Technical Validation

Mission-critical Workload Performance Testing of Different Hyperconverged Approaches on the Cisco Unified Computing System Platform (UCS)

By Tony Palmer and Kerry Dolan, Senior Validation Analysts

February 2019

This ESG Lab Report was commissioned by Cisco and is distributed under license from ESG.

Contents

ESG Validation Reports

The goal of ESG Validation reports is to educate IT professionals about information technology solutions for companies of all types and sizes. ESG Validation reports are not meant to replace the evaluation process that should be conducted before making purchasing decisions, but rather to provide insight into these emerging technologies. Our objectives are to explore some of the more valuable features and functions of IT solutions, show how they can be used to solve real customer problems, and identify any areas needing improvement. The ESG Validation Team's expert third-party perspective is based on our own hands-on testing as well as on interviews with customers who use these products in production environments.

Introduction

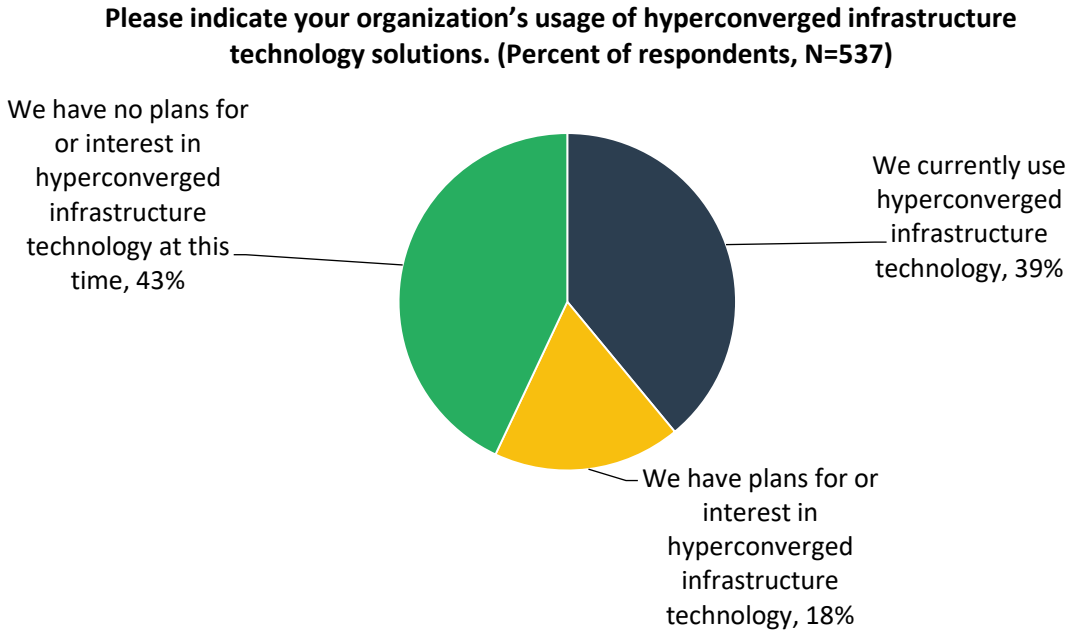
This report documents an ESG Lab audit and validation of Cisco HyperFlex hyperconverged infrastructure (HCI) performance testing, which focused on comparisons of Cisco HyperFlex fully engineered all-flash solutions on Cisco UCS against two software-only HCI offerings from leading vendors independently validated to run on Cisco UCS hardware servicing mission-critical workloads.

Background

Organizations today must be extremely agile and flexible in their ability to add applications and virtual machines (VMs) to mission-critical production environments quickly to handle the speed of business. This level of agility is extremely difficult to achieve with silos of compute, network, and storage gear that are static and require individual management. This is one reason for the popularity of hyperconverged infrastructures (HCI). HCI offers a single, centrally managed solution with software-defined compute, network, and storage that is flexible, scalable, and easy to deploy.

Adoption of HCI has grown significantly since coming to market and ESG research continues to confirm the popularity of HCI: in an ESG research study, 57% of respondents reported that they currently use or plan to use HCI solutions.[1] This is not surprising given the factors driving them to consider HCI. Deployment drivers most cited by respondents include improved scalability (31%), total cost of ownership (28%), ease of deployment (26%), and simplified systems management (24%).

Figure 1. Organizations' Usage of Hyperconverged Infrastructure Solutions



Source: Enterprise Strategy Group

Powering Tier-1 workloads on HCI

Organizations looking to move mission-critical workloads traditionally reserved for three-tiered architecture or converged infrastructure (CI) solutions to HCI need to give careful consideration to the solution they choose. Powering complex workloads can expose architectural deficiencies in an HCI solution not optimized to handle the workload requirements. An HCI platform deployed to support a tier-1 workload needs to not only provide high IOPs and low read/write latency, it

[1] Source: ESG Master Survey Results, *Converged and Hyperconverged Infrastructure Trends*, October 2017.

needs to do so in a consistent, predictable manner. Predictable performance and low VM performance variability are critical to maximize end-user productivity across an organization.

Key Metrics to Consider when Evaluating HCI Solutions

Simplicity is no longer the only priority; as more HCI solutions have come to market, the key buying criteria have expanded to include performance, but many solutions still cannot deliver the consistent high performance that mission-critical workloads demand. While first generation HCI architecture ran on x86 servers connected through commodity grade switches, the mission-critical nature of tier-1 workloads has led to software-only HCI companies validating their software on trusted enterprise grade hardware like Cisco UCS.

Input/output operations per second (IOPS)—Adoption of flash-based storage has greatly reduced I/O challenges in traditional shared-storage environments, but in a clustered environment like HCI, total IOPS can vary greatly depending on the network connection between nodes as well as the software layer powering the HCI solution. For HCI deployments, it's important to evaluate both the total number of IOPS delivered by the cluster as well as the IOPS consistency that is delivered. Consistent VM performance has been a challenge since the beginning of virtualized computing, but "noisy neighbor" VM performance can be even more pronounced with HCI deployments based on how the software layer writes data across the cluster.

Latency—While IOPS are an important performance indicator, latency as it relates to the application should also be considered when purchasing an HCI solution. Clustered environments like HCI can have multiple bottlenecks like storage performance, software responsiveness, and network throughput, all of which can contribute to application latency. Increased latency means decreased responsiveness of applications for users.

- Read latency—The time required for the storage controller to find and deliver the proper data blocks. For flash storage as evaluated in this paper, this includes the time for the flash subsystem to find the required data blocks and prepare to transfer them, and the transit time through the network.
- Write latency—This is the time it takes for the storage controller to perform all the activities required to write data blocks, including determination of the proper location for the data, performance of overhead activities—block erase, copy, and "garbage collection," then writing and acknowledging the write back to the host.
- Total latency—Total latency is simply a combination of the read and write latencies calculated using the ratio of reads and writes used by the application. For example, for a workload that consists of 70% reads and 30% writes, the total latency is the average of the read and write results, weighted according to the percentage of each.

