glass.

With Ignite, you can create multiple virtual data centers, control resources for individual applications and users, manage the hardware, deploy and secure applications with microsegmentation, monitor application health, and much more.

Ignite is the reason that you're able to eliminate management silos in your data center. What used to live in multiple onsite management silos can now be managed centrally through the cloud. Whether you are deploying a new VDI instance or a new Docker environment across multiple sites, Ignite is powerful enough to meet the demands of IT professionals and easy enough for someone with little network experience to manage and run your apps at peak performance.

Resiliency and Business Continuity

Reliability and consistency are really important in IT systems. Any architecture that can't provide basic uptime isn't going to be acceptable to, well, anyone. In the Cloudistics environment, the architecture is designed so that there is no single point of failure. The solution sports redundant storage controllers, drives, and network switches. Recovery times are minimal when failures do occur.

Let's dig a bit deeper.

The EBF storage controller features an active-passive design, and any single controller has sufficient capacity to support the I/O to all the drives behind the pair of controllers. In the event of a controller failure, the system simply switches to the passive controller, so there is no loss in performance.

The disks themselves are configured to use RAID. While other solutions on the market today make copies of data rather than use RAID, Cloudistics feels that RAID's smaller capacity overhead provides an all-around better solution. As previously mentioned in this book (but worth reiterating), each drive uses erasure coding to handle uncorrectable errors. Each group of seven drives is organized in a 6+P RAID-5 configuration, there are two global hot spares across 30 drives, and RAID-0 is implemented across the group of 30 drives. Boom! Instant performance and capacity.

One of the times of highest risk in any data center environment is during firmware and other core software updates. Particularly in a traditional SAN scenario, a failed firmware update can wreck your lunch hour. To prevent a single software update from bringing utter destruction to your data center, Cloudistics supports rolling software upgrades. A group of nodes can be defined as belonging to a *migration zone*, and all the nodes in this zone can be updated to the new release of software, while other migration zones continue to operate using the previous release. Up to three releases of software can be running at the same time in different migration zones, so you can do some pretty intense testing. Finally, software updates do not require you to reboot the node, thus making software updates nondisruptive.

The hypervisor bundled with the solution also provides additional availability options, just like you've come to expect from other commercial hypervisors. If the included hypervisor fails, virtual machines can be restarted on alternative nodes.

Another risky situation comes into play when attempting to scale traditional infrastructure environments. With solutions based on composable technology, this risk goes away. Cloudistics supports nondisruptive hardware upgrades and nondisruptive expansion.

Finally, in the area of ongoing availability and recovery, the platform supports very low recovery time objectives (RTO) and recovery point objectives (RPO) and provides for application consistent replication and snapshots. RPO/RTO are foundational to the Cloudistics Platform. Whether at a VM level or for an entire VDC (Virtual Data Center), snapshots are created as often as every 15 minutes. These snapshots can be automatically replicated to another Cloudistics platform in a separate location. Replicants have their own unique retention policy and can be stored as long as needed. Individual virtual machines can also be backed up using most backup tools.

Open and Future-Ready

There's one more critically important component for any infrastructure you put in place today: future-proofing. Technology doesn't stand still. It continues to develop over time and really smart engineers at really smart companies are always throwing new things out into the marketplace hoping that something will stick. These innovations are oftentimes small, but they are sometimes pretty massive.

You need an infrastructure that allows you to adopt the technologies that prove to be transformative.

At its core, Cloudistics uses entirely off-the-shelf hardware such as x86 server nodes, open network switches, and SSD drives. As vendors whip up newer versions of these technologies, they can be seamlessly incorporated into the design in the future. Cloudistics' hypervisor, which is built on top of the open source KVM hypervisor, uses the industry-standard iSCSI protocol internally to connect with Elastic Block Flash. EBF is built on Linux and uses many battle-hardened and robust subsystems from the Linux kernel. Cloudistics also integrates with existing data center networking and provides redundant network switching using the Multi-Chassis Link Aggregation (MLAG) open Ethernet standard.

With NVMe quickly gaining traction in the market, you need to make sure that you can adopt this technology and other innovations as they become available. Cloudistics is ready to support NVMe since the network adapters used are already remote direct memory address (RDMA) capable, which is required by the emerging NVMe fabric standard. Moreover, it can leverage next-generation 3D Xpoint/Optane storage-class memory as it becomes available from Intel and Micron.

Table 4-1 shows how Cloudistics meets all of the requirements laid out in this book and contrasts the Cloudistics solution to the capabilities of hyperconverged infrastructure systems.

