

Leading Hyperconverged Infrastructure Solutions Diverge Over QoS

Hyperconvergence is Reshaping the Enterprise Data Center

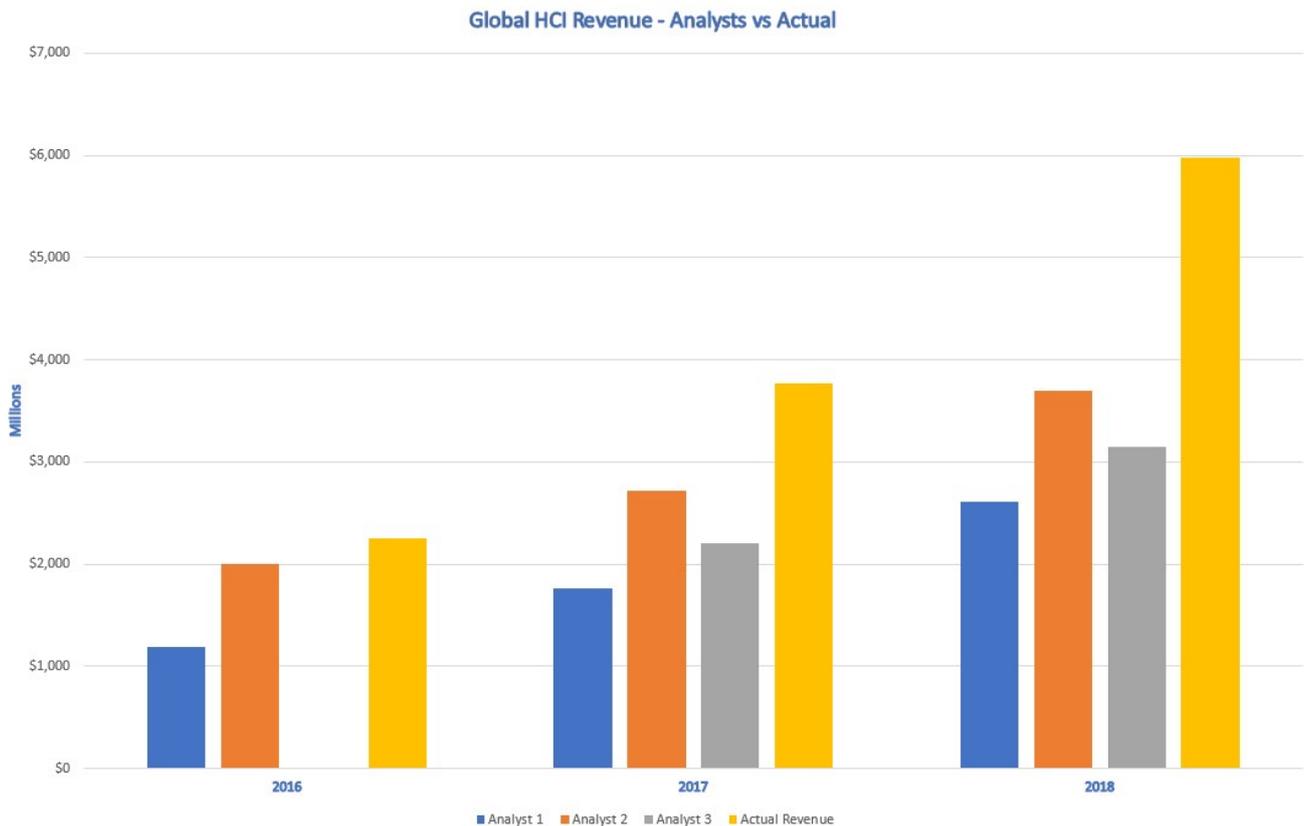
Virtualization largely shaped the enterprise data center landscape for the past ten years. Hyper-converged infrastructure (HCI) is beginning to have the same type of impact, re-shaping the enterprise data center to fully capitalize on the benefits that virtualizing the infrastructure affords them.