Industry Approaches to HCI—Software Validated versus Fully Engineered

HCI was conceived as the next step in the evolution of the modular data center concept. The goal was to simplify rack-level converged infrastructure (CI) to node-level deployments. Rather than three tiers of infrastructure managed through a common software platform, HCI combines virtualized compute and software-defined storage integrated through the software layer and deployed on a single chassis to create a node. Nodes are connected through network switches to form a shared pool of resources that can be scaled on demand by adding a new node to the cluster. There are distinct approaches that vendors have taken to bring HCI solutions to market that should be considered.

HCI Distribution Models:

Software-only HCI: This model focuses on the software layer used to integrate compute and storage into a single node. Users purchase the HCI software, which can then either be installed in-house or by a third party on industry-standard servers. Initial HCI deployments tended to support tier-2 or even tier-3 workloads, so it was common for the software to be

deployed on off-the-shelf servers and connected through commodity switches to keep costs low. As HCI has matured and more critical workloads are being deployed, organizations have begun to demand that HCI run on trusted hardware platforms. It is important to note that not all hardware manufacturers carry the same validations, so it would be wise for potential users to read the fine print for these types of deployments. Deploying software on a hardware platform that does not carry validation from all parties can open the door to finger pointing and add a degree of risk that might be more than some organizations are willing to accept.

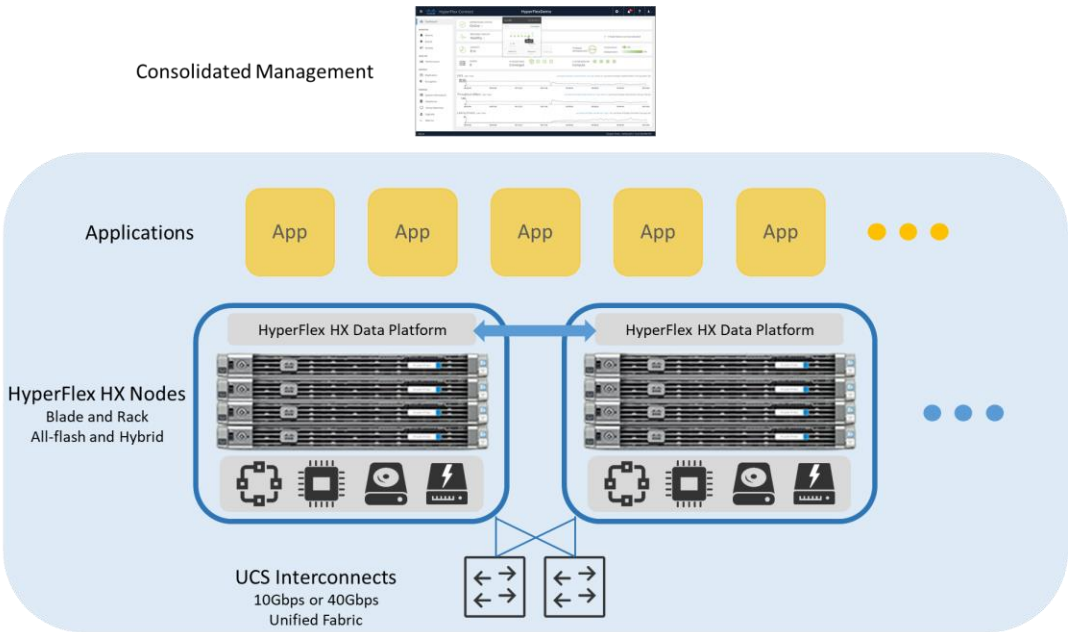
Fully engineered HCI: This model is designed to provide the simplest way to deploy an HCI solution. Predominantly sold by tier-1 vendors, users selecting this approach get appliances built on trusted hardware platforms that are shipped with the software preinstalled. Vendors choose this approach because it enables them to engineer and optimize every component of HCI across computing, networking, and storage into a simple appliance offering. Fully engineered HCI solutions remove a level of risk by ensuring users are delivered factory-validated configurations whose operation and support are guaranteed by a single source for computing, networking, and storage software. It is important to note that some appliances are created through partnerships between hardware and software vendors and the level of optimization between the hardware and software layers can vary and affect overall performance.

Cisco's Fully Engineered HCI Approach

Cisco HyperFlex is a fully engineered hyperconverged system that combines compute and software-defined storage, as well as fully integrated networking optimized for the east-west traffic flow of an HCI platform. This fully integrated platform is designed to scale resources independently and deliver consistent high performance. Cisco HyperFlex is engineered on Cisco UCS, combining the benefits of the UCS platform (such as policy-based automation for servers and networking) with those of the HX Data Platform's distributed file system for hyperconvergence.

It supports edge to edge workloads from mission-critical core data center applications to remote locations. The latest HX 3.0 update adds support for Microsoft Hyper-V in addition to VMware ESXi along with support for multi-cloud and containerized environments. HyperFlex deployments require a minimum three-node cluster for high availability, with data replicated across at least two nodes, and a third node to protect against single-node failure.

Figure 2. Cisco HyperFlex Hyperconverged Infrastructure



Source: Enterprise Strategy Group

HyperFlex HX-Series Nodes are engineered on the Cisco UCS platform and powered by the latest generation of Intel Xeon processors, and comprise:

- Cisco HyperFlex HX Data Platform. The core of any HCI solution is the software platform, and the HX Data Platform was engineered specifically for HCI software-defined storage. Operating as a controller on each node, the HX Data Platform is a high-performance, distributed file system that combines all SSD and HDD capacity across the cluster into a distributed, multi-tier, object-based data store, striping data evenly across the cluster. It also delivers enterprise data services such as snapshots, thin provisioning, and instant clones. Policy-based data replication across the cluster ensures high availability. Dynamic data placement in memory, cache, and capacity tiers optimize application performance, while inline, always-on deduplication and compression optimize capacity.
  - The HX Data Platform handles all read and write requests for volumes accessed by the hypervisor. By striping data evenly across the cluster, network and storage hotspots are avoided, and VMs enjoy optimal I/O performance regardless of location. Writes go to local SSD cache and are replicated to remote SSD in parallel before the write is acknowledged. Reads are from local SSD if possible or retrieved from remote SSD.
  - The log-structured file system is a distributed object store that uses a configurable SSD cache to speed reads and writes, with capacity in HDD (hybrid) or SSD (all-flash) persistent tiers. When data is de-staged to persistent tiers, a single sequential operation writes data to enhance performance. Inline deduplication and compression occur when data is de-staged; data is moved after the write is acknowledged so there is no performance impact.
- Cisco UCS compute-only nodes. Both UCS blade and rack servers can be combined in the cluster, with a single network hop between any two nodes for maximum east-west bandwidth and low latency. HyperFlex lets you alter the ratio of CPU-intensive blades—compute nodes—to storage-intensive capacity nodes—HX nodes—so users can optimize the system as application needs shift. All-flash and hybrid nodes are available.
- Cisco Unified Fabric—UCS 6200/6300 Fabric Interconnects enable software-defined networking. High bandwidth, low latency, and 40Gbps and 10Gbps connectivity in the fabric enable high availability as data is securely distributed and replicated across the cluster. The network enables HX clusters to scale easily and securely. The single hop architecture is designed to maximize the efficiency of the storage software to enhance overall cluster performance.
- Cisco Application Centric Infrastructure (ACI) for automated provisioning. ACI enables automation of network deployment, application services, security polices, and workload placement per defined service profiles. This provides faster, more accurate, more secure, lower cost deployments. ACI automatically routes traffic to optimize performance and resource utilization and re-routes traffic around hotspots for optimal performance.
- Choice of-industry leading hypervisors including VMware ESXi and vCenter as well as Microsoft Hyper-V. The hypervisor and management application come pre-installed, providing a familiar management interface for all hardware and software.

