NetApp HCI

NetApp HCI is a turnkey, compute and storage solution based on software that runs the SolidFire scale-out all-flash arrays. Like other hyperconverged infrastructure appliances, the NetApp HCI pools the flash capacity on each node, providing storage to VMs running on nodes in the cluster. It runs on industry-standard Intel x86 server hardware and commodity storage devices, but has a disaggregated architecture that separates storage and compute resources into separate nodes, allowing for a more efficient configuration and more granular expansion than most other HCI solutions.

There are over a dozen vendors marketing HCI appliances, besides NetApp. Competitive solutions include Nutanix, Cisco HyperFlex, HPE SimpliVity and Dell EMC VxRail, among others. These and the other leading products are included in the Evaluator Group Research Hyperconverged comparison matrix and associated Product Briefs, Analyses and Evaluation Guide.

Highlights

- Characteristics
  - Performance – SolidFire software stack designed for scale-out, all-flash storage nodes
  - Efficient Scaling - Separate Storage and Compute nodes eliminates wasted resources
  - Scale-out Expansion – Compute scales from 2 to 64 nodes per cluster, storage scales from 4 to 40 nodes
  - Flexible Scaling – NetApp HCI can support mixed node configurations
  - Non-disruptive scalability - automatic data distribution and load balancing
  - Quality of Service – QoS features assure consistent performance
  - Data Resiliency – dual redundant copies of data distributed to all nodes, automated drive rebuilds
  - Data Protection – native snapshot-based backup and restore functionality to object storage via S3 or SWIFT compatible API
  - Replication – Synchronous, asynchronous and snapshot replication locally and between remote clusters
  - Thin Provisioning – Reservation-less thin provisioning
  - Deduplication – Global, Inline deduplication is always on
  - Compression – In-line and post-process compression, together with deduplication provides 5-10x data reduction (SolidFire claim)
  - High Availability – Automated failover and failback available between clusters
  - Encryption – 256-bit data-at-rest encryption
  - Connectivity – Storage connectivity via iSCSI, VMware VVOLs, and container native storage via NetApp Trident
  - Deployments – Public cloud deployments and cloud service providers in enterprise environments
- Applications

  - Public cloud deployments and cloud service providers
  - Private cloud deployments in enterprise environments
  - Hyper-scale deployments that support high-performance, multi-tenant environments
- Integrations
  - VMware vCenter plug-in, VAAI, SRM/SRA, Microsoft VSS Provider, PowerShell
  - Integrated with OpenStack, CloudStack, Flexiant, and OnApp frameworks
- Deployment and Administration
  - NetApp HCI turnkey appliances
  - Designed for automated management environments with comprehensive API

Overview of System

SolidFire, a start-up company that was acquired by NetApp in 2016, developed and marketed a line of scale-out, all-flash appliances targeting cloud service providers and enterprise environments. The operating system for these SF Series appliances, called “Element OS”, is the software foundation for the NetApp HCI.

A NetApp HCI cluster is comprised of a minimum of four storage and two compute nodes, but can expand up to 100 nodes in most any combination. This “disaggregated” architecture allows the NetApp HCI to be configured with the optimum storage and compute capacity to support a given set of workloads, and to scale more efficiently, adding only the resources required. The SolidFire OS was designed for all-flash storage systems and has the performance to support more VMs per storage node than most other HCIs.

Inline data deduplication and data compression are used to increase the effective capacity of the cluster. Other advanced features include remote replication, read-only snapshots and encryption, plus a backup and restore capability where blocks from a LUN can be written to any S3- or Swift-compatible object storage system without the use of host server compute resources or backup software.

Figure 1: NetApp H410 Hyperconverged Infrastructure Appliance

Hardware Architecture

NetApp HCI appliances are 1U or 2U server chassis that contain up to four independent compute and/or storage nodes. Storage nodes do not require a hypervisor as they run the SolidFire Element OS on bare-metal, not in a VM as most HCIs do. Compute nodes come with dual CPUs in multiple core and memory configurations and do require a hypervisor, currently supporting VMware only. (See Table below).

Each node contains drives for data storage and data management (metadata drives), as well as an NVRAM write cache for reduced write latency. Nodes are connected to a 10GbE storage and a 1GbE management network. The management network can be configured with IPv4 or IPv6.

Each NetApp HCI cluster must contain four storage nodes and two compute nodes but can contain any combination of nodes after that, up to a maximum of 40 storage and 64 compute nodes. Hardware expansion can occur in single-node increments, non-disruptively, with automatic load balancing. This is a core advantage of NetApp's disaggregated architecture, giving it the ability to expand storage or compute power independently. When first released, NetApp HCI appliances were 2U server chassis that contained up to four ½ RU x 1U nodes, the H410C compute and H410S storage nodes. The company has added a 1U dedicated storage node with up to 12 SSDs as well as 1U and 2U dedicated compute nodes with optional GPUs.

