3 Reasons to beware the hype around hyperconverged

Tintri

3 REASONS TO BEWARE THE HYPE AROUND HYPERCONVERGED 170922T10367

Table of Contents

HCI and the Enterprise

It seems as if every few years there's a new infrastructure approach that promises to revolutionize the enterprise data center. You probably know from experience that many of these trends don't live up to the initial hype and some even end up taking your operations in the wrong direction. So it is with conventional hyperconverged infrastructure (HCI) today.

Conventional HCI combines storage, compute, and virtualization on each node—and placing storage on top of the hypervisor—in an approach that promises simplicity and lower cost. While this may make sense for small deployments, Tintri strongly believes that most enterprises will be better served with separate servers and virtualization-centric storage when deploying infrastructure at scale.

When Tintri was founded, we evaluated various architectures and decided that a virtualization-centric, federated pool of storage driven by analytics was the best approach in terms of balancing cost, performance, and complexity. This paper examines why Tintri's CONNECT architecture with its web services building blocks is far better suited to address enterprise needs and for creating an enterprise cloud.

Cost

Despite vendor claims to the contrary, conventional HCI can increase deployment costs in a number of ways:

- Requirement for balanced nodes
- Increased software licensing costs
- Increased storage costs

Balanced Nodes

Conventional HCI implementations generally require you to have a similar CPU, memory, and storage configuration on all the nodes in a cluster. While it may be possible in some implementations to have storage-heavy or compute-heavy nodes, they are not considered a best practice because the imbalance can cause storage hot spots and bottlenecks. For example, balanced nodes are still considered a best practice for both Nutanix and VMware vSAN.

As a result, anyone that follows best practices ends up purchasing storage when they need compute or compute when they need storage. You end up spending more—and having valuable resources sitting idle. This effect is so well known that it is often referred to as the "HCI tax."

Storage Costs

Many conventional HCI implementations, including Nutanix, VMware vSAN, and NetApp HCI, store multiple (two or three) copies of each block of data to protect against failures. Naturally this increases the total amount of storage you'll need—and thus your total cost.

Licensing Costs

HCI may also add to your licensing costs. In most implementations, storage on each node is controlled by a dedicated virtual machine. So, the more storage you have, the higher your virtualization licensing costs.

Other software licensing costs can go up as well. For example, Microsoft SQL Server and Oracle database are licensed based on the number of CPUs on a node. It doesn't matter if those CPUs are actually being used for storage, your license costs are still based on total CPUs on the node. Because you're dedicating a lot of resources on each node to storage—you end up needing more nodes to get enough vCPUs for all your database instances.

The Advantage of Best-of-Breed Architecture

Infrastructure that is architected with separate, best-of-breed servers and virtualization-centric storage avoids these cost-related challenges. With storage independent from compute, it's much easier to get the right mix of resources and you have more flexibility to pick the best compute and storage to support your particular workloads. Compute and storage gets better every year in terms of performance and density. Purchasing them separately gives you more flexibility to purchase the latest equipment and mix new and old hardware as needed.

Performance

Performance is a key consideration in almost any IT infrastructure deployment. Enterprise data centers have quickly adopted all-flash storage as a means to deliver the IO performance needed to power applications of all kinds—especially analytics and new mobile and customer-facing applications.

Although hyper-converged infrastructure (HCI) has gotten a lot of attention in industry press, conventional HCI architectures lag behind best-of-breed external storage systems—both in hybrid flash and all-flash configurations—in a number of important performance metrics:

- Latency
- IOPS (especially with all-flash)
- Predictability

Latency

The latency of IO operations on conventional HCI implementations suffers in comparison to external storage systems because of the requirement to store multiple copies of each block of data. All data must be mirrored or copied across the network to one or two other nodes. Some vendors support erasure coding, but it comes with a high performance and latency penalty. Others support post-process erasure coding, but only for cold data.