Cisco HyperFlex delivers numerous benefits, including:

- High performance. In addition to performance features mentioned above, HyperFlex Dynamic Data Distribution securely and evenly distributes data across all cluster nodes to reduce bottlenecks.
- Fast, easy deployment. This pre-integrated cluster can be deployed just by plugging into the network and applying power. Node configuration and connection is handled through Cisco UCS service profiles. Cisco says that customers report typical deployment times of less than one hour.

- Consolidated management. Systems are monitored and managed through Cisco HyperFlex Connect or Cisco Intersight, which eliminates separate management silos for compute and storage. HyperFlex Connect lets organizations manage and monitor clusters from anywhere and at any time with metrics and trends to support the entire management lifecycle. Intersight is an optional cloud-based platform that allows users to manage all their Cisco HyperFlex and Cisco Unified Computing System (Cisco UCS) infrastructure including traditional, hyperconverged, edge, and remote/branch offices through a single cloud-based GUI.

- Independent scaling. Different from other HCI systems, HyperFlex can independently scale compute and storage resources without the need to add full nodes to the cluster. Users can easily incorporate compute-only nodes with bare UCS servers through the Fabric Interconnects to add additional compute to the cluster, or if more storage is needed, add individual drives to each node; data is automatically rebalanced. This provides the right resources for different application needs, instead of scaling in pre-defined node increments that also add additional software licensing costs.

ESG Technical Validation

Testing was conducted using industry-standard tools and methodologies and was focused on comparing the performance of Cisco's fully engineered HCI solution—HyperFlex—with two software-only HCI offerings from leading vendors validated to run on Cisco UCS hardware within their listed hardware compatibility guidelines. The bulk of the testing used HCIBench and HXBench, tools designed to test the performance of HCI clusters running virtual machines. Both tools leverage Oracle's Vdbench tool and automate the end-to-end process that includes deploying test VMs, coordinating workload runs, aggregating test results, and collecting data.

This extensive testing was executed using a stringent methodology including many months of baselining and iterative testing. While it is often easier to generate good performance numbers with a short test, benchmarks were run for long periods of time to observe performance as it would occur in a customer's environment. In addition, tests were run many times, never back-to-back but separated by days and weeks, and the results averaged. These efforts add credibility by reducing the chances that results were influenced by chance circumstances. Also, testing was conducted using data sets large enough to ensure that data did not remain in cache but leveraged the back-end storage across each cluster.[2]

Mission-critical Workload Testing

The test bed included a four-node HyperFlex HX220c version 2.6 cluster. Comparative software-only HCI solutions were running UCS C220 and C240 four-node systems with similar configurations. Configuration details are listed in Table 1.

[2] When evaluating technology solutions, customers would be wise to understand the details behind vendor testing. Timing of test runs, volumes of data, and other details will impact performance results; these results may or may not be relevant to the customer environment.

Table 1. Tested HCI Configurations

Platform	Nodes	Processors/Cores Per Node	RAM Per Node	Cache Per Node	Storage Capacity Per Node	Hypervisor
Cisco HyperFlex – Fully Engineered HCI with Cisco UCS	Four	2x E5-2680, 28 Cores	512GB	800GB Performance	6x 960GB SSD Value	VMware vSphere 6.5
Vendor A Software-only HCI Validated on Cisco UCS	Four	2x E5-2695, 36 Cores	512GB	Note[3]	6x 1.6TB Performance	VMware vSphere 6.5
Vendor B Software-only HCI Validated on Cisco UCS	Four	2x E5-2680, 28 Cores	256GB[4]	800GB Performance	6x 960GB SSD Value	VMware vSphere 6.5

Source: Enterprise Strategy Group

OLTP tests were run with four VMs and a 3.2TB working set, while the mixed workload test used 140 VMs (35 VMs per node), each with 4 vCPUs, 4 GB RAM, one 40GB disk, and running RHEL version 7.2. The working set size was 5.6 TB. Tests were run for a minimum of one hour and up to five hours, with a five-minute ramp-up before each test and a minimum one-hour cool-down between tests. Before every test was run, each VM was primed with written data by the test tool. This ensures that the test is reading "real" data and writing over existing blocks and not simply returning null or zero values directly from memory. This happens when data is not primed so it is an important step to ensure that the test accurately reflects how data is read and written in an application environment. Priming of this large working set can take many hours to complete but is a wise investment in time to get more accurate performance results.

Testing was performed using I/O profiles designed to emulate complex, mission-critical workloads, including OLTP using Oracle and SQL Server backends, as well as virtual application server and desktop activity. Block sizes were assigned according to the applications being emulated, with 100% random data access. VMs by nature generate random I/O by combining I/O from multiple applications and workloads. It is important to note that all tests were run with compression and deduplication active on the Cisco HX cluster. Alternative vendor's solutions offer the ability to disable compression and deduplication, so tests were run in both modes for those systems.

Aggregate Testing IOPS from the Vdbench Tool

The Vdbench tool uses a specific methodology to derive an aggregated IOPS result during benchmark testing. Aggregate testing IOPS are calculated by taking the average IOPS delivered to test virtual machines (VMs) at various workload levels—12 curves ranging from 20% to 100% loads. The average IOPS of each test VM are then aggregated to derive the aggregated testing IOPS in each test—for example, aggregated IOPS from four test VMs and each of their 12 load curves.

Note: Aggregate testing IOPS cannot be used to size workloads for specific applications.