Storage Nodes	Rack Units	SSDs /node	Max SSD Capacity	Storage Performance
H410S	1U x 1/2 RU	6	1.92 TB	50K - 100K IOPS
H610S	1U	12	3.84 TB	100K IOPS
Compute Nodes	**Rack Units**	**Max cores/node (all dual CPUs)**	**Memory Capacity**	**Optional GPUs**
H410C	1U x 1/2 RU	40	1 TB	-
H610C	2U	32	512 GB	2 x NVIDA Tesla M10
H615C	1U	48	1.5 TB	3 x NVIDIA Tesla T4

Table 1: NetApp HCI Storage and Compute Node Configurations

Software Architecture

Data Layout and Scaling

NetApp HCI uses Element OS, a software-defined storage system that runs on each node and creates a shared pool of block storage for the cluster. The maximum volume size is 16 TB. To provide file services, ONTAP Select software is installed on a storage volume (see "ONTAP Select").

Element OS uses a distributed replication algorithm called "Helix" that parses volumes into blocks, duplicates each block and spreads them evenly across *all* the available drives in the cluster, with each copy residing on a separate node. This layout creates a single, shared pool of block storage that automatically balances data as nodes are added or removed from the cluster. Each storage node adds capacity and storage processing power, allowing the cluster to scale performance linearly to the millions of IOPS, and newly added capacity is available to every volume without reallocating volumes over new drives. This non-disruptive scaling enables incremental expansion of the cluster, without impacting operations. Clusters can also contain nodes with any valid configuration of storage devices, allowing older nodes to be replaced with newer configurations. While important in scale-out architectures, not all HCIs offer this kind of flexibility.

Evaluator Group Comment

NetApp SolidFire originally used the term "shared nothing" to describe their scale-out storage system in which each node contained storage processing and storage devices, instead of sharing a common pool of drives or a pair of storage controllers. For SolidFire, the all-flash storage system that was competing mostly with scale-up architectures, this was an important distinction. However, the software-defined storage layer in most HCIs has a similar "shared-nothing" architecture. Note: NetApp HCI's disaggregated architecture is a different aspect of this product's technology, one that is NOT shared by most HCIs.

Data Resiliency

Protection against device or node failures is provided by a "RAID-less" process of distributing data blocks around the cluster. The Helix architecture assures that redundant data blocks are maintained on at least two storage nodes. Most software-defined storage systems use some form of RAID to provide protection against device and node-level failures, copying each drive's data to the other drives in the RAID set (typically 2 to 12 drives). This increases the IO load on those remaining drives 2-5x during rebuild putting system performance in a degraded mode.

In a NetApp HCI system there are no spare drives or nodes, all resources are active and drive rebuilds are automatic. When a failure occurs, Helix recreates the affected blocks across the remaining nodes to maintain data availability. Also, Helix doesn't rebuild a failed drive to a single device; instead, data is

rebuilt to the free space on all remaining drives, simultaneously. The result is a rebuild process that's much faster than traditional RAID reducing the potential vulnerability from another failure and improving performance consistency.

Element OS supports Protection Domains on NetApp HCI clusters of three or more chassis with two storage nodes in each. When enabled (which is automatic with the appropriate node and chassis configuration), the system will lay out data such that a chassis failure can be sustained without data loss.

Performance

HCIs have two performance components, storage and compute, that must be supplied in sufficient amounts to support a given number of application VMs. Storage performance, the measure of a system's ability to provide storage IO to a VM workload within an acceptable response time, is a function of the storage devices and storage software stack. Compute performance is the system's ability to process program instructions, based on the product of a system's clock speed and available CPU cores, for a given class of processors.

While enough storage and compute performance must be be available for *each* workload, HCI environments typically support multiple workloads, usually a mixed collection of workloads. This makes sizing and configuration among the most important design decisions and VM density, or the number of virtual machines a node can support, the most useful measure of performance (see "Evaluator Group Comment" below).

Performance and Cost

When sizing an HCI cluster for a given set of workloads, storage and compute performance must be considered, since both are obviously required. Most HCIs include storage and compute resources in each node, meaning the lower of these two performance numbers determines the minimum node count to support a given set of workloads. This can lead to overprovisioning, since a node with the minimum required storage performance may have more compute power than is needed, or vice versa.

The NetApp HCI, with its disaggregated architecture, can address this problem by enabling a cluster to be configured with the minimum number of storage nodes and compute nodes required to support these independent functions. Separating storage from compute functions also improves performance, because the same CPU isn't being used for both compute and storage IO tasks.

Of the two, storage performance is the biggest differentiator of a particular HCI's overall performance, since it is comprised of the software defined storage layer and the IO stack, that are unique to most HCIs. Many HCIs run their SDS software as a VM on each node. This architecture incorporates the hypervisor into the software stack and can impact storage performance. The NetApp HCI runs Element OS on bare-metal, eliminating that potential impact on performance. Element OS was originally developed for (and is still used with) the SolidFire all-flash storage system, so its architecture is designed for all-flash. This creates a storage system that can fully utilize the performance of solid-state devices, compared with other SDS technologies that were originally developed for systems with hard disk drives. All of these factors add up to better VM density (fewer nodes required) and a lower cost (see Evaluator Group Comment below).