Hyperconverged Infrastructure Defined

[DCIG](#) defines a hyperconverged infrastructure (HCI) as a solution that pre-integrates virtualized compute, storage and data protection functions along with a hypervisor and scale-out cluster management software. HCI vendors may offer their solutions as turnkey appliances, installable software or as an instance running on public cloud infrastructure. The most common physical instantiation of—and unit of scaling for—hyperconverged infrastructure is a 1U or 2U rack-mountable appliance containing 1–4 cluster nodes.

HCI Adoption Exceeding Analyst Forecasts

Hyperconverged Infrastructure (HCI)—and the software-defined storage (SDS) technology that is a critical component of these solutions—is still in the early stages of adoption. Yet according to IDC data, spending on HCI already exceeds \$5 Billion annually and is growing at a rate that substantially outpaces many analyst forecasts.



HCI Requirements for Next-Generation Datacenter Adoption

The success of initial HCI deployments in reducing complexity, speeding time to deployment, and lowering costs compared to traditional architectures has opened the door to an expanded role in the enterprise data center. Indeed, HCI is rapidly becoming the core technology of the next-generation enterprise data center. In order to succeed as a core technology these HCI solutions must meet a new and demanding set of expectations. These expectations include:

- Simplified management, including at scale
- Workload consolidation, including mission-critical

The Role of Quality of Service in Simplifying Management and Consolidating

Workloads

Three performance elements that are candidates for quality of service (QoS) management are latency, IOPS, and throughput. Some HCI solutions address all three elements, others manage just a single element.

HCI solutions also take varied approaches to managing QoS in terms of fixed assignments versus relative priority. The fixed assignment approach involves assigning minimum, maximum and/or target values per volume. The relative priority approach involves assigning each volume to a priority group—like Gold, Silver or Bronze.

Superior QoS technology creates business value by driving down operating expenses (OPEX). It dramatically reduces the amount of time IT staff must spend troubleshooting service level agreement (SLA) related problems.

Superior QoS also creates business value by driving down capital expenses (CAPEX). It enables more workloads to be confidently consolidated onto less hardware. The more intelligent it is, the less over-provisioning (and over-purchasing) of hardware will be required.

Finally, QoS can be applied to workload performance alone or to performance and data protection to meet service level agreements in both domains.

How Some Popular Hyperconverged Infrastructure Solutions Diverge Over QoS

DCIG is in the process of updating its research on hyperconverged infrastructure solutions. In the process we have observed that these solutions take very divergent approaches to quality of service.

[Cisco HyperFlex](#) offers QoS on the NIC, which is useful for

converged networking, but does not offer storage QoS that addresses application priority within the solution itself.

[Dell EMC VxRail](#) QoS is very basic. Administrators can assign fixed IOPS limits per volume. Workloads using those volumes get throttled even when there is no resource contention, yet still compete for IOPS with more important workloads. This approach to QoS does protect a cluster from a rogue application consuming too many resources, but is probably a better fit for managed service providers than for most enterprises.

[Nutanix](#) “Autonomic QoS” automatically prioritizes user applications over back end operations whenever contention occurs. Nutanix AI/ML technology understands common workloads and prioritizes different kinds of IO from a given application accordingly. This approach offers great appeal because it is fully automatic. However, it is global and not user configurable.

[Pivot3](#) offers intelligent policy-based QoS. Administrators assign one of five QoS policies to each volume when it is created. In addition to establishing priority, each policy assigns targets for latency, IOPS and throughput. Pivot3’s Intelligence Engine then prioritizes workloads in real-time based on those policies. The administrator assigning the QoS policy to the volume must know the relative importance of the associated workload; but once the policy has been assigned, performance management is “set it and forget it”. Pivot3 QoS offers other advanced capabilities including applying QoS to data protection and the ability to change QoS settings on-the-fly or on a scheduled basis.

QoS Ideal = Automatic, Intelligent and Configurable

The ideal quality of service technology would be automatic and intelligent, yet configurable. Though none of these

hyperconverged solutions may fully realize that ideal, Nutanix and Pivot3 both bring significant elements of this ideal to market as part of their hyperconverged infrastructure solutions.

Enterprises considering HCI as a replacement for existing core data center infrastructure should give special attention to how the solution implements quality of service technology. Superior QoS technology will reduce OPEX by simplifying management and reduce CAPEX by consolidating many workloads onto the solution.