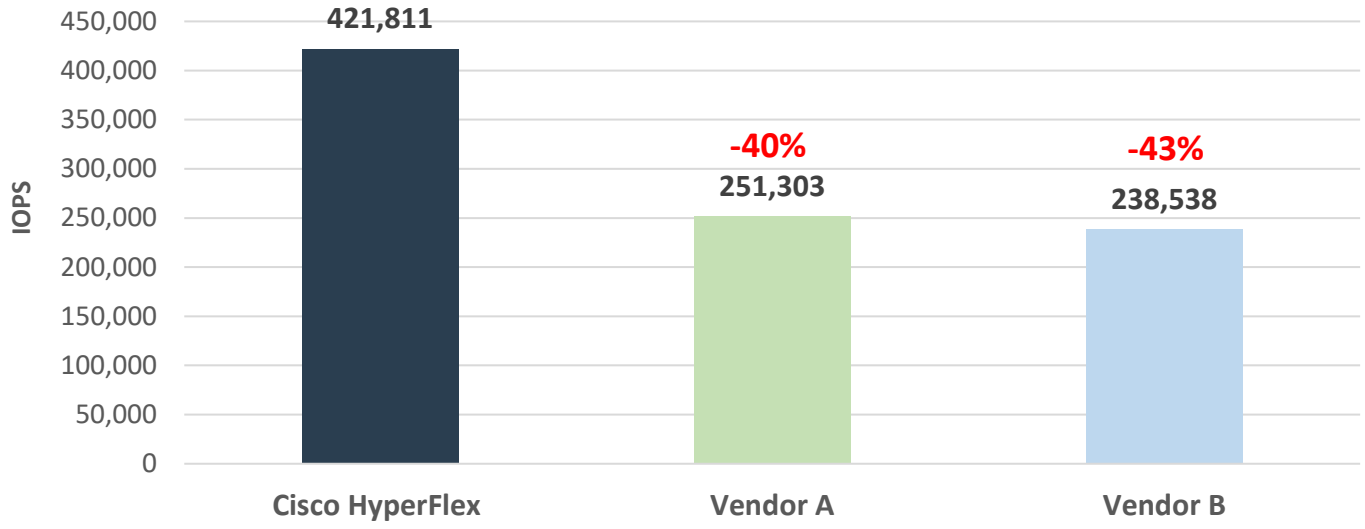
ESG Testing

First, ESG Lab looked at an OLTP workload designed to emulate an Oracle environment.[5] Vdbench was used to create a workload that exercised different transfer sizes and read/write ratios. In the Vdbench profile, the deduplication ratio was set to 3 with a unit size of 4 KB and the compressibility ratio also set to 3. The test was run with four virtual machines.

[3] Note: Vendor validated configuration requires all Enterprise Performance SSDs only, no cache.
[4] The amount of CPU resources and memory available had no measurable impact on performance across vendors. CPU and memory resource utilization on each node for all vendors was far below available capacity.
[5] A publicly available Vdbench profile was used to simulate the I/O and data patterns produced by Oracle and these results should not be interpreted as Oracle application measurements.

Over the course of the four-hour test, HyperFlex was able to aggregate more than 420,000 testing IOPS in Vdbench with a total response time of just 447 µsec, as seen in Figure 3. Software-only HCI vendors A and B were able to support just 238,000 and 251,000 testing IOPS respectively.

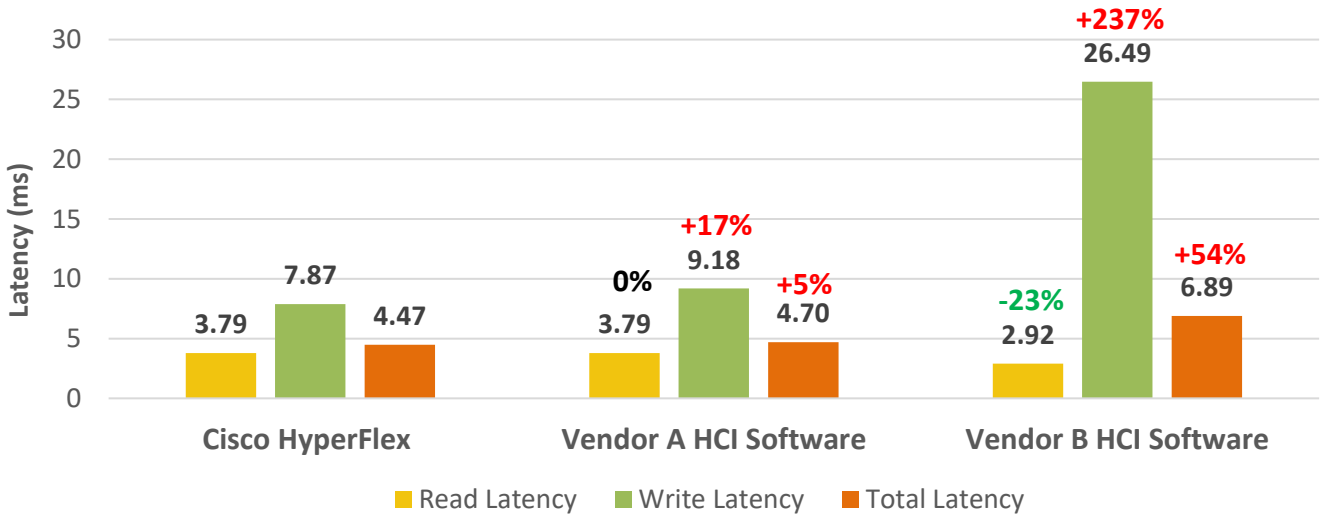
Figure 3. Oracle OLTP Workload—Aggregate Testing IOPS



Source: Enterprise Strategy Group

Response times were reasonably comparable across systems, with the notable exception of write latency on Vendor B, which averaged 26.49ms. Compression and deduplication was active on all systems.

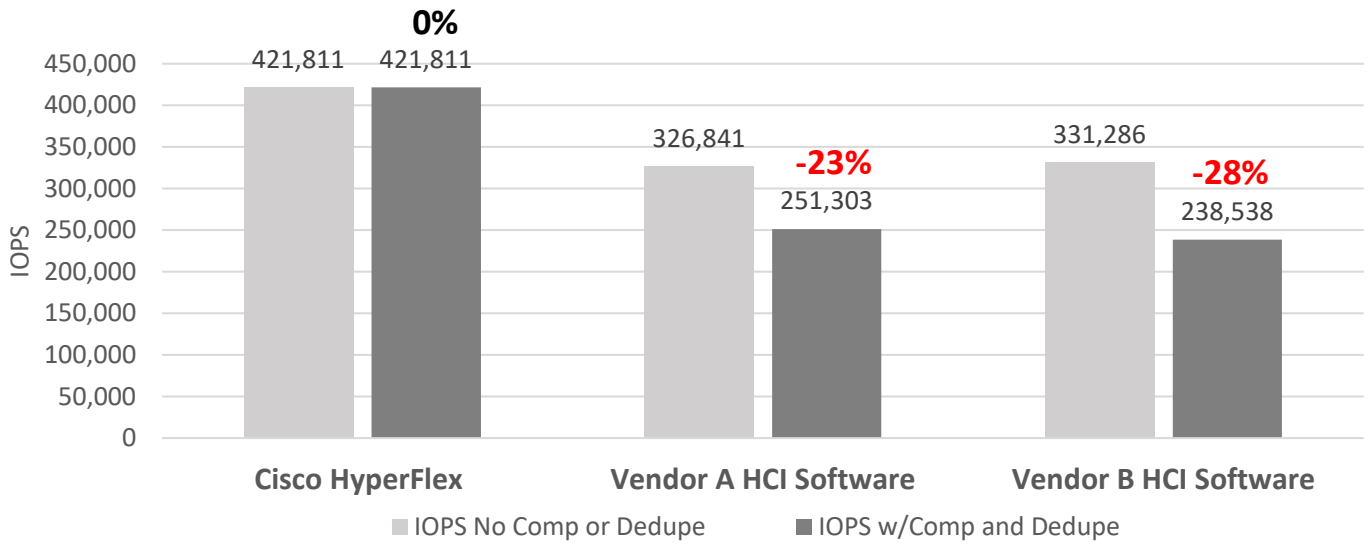
Figure 4. Oracle OLTP Workload—Response Time



Source: Enterprise Strategy Group

ESG Lab also examined the same workload on the two alternative systems with deduplication and compression disabled, to determine the potential impact of these technologies running an Oracle workload.