Mirroring affects write latencies and may affect read latencies as well. In a recent study, ESG compared the performance of several HCI platforms under different conditions. The best latency achieved by any solution was around 5ms, which is far slower than best-of-breed all-flash arrays.

Apart from mirroring and erasure coding, activities like VMware vMotion, HA events, maintenance on nodes and node failures can cause increased latencies for workloads because of their noisy nature and the reduction in total available resources.

IOPS Performance

The IOPS performance that storage can deliver, especially all-flash storage, correlates directly to how much CPU you have. Most standalone all-flash arrays use 28-40 cores per controller for 13-24 SSDs. (Some arrays scale higher, but Tintri believes this negatively impacts the IO density of all-flash and, as a result, performance predictability.)

HCI implementations limit the amount of CPU available for storage. Up to 8 vCPUs or 20% of available CPU are typical limits. This is not enough horsepower to deliver full performance from the flash drives on each node (6-24), resulting in a lot of wasted flash IOPS. Enabling data reduction on HCI platforms consumes even more CPU, making the situation that much worse. That's why data reduction is optional on many HCI implementations.

Increasing the amount of CPU dedicated to storage, if possible, will end up having a big impact on licensing costs as discussed above. You don't want to be stuck paying for expensive hypervisor, SQL Server, and/or Oracle licenses on CPUs dedicated to storage functions.

Predictability

For many applications, delivering performance that's predictable is just as important as low latency or raw IOPS. With HCI, storage and compute activities run on the same nodes. In most cases, storage software runs inside a VM as a virtual appliance at the guest layer. Each IO operation flows through the hypervisor's CPU scheduler four times: twice for the IO, twice for IO acknowledgment.

This is less of a problem when system utilization is low, but—because the CPU resources are shared—it becomes a major bottleneck when utilization becomes moderate or heavy. You could apply CPU reservations, but per-VM reservations introduce a new set of challenges all the way up to impacting cluster-wide HA-failover policies. And CPU reservations don't guarantee the virtual appliance will have instant access to the CPU. If another vCPU is scheduled, it is allowed to finish its operation, causing IO delays within the virtual appliance.

The result is less predictable latency with unexpected spikes when a node or cluster becomes busy. Applications may see latency that varies widely from one IO to the next, which can be a disaster for those that are latency-sensitive. This is exacerbated in an enterprise cloud environment where an organization is managing thousands of virtual machines and/or containers. Despite claims of delivering "web-scale," conventional HCI is rarely able to meet performance expectations at scale.

Caveat Emptor

HCI promises significant benefits, but as with any major infrastructure decision, let the buyer beware. Enterprise IT teams want—and in many cases, need—to use all the functionality infrastructure can deliver.

With HCI, enabling new functionality can increase resource utilization beyond acceptable levels. As new features like snapshots, replication, deduplication, compression, and so on are enabled, you either add more hardware or it impacts the predictability and performance of your infrastructure. Making these trade-offs can become an almost daily fact of life for HCI admins.

A Best-of-Breed Architecture Offers Better Performance—Especially with All-Flash

The main reasons for choosing all-flash storage are:

- Dramatic reductions in the latency of each IO operation
- Big increases in total IOPS
- More predictable performance for every IO

External storage systems do a much better job delivering all three of these than HCI.

Tintri's CONNECT architecture goes further—we automatically assign every virtual machine and container to its own lane to its own lane to eliminate any conflict over resources. That makes it simple to set minimum and maximum quality of service on individual virtual machines, and guarantee application performance. If performance is a primary consideration, you'll want to evaluate all options carefully and think twice before choosing HCI.

Complexity & Risk

One of the most often-claimed advantages of conventional HCI is that it decreases complexity. The "simplicity" of HCI architectures comes at a price:

- Increased troubleshooting complexity
- Greater operational risk

Troubleshooting Difficulties

The tightly coupled architecture of HCI makes it more difficult to troubleshoot performance issues. Because everything is layered together on each node, it becomes almost impossible to isolate the source of a performance bottleneck.