	HCI	Cloudistics
Workload Diversity		
Independent scaling	X	✓
Predictable performance	X	✓
Efficiency & Cost		
Low cost	✓	✓
Full virtualization	=	✓
Efficient redundancy	=	✓
No need for 3rd party SW	X	✓
Performance		
High performance	✓	✓
Avoid VM storage xlation	X	✓
Network microservices	X	✓
Predictable performance	X	✓
Simplicity		
Single silo management	=	✓
Software defined	✓	✓
Eliminates need to choose	X	✓
VM-centric (vs. LUN-centric)	✓	✓
Composable	X	✓
Resiliency		
Nondisruptive upgrades	✓	✓
Site failover	✓	✓
Low RTO/RPO	✓	✓
Efficient storage failure	X	✓
Open/Future-proof		
Standard hardware	✓	✓
Open standards	✓	✓
Native Docker support	=	✓

Key	✓ - strong	X - weak	= - middle

Table 4-1. Comparing hyperconverged infrastructure with the Cloudistics composable infrastructure platform

Chapter 5

About Cloudistics

It's time for you to take what you've learned in this book and begin to apply it for yourself by discovering the world's only modular cloud-managed infrastructure. The Cloudistics solution is an on-premises cloud that makes it easy to deploy, secure, and manage applications at a fraction of the time and cost of the public cloud.

Everything Is Included

The platform is a *complete* on-premises cloud solution. With Cloudistics, network, storage, compute resources, hypervisor, and cloud controller work seamlessly to bring the cloud to your on-premises environment in a fraction of the time and at approximately a third of the cost of the public cloud.

Everything is included—all the software and all the hardware to run your cloud. You just rack it, stack it, and add your applications and go! You can choose to create your own application templates or just download pre-configured ones for quick and simple deployment of your most complex applications.

Modular Design: Pay as You Grow

Just choose the resources that are optimized for your application demands—Cores, Memory, IOPS, Latency, and Storage Capacity—Ignite will launch the app and manage its resources for optimal performance.

And don't worry, Cloudistics can scale network, compute, and storage resources independently as your application demands change in real time and without downtime, saving you time and money.

Cloud Managed

What used to live in multiple IT management silos can now be managed centrally through the cloud with the free Ignite cloud console. Whether you are deploying a single site or across multiple sites, Ignite is powerful enough to meet the demands of IT professionals yet easy enough for someone with little infrastructure management experience. And the price? It's free! Every piece of Cloudistics hardware includes a perpetual Ignite license.

Through the Ignite cloud controller, you can control the entire infrastructure from a single, simple-to-use pane of glass.

- Manage everything—networking, storage, compute and virtualization all controlled from a single console.
- Create your own VM templates or deploy pre-configured ones from our application marketplace for quick and simple deployments of the most complex applications.
- Create resource profiles for each application to ensure predictable performance.
- Secure applications within virtual networks in just a few clicks.

Network Simplified

With Cloudistics, you can deploy applications in a secure and micro-segmented virtual network with just a click—no need for manual, error-prone manipulation of switch, VLANs, MAC/IP tables, ports, VTEPs, or firewall configurations. Cloudistics virtual networking makes networks as easy to manage as compute and storage, leading to true data center agility.

Each network can be named and supports network function virtualization services such as firewalls, load balancing, NAT, and DHCP. And, the virtual network works as fast as the real network with no performance loss; you can create 100 Gb/s virtual networks from 100 Gb/s physical networks!

All-Flash Speed

The Cloudistics native elastic block storage is optimized for high-speed applications. Each 2U storage block supports 200,000 IOPS with 150 microsecond latency (assumptions: 4K block size with a 70%/30% read/write mix). And multiple storage blocks can be federated into larger flash pools to scale performance linearly. High availability, snapshots and disaster recovery are built in, assuring maximum reliability.

Revolutionary TCO

The Cloudistics platform is a fraction of the cost of the public cloud or hyperconverged solutions. Designed from the ground up to fit even the tightest budget, along with an integrated cloud controller that's completely free, Cloudistics is the obvious choice for cost-sensitive deployments. But don't let the low price fool you: Cloudistics exceeds the performance of other solutions costing much more.

Next Steps

At the end of a book, you might be looking for something witty and incredibly insightful. But, I realize that your plane ride may be coming to an end, or you might be ready to head home from work, so I'm going to leave you with a very simply homework assignment: head to www.cloudistics.com and experience the platform for yourself! Take a test drive in Cloudistics' virtual lab, register for a live demo, or even get a test unit delivered right to your office so you can put the platform through its paces and see how well it can support even your most intense workloads.

About

About Cloudistics

Launched in 2016, Cloudistics helps anyone bring the power of the cloud to the data center in an easy-to-use, on-premises cloud platform that automatically provides high-performance resources for all types of applications: Docker, Splunk, Hadoop, Citrix® VDI, and many other high-performance workloads. With no onsite controllers to install or maintain, it's easy to scale across a large site or multiple locations—all from a single, centralized dashboard.

Everything is included, all software and all hardware to run your cloud—just download your applications and go.

Our software-driven platform offers with the same ease of use, elasticity, and pay-as-you-grow economics as the public cloud while addressing enterprise requirements of predictable performance, predictable costs, and security and control of applications and data.

To learn more about Cloudistics, visit http://www.cloudistics.com

About the Author

Scott D. Lowe, vExpert

Scott Lowe is a vExpert and a partner in ActualTech Media. Scott has been in the IT field for over twenty years and spent ten of those years filling the CIO role for various organizations. Between books, articles, and blog posts, Scott has written thousands of pieces of content and, today, along with the motley crew at ActualTech Media, focuses on educating IT buyers and connecting enterprise IT companies with potential customers. You can learn more about ActualTech Media by visiting www.actualtechmedia.com.