Figure 5. Impact of Compression and Deduplication-Oracle OLTP

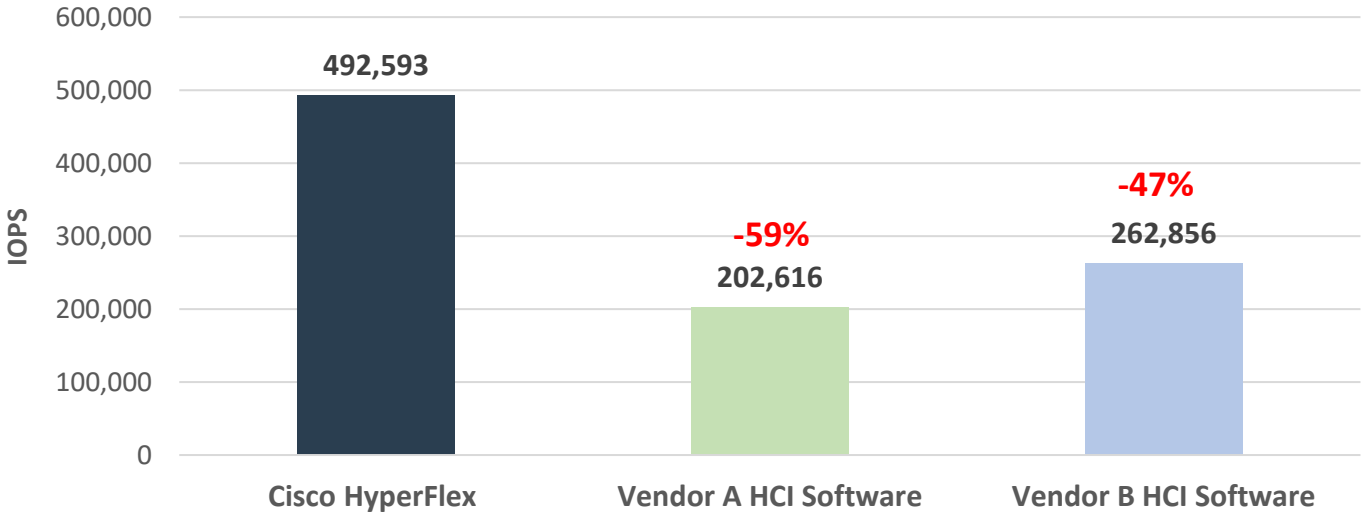


Source: Enterprise Strategy Group

As seen in Figure 5, for software-only HCI vendors, compression and deduplication reduced performance by up to 28%. Compression and deduplication are always on and inline for Cisco HyperFlex, so both results are with compression and deduplication enabled.

Next, we looked at an OLTP workload designed to emulate a Microsoft SQL Server environment.[6] There are subtle but potentially significant differences that warranted testing against both Oracle and SQL workloads. Vdbench was used to create a workload that exercised different transfer sizes and read/write ratios. In the Vdbench profile, the deduplication ratio was set to 2 with a unit size of 4 KB and the compressibility ratio also set to 2. Again, the test was run with four virtual machines.

Figure 6. SQL Server OLTP Workload—Aggregate Testing IOPS

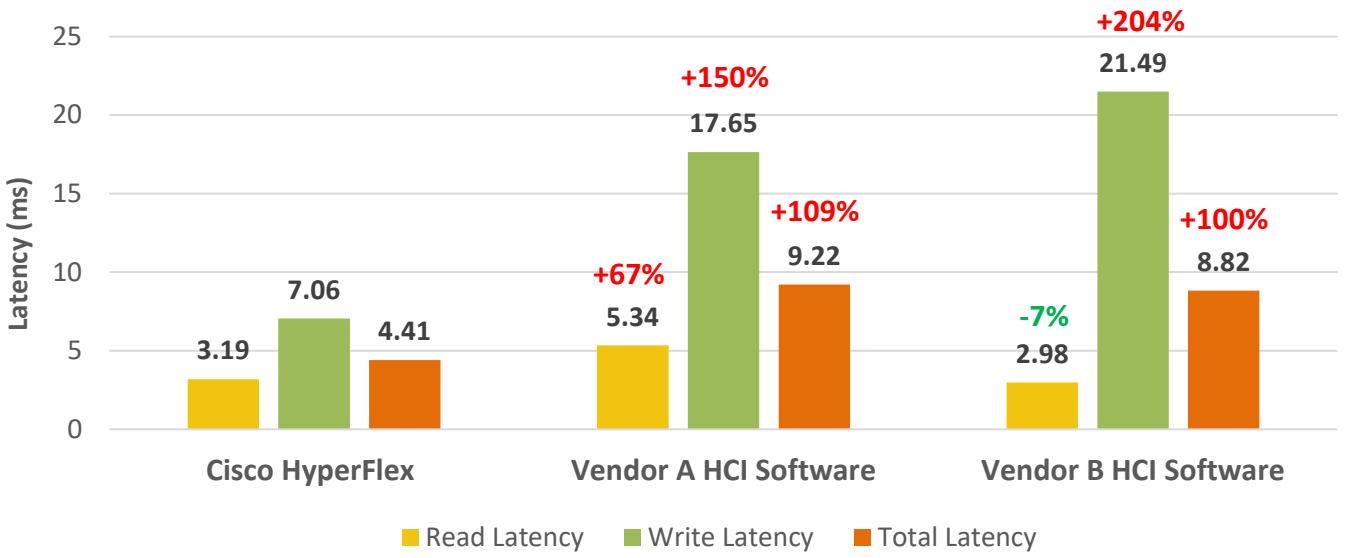


Source: Enterprise Strategy Group

[6] A publicly available Vdbench profile was used to simulate the I/O and data patterns produced by SQL Server and these results should not be interpreted as SQL application measurements.

As Figure 6 shows, the Cisco HyperFlex cluster more than doubled the testing IOPS of software-only HCI Vendor A and nearly doubled the IOPS of Vendor B.