If increasing a VM's memory and CPU resources doesn't solve the problem, you then have to assume the problem is IO.

- Where is the IO bottleneck? Is it in the host, network, or storage?
- Are too many data services (deduplication, erasure coding, replication, etc.) increasing metadata and affecting performance?
- Since guest VMs and the storage VM share the same resources, how do you isolate the problem?
- Does the problem result from IO to internal storage or storage on another node?
- If it's internal storage, can you throttle or migrate workloads? Will migrating workloads fix or increase the performance problems?
- If it's storage on another node, is it a network bottleneck or is it the other node? Are multiple external nodes involved? In some cases, data for a single VM could be spread across many nodes.

As you can see, the process gets complicated quickly and that complexity grows with the size of your HCI cluster. Often, the only solution to the above scenario is to add another node.

Higher Risks

Virtualization helped solve many traditional infrastructure issues such as hardware maintenance and patching. With external storage, you can easily move VMs to another host by moving the compute and memory state using VMotion or Hyper-V live migration. With conventional HCI architectures, storage is more tightly coupled with compute so there's a lot more to think about:

- **Data evacuation before maintenance.** Although most vendors allow maintenance without data evacuation, it is not a best practice because it introduces risk. When you evacuate, you're spreading the entire load from the node—both compute and storage—across other nodes, increasing the potential for bottlenecks and noisy neighbor problems.
- **Reduction in amount of storage.** When a node goes offline for maintenance, a big chunk of your storage goes offline too, potentially leaving your cluster constrained.
- **Reduction in amount of available flash.** Especially in hybrid configurations, when a node goes offline that also means a big chunk of flash goes offline. Flash is highly important as a cache, so flash misses go up and performance goes down.

All this means that your IT team needs to be extremely careful about taking maintenance windows, and, in many cases, you lose the independence to do maintenance activities because of its broader impact. Doing maintenance becomes risky, but you know the risks of not doing patching and other maintenance all too well.

The HCI Snowball Effect

A single HCI failure can trigger much larger problems. When a host fails for any reason, it has the following effects:

- Reduces available resources for compute and storage.
- Reduces available flash in hybrid configurations, resulting in a double dip. Flash for VMs from the failed node must be rewarmed. The extra pressure causes data from existing VMs to be evicted. Thus, VMs from both the failed and surviving nodes suffer.
- The process repeats on the reintroduction of the failed node.

Failure of even a single component, such as a flash drive, can cause an entire node to collapse. The result is a far greater impact on operations than when storage is decoupled from the host.

A Best-of-Breed Architecture Has Lower Risk

Conventional HCI destroys the stateless nature of virtualization and increases risk. Performance problems are much easier to troubleshoot with a decoupled architecture, especially when the architecture has been built from the ground up to provide workload-granular analytics. For example, Tintri allows you to see the root cause of any latency issue across compute, network, and storage, giving you a comprehensive view of your infrastructure at the VM or container level.

Because storage and compute are physically and logically separated, none of the risks described above affect the Tintri enterprise cloud platform, making it a lower-risk option for large-scale enterprise infrastructure deployments.

Which Choice is Right for You?

Conventional HCI promises many benefits, but it struggles to deliver those benefits at scale because of the inherent challenges associated with its design. It simply hasn't proven to be a great design for enterprises running mixed workloads at scale. Many current Tintri customers learned this lesson the hard way before coming to us.

At Tintri, we believe that to build a true enterprise cloud, you must be able to control costs, deliver predictable performance, and minimize risks by deploying separate best-of-breed servers and storage. Tintri offers all-flash storage arrays, cloud management software, and web services. Together, these building blocks deliver virtualization-centric operations with guaranteed performance, in-depth analytics, and a federated scale-out architecture. Our web services approach to infrastructure simplifies your data center and makes autonomous operations a reality.