Figure 7. SQL Server OLTP Workload—Response Time

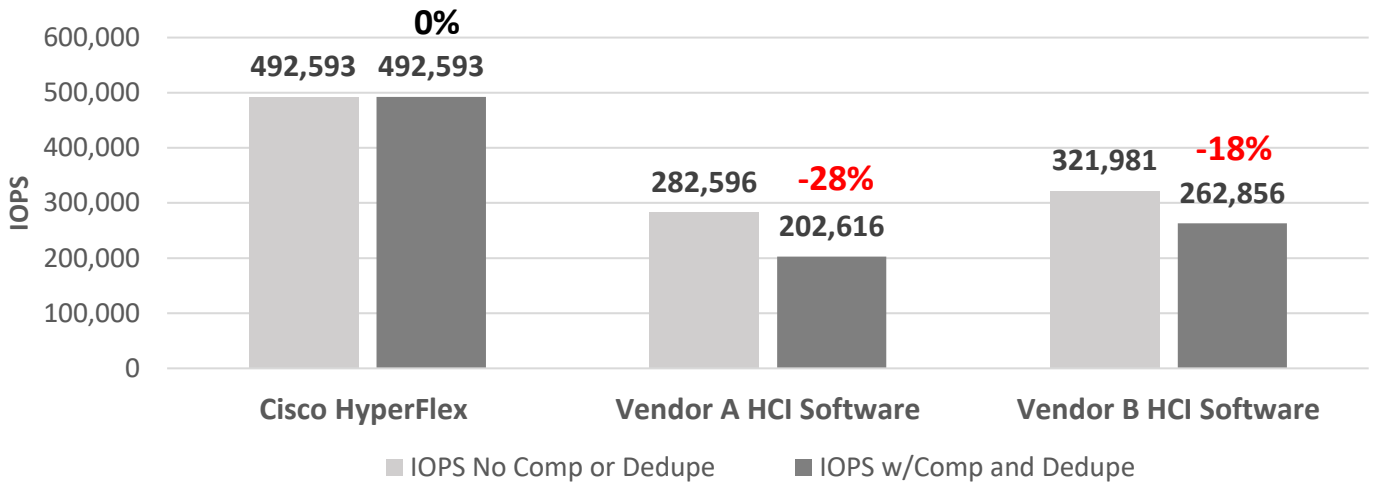


Source: Enterprise Strategy Group

Cisco HyperFlex posted an average response time of 4.41ms. By way of comparison, software-only HCI Vendor A's average response time was 9.22ms and Vendor B's was 8.82ms. This time, both Vendor A and Vendor B posted very high write latency for all-flash systems, averaging 17.65 and 21.49ms respectively.

Again, we examined the same workload on the two alternative systems with deduplication and compression disabled, to determine the potential impact of these technologies running a SQL Server workload. As seen in Figure 8, compression and deduplication again reduced performance by up to 28%. Compression and deduplication are always on and inline for Cisco HyperFlex, so both results are with compression and deduplication enabled.

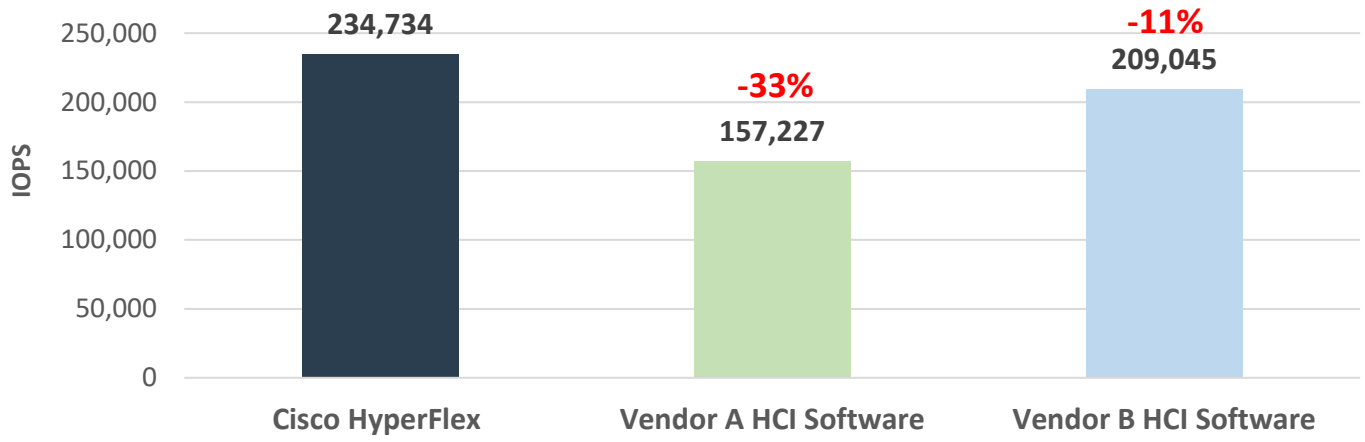
Figure 8. Impact of Compression and Deduplication-SQL Server OLTP



Source: Enterprise Strategy Group

Next, we looked at a mixed workload designed to emulate a virtualized environment with multiple VMs running different applications. Vdbench was used to create a workload that exercised transfer sizes from 4 KB to 64 KB. We ran two sets of tests, with a read/write ratio of 70/30 and with a read/write ratio of 50/50. These tests were run using HCIBench against 140 VMs in each cluster—35 per node, emulating a mixed workload environment with many virtual machines running a variety of applications. In the Vdbench profile, the deduplication ratio was set to 2 with a unit size of 4 KB and the compressibility ratio also set to 2.

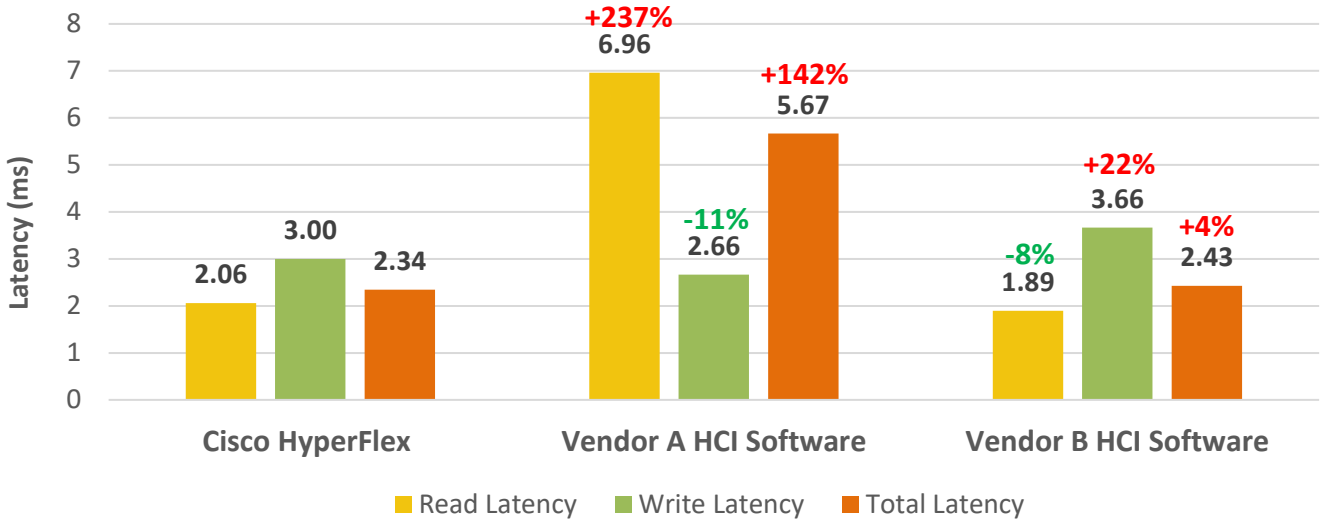
Figure 9. 70/30 Mixed Workload—Aggregate Testing IOPS



Source: Enterprise Strategy Group

As Figure 9 shows, the Cisco HyperFlex cluster sustained more aggregate testing IOPS over the five-hour test than software-only HCI Vendor A or Vendor B.

Figure 10. 70/30 Mixed Workload—Response Time



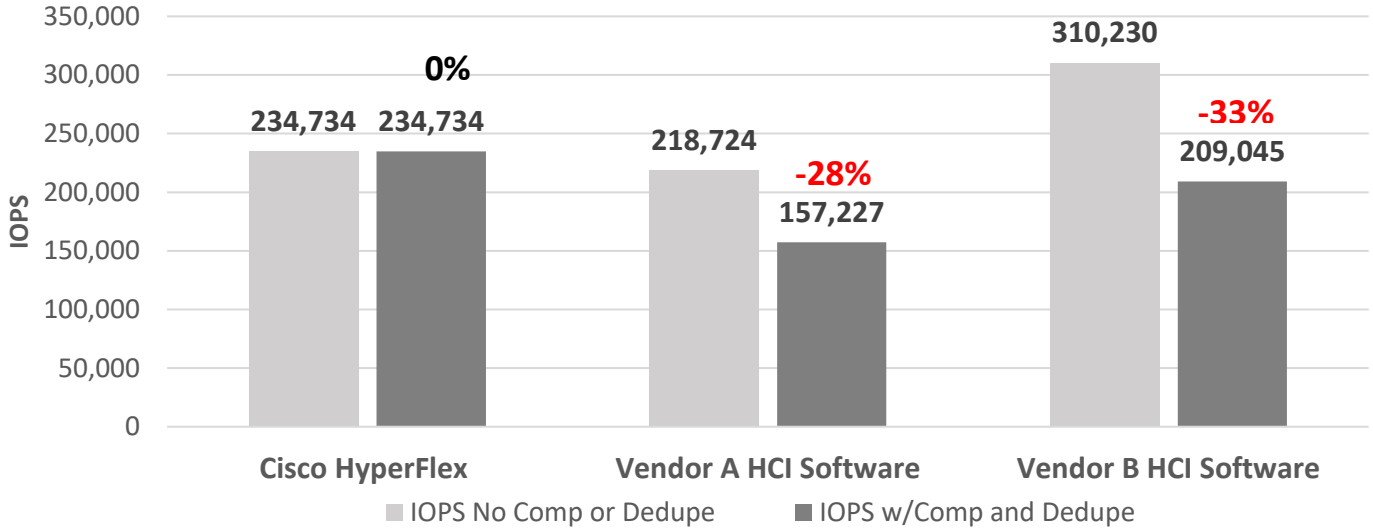
Source: Enterprise Strategy Group

Cisco HyperFlex posted an average response time of 2.34ms. By way of comparison, software-only HCI Vendor A's average response time was 5.67ms and Vendor B's was 2.43ms.

Again, we examined the same workload on the two alternative systems with deduplication and compression disabled, to determine the potential impact of these technologies running a mixed workload. As seen in Figure 11, compression and

deduplication again reduced performance by up to 33%. Compression and deduplication are always on and inline for Cisco HyperFlex, so both results are with compression and deduplication enabled.

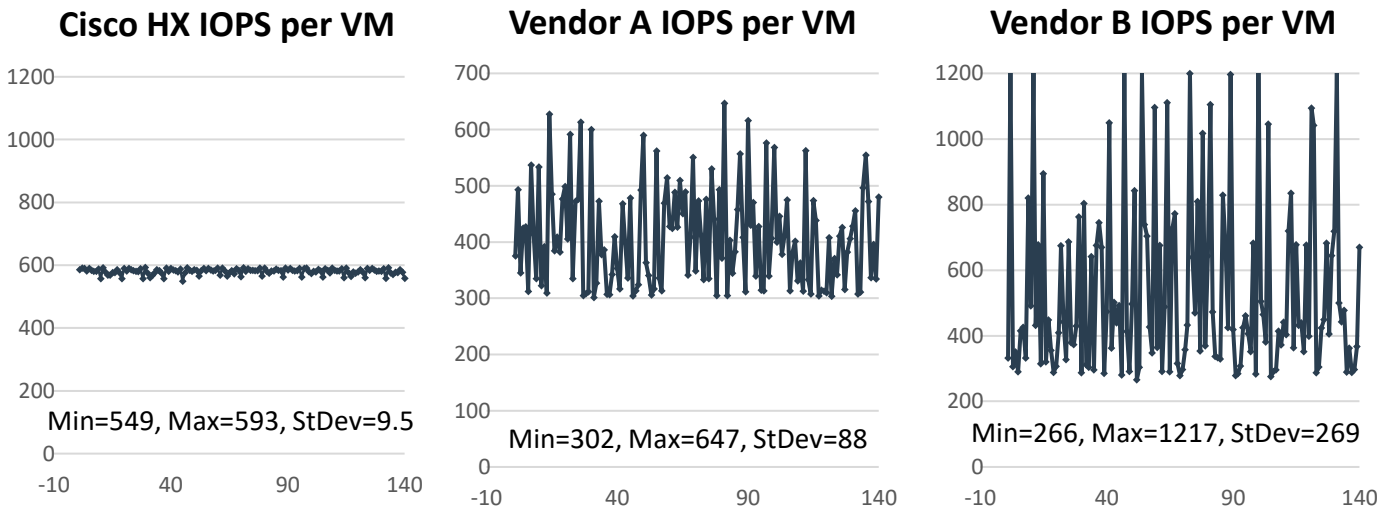
Figure 11. Impact of Compression and Deduplication-Mixed Workload 70/30



Source: Enterprise Strategy Group

An interesting observation was made during mixed workload testing. Software-only HCI vendors A and B both showed considerable variability in performance from VM to VM. While Cisco HyperFlex showed little variation across all 140 VMs—aggregate testing IOPS stayed very close to the target of 600—Vendor A testing IOPS (see Figure 12) varied wildly, from a low of 302 to a high of 647 IOPS, while Vendor B showed even more variability, swinging between 266 and 1,207 IOPS. We saw the same levels of variability in the 50/50 tests.

Figure 12. Mixed Workload, 70% Read, 100% Random—140 Virtual Machines



Source: Enterprise Strategy Group

It's important to note that this variability was observed in every iteration of testing and that no form of storage QoS was used during these test runs on any of the clusters. Network QoS was used for all systems.

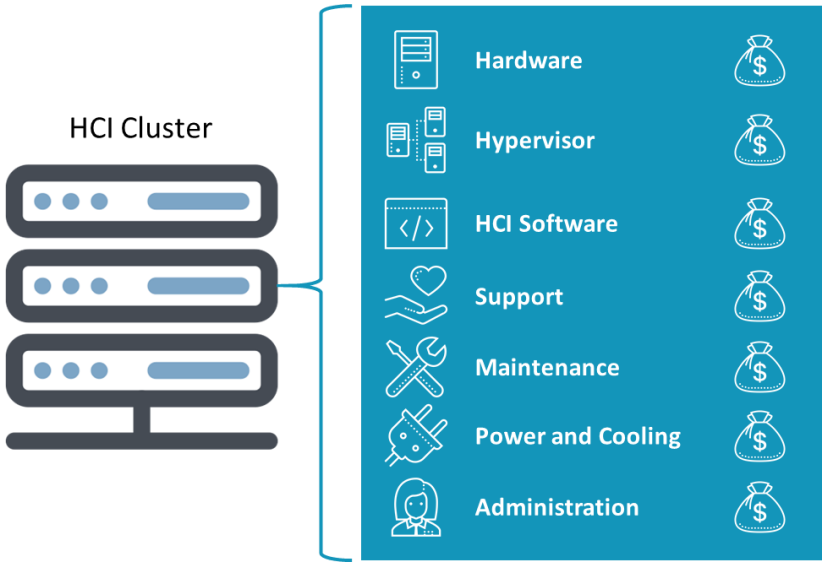
Inconsistency like this could be quite problematic for administrators, who would likely need to use some form of QoS (if available from the HCI vendor) to attempt to control the VMs that are consuming more than their share of resources so others are not starved.

These results prompted a revisit of the Oracle and SQL tests, using larger numbers of virtual machines. As the number of VMs running the SQL and Oracle workloads were increased, first to eight, and then to 16—keeping the number of threads and the working set the same size as in the original tests—performance became unpredictable for both Vendor A and Vendor B with wide variability across VMs, while Hyperflex maintained the same levels of performance and consistency across VMs that we saw in the mixed workload tests.

Differences in HCI Performance Directly Impact Solution Costs

Achieving high levels of performance is important to meeting the requirements of organizations looking to adopt HCI technologies and it is essential to achieve that performance cost-effectively. While the performance delivered by an HCI solution affects application responsiveness and end-user experience, it also factors greatly into the overall cost of the solution being deployed. HCI's node-based architecture means that it can easily scale by adding additional nodes to meet performance needs, but each node carries an upfront capital expenditure (CapEx) burden based on the cost of the hardware platform, HCI software and hypervisor licensing, as well as ongoing maintenance and support plans.

Figure 13. HCI Node Costs



In traditional IT infrastructure, higher performance platforms command cost premiums. With HCI solutions, the performance delivered per node will determine the total number of nodes needed to attain defined workload performance requirements—the fewer the nodes required, the lower the total upfront cost.

ESG used the IOPS per node performance data gathered from the mixed workload, 70% read, 100% random test (see Figures 9-12) to extrapolate how many nodes would be required per cluster to support higher levels of aggregate IOPS. Our goal was to determine the relative CapEx cost of a cluster to support each level of performance.

To do this, we made two assumptions: First, that each cluster would scale linearly, and second, that each solution would have the same cost per node. As Table 2 shows, both software-only solutions require at least one and as many as four more nodes in the 500,000 IOPS performance category, and as many as eight more nodes at 1 million IOPs, as compared with Cisco HyperFlex to support a given mixed workload.

This translates to a potential cost savings of up to 30% in the examples shown, but as a cluster scales, the savings may be higher, as not all HCI systems scale perfectly linearly and not all solutions will be priced at the same cost per node.

Table 2. Extrapolated Nodes Required for Increasing IOPS Levels—70% Read Mixed Workload

Platform	500,000 Aggregate IOPS		750,000 Aggregate IOPS		1,000,000 Aggregate IOPS	
	Calculated Nodes	Total Nodes Required	Calculated Nodes	Total Nodes Required	Calculated Nodes	Total Nodes Required
Cisco HyperFlex – Fully Engineered HCI with Cisco UCS	8.52	9	12.78	13	17.04	18
Vendor A Software-only HCI Validated on Cisco UCS	12.72	13	19.08	20	25.44	26
Vendor B Software-only HCI Validated on Cisco UCS	9.57	10	14.35	15	19.13	20

Source: Enterprise Strategy Group

While a CapEx savings of 30% is significant, it's also important to note that the additional nodes required to meet these performance requirements drag operational expenditures (OpEx) along with them in the form of greater staff time to manage a higher node count, added maintenance, additional power and cooling, potential rack space cost if the cluster is in a hosted environment, and additional software licensing for applications licensed by core count. These areas were not analyzed for this report, but it's important to note that the true TCO savings extends beyond the upfront cost of the nodes.

Why This Matters

ESG research asked 306 IT managers and executives what benefits their organizations have realized as a result of deploying a hyperconverged infrastructure technology solution and the top two most-cited reasons were improved scalability and improved total cost of ownership.[7] Executives want IT to purchase new technologies to modernize their infrastructures and meet business requirements, but they prefer to not spend a lot to do so.

ESG Lab validated that Cisco HyperFlex all-flash systems delivered higher performance than other similarly configured HCI solutions using simulated OLTP, SQL, and mixed workloads. HyperFlex not only outpaced competitors in terms of IOPS and latency, but it also offered more consistent, predictable performance per VM and per node than both software-based systems. This translates directly to lower upfront and ongoing costs, because a given workload can potentially be serviced by a smaller number of Cisco HyperFlex nodes.

[7] Source: ESG Master Survey Results, *Converged and Hyperconverged Infrastructure Trends*, October 2017.

The Bigger Truth

Hyperconverged infrastructures, while becoming mainstream, have long been considered more appropriate for tier-2 workloads. When asked in 2016 why they would choose converged infrastructure over hyperconverged, ESG research survey respondents' most-often-cited (54%) response was better performance. In addition, 32% of respondents believed converged, i.e., loosely integrated independent components packaged together, was better for mission-critical workloads.[8]

Fast forward to 2018, and the picture has shifted, with only 24% of respondents citing performance as a reason to choose converged, while just 22% believe converged is better suited to tier-1 workloads.[9]

Cisco has an answer to those assumptions. HyperFlex provides the typical benefits of HCI—it is cost-effective, simple to manage, and lets organizations start small and scale. But it also provides the performance that mission-critical, virtualized workloads demand. The *consistency* of performance over time and across all VMs in a cluster was particularly notable. In addition, its independent resource scalability enables organizations to adapt quickly to changing requirements, as today's environments demand.

Cisco HyperFlex HCI solutions are highly integrated, fully engineered systems powered by the latest generation of Intel Xeon processors and provide pre-integrated clusters that include the network fabric, data optimization, unified servers, and choice of hypervisor including VMware ESXi/vSphere and Microsoft Hyper-V, enabling fast deployment. This makes them simple to manage and scale. ESG Lab validated that HyperFlex provides consistent high performance for VMware environments running mission-critical workloads. HyperFlex outpaced multiple anonymous competitive solutions with higher IOPS, lower latency, and better consistency over time and across VMs.

The test results presented in this report are based on applications and benchmarks deployed in a controlled environment with industry-standard testing tools. Due to the many variables in each production data center environment, capacity planning and testing in your own environment are recommended. While the methodology in these tests was more stringent than most, customers are well advised to always explore the details behind any vendor testing to understand the relevance to your environment.

When market evolution changes the buying criteria in an industry, there is often a mismatch between what customers want and what they can get. Vendors that can see what's missing and fill the void gain an advantage. Cisco delivers an HCI solution that provides the essential simplicity and cost-efficiency features of HCI, but also the consistent high performance that has been missing—and that customers need for mission-critical workloads. HyperFlex supports VMware and Microsoft on-premises virtualized environments, and expansion to bare metal, containerized, and multi-cloud environments.

HCI solutions have been focused on second tier workloads, but the consistent, high performance offered by Cisco HyperFlex is extremely well-suited to tier-1 production workloads. Organizations seeking cost-effective, scalable, high performance infrastructure solutions for mission-critical workloads would do well to take a close look at Cisco HyperFlex.

[8] Source: ESG Research Report, *The Cloud Computing Spectrum, from Private to Hybrid*, March 2016.
[9] Source: ESG Master Survey Results, *Converged and Hyperconverged Infrastructure Trends*, October 2017.

Enterprise Strategy Group is an IT analyst, research, validation, and strategy firm that provides market intelligence and actionable insight to the global IT community.

 www.esg-global.com

 contact@esg-global.com

 P. 508.482.0188