Evaluator Group Comment:

When determining VM density, performance testing with actual workloads is ideal, although not realistic. Therefore, the testing platform should use workloads that are representative of the intended environment and published, so that results can be repeated. The IOmark platform uses VMware's VMmark workloads to calculate storage and compute performance, measuring the minimum number of virtual machines that can be supported for a given node configuration, while staying under a latency threshold (complete details are available at IOmark.org).

Quality of Service

Quality of Service (QoS) is a term used to describe a storage system's ability to guarantee a minimum level of performance, based on IOPS, for a host or application in a multi-tenant environment. Element OS allows maximum and minimum IOPS levels to be set for each volume, with the ability to burst IOPS consumption so long as it doesn't impair other volumes' abilities to get their assigned minimums. The following IOPS levels can be set with policies and can be changed on the fly.

Minimum IOPS is the guaranteed threshold of performance assigned each storage volume. When the system has been overprovisioned, the total min IOPS for active volumes exceeds system capacity, performance for all volumes will be scaled down proportionate to each volume's minimum IOPS, making this essentially a prioritizing function.

Maximum IOPS is the sustained performance limit a volume will be allocated over a period of time. It has the effect of keeping certain workloads in a multi-tenant environment from becoming "noisy neighbors".

Burst IOPS is a temporary performance limit assigned to each volume that allows for transient spikes in demand, without impacting the ability of the system to support long-term resource sharing.

Evaluator Group Comment:

Quality of Service has become a "checkbox" item for many software-defined storage solutions, in which most products implement this feature by limiting the number of IOPS they allow each host to consume. However, placing thresholds on some hosts is an indirect way to ensure that resources are available for others. The SolidFire software in the NetApp HCI has a developed a more sophisticated process that directly controls performance maximum and minimum levels and allows for spikes in demand. The result is greater QoS control and better efficiency.

Data Reduction

The NetApp HCI's Element OS provides always-on, in-line deduplication, compression and thin provisioning. The company claims these technologies deliver between 5x and 10x data reduction. Each write is divided into 4KB blocks which are hashed and distriubted across the cluster (see "Data Resiliency"), each with a unique block ID that's compared with subsequent writes to confirm when blocks are duplicates. This global comparison is made with blocks across the entire cluster to improve dedupe effectiveness. In addition to in-line compression, SolidFire employs a post-process compression algorithm to further reduce the amount of data written.

Networking

Deployment of the NetApp HCI cluster requires a network that supports MTU 9000 on all 10/25GbE ports, separate VLANs for management, vMotion and iSCSI, and all host ports with spanning tree configured to PortFast. For maximum performance, the NetApp HCI cluster should be configured with six cables per node, two each for management, virtual machines and storage. Otherwise, a two-cable-per-node configuration can be used with VMware vSphere Distributed Switch which requires a VMware Enterprise Plus license.

Management

Management Node for Element Software

The Management Node is a brower-based user interface to the Element OS that provides system services and upgrades, manages setting and resources, conducts system tests and configures Active IQ for monitoring (see "Active IQ"). It also has a REST-based API that enables system control via 3rd party management systems or through a custom-built solution. The API has been intergrated with the OpenStack, CloudStack, Flexiant and OnApp frameworks.

The RedHat and NetApp HCI solutions combine technologies from Red Hat, such as OpenShift Container Platform, OpenStack® Platform and Ansible®, enabling the development and deployment of applications through virtual traditional or cloud-native methods, with single point of management, orchestration, and monitoring. With NetApp OpenStack contributions DevOps can provision compute, storage, and networking resources for enterprise storage in production-ready private clouds powered by OpenStack.

vCenter Plug-In

The NetApp HCI supports the VMware vCenter Server plug-in (VCP) through the NetApp Element Plug-in for vCenter for configuring and managing NetApp HCI clusters. The VCP can discover HCI nodes, configure and expand storage clusters and allocate storage to configure datastores and virtual datastores. With this plug-in, admins can also set up and run data protection tasks and manage QoS policies while monitoring the infrastructure with real-time reporting and alerts.

Active IQ

Active IQ is a Software-as-a-Service offering that enables IT administration to address potential threats before they arise, and to make smarter decisions about storage for optimal data management. It collects data from NetApp HCI systems and uses predictive analytics to generate actionable intelligence. Active IQ leverages the collective installed base of Element OS systems (over 300,000 worldwide) to create and maintain baselines from which to compare each system and predict problems. Custom access to the data is also possible for more detailed analysis

Continuous risk assesement, automated case opening and predictive alerts for certain conditions is also included in the reporting information. In additon, growth of capacity is reported and projections made for planning purposes, doing some of the storage resource management and optimization functions.

Fabric Orchestrator

NetApp Fabric Orchestrator is a cloud-based data manager that discovers, tags and organizes data from ONTAP systems, Cloud Data Services, NetApp Kubernetes Service or NetApp HCI clusters. This extensible cloud service supports collaboration, protection and compliance while reducing risk and improving resource optimization for cloud-based or on premises data assets.

Cloud Insights

Cloud Insights is an infrastructure monitoring and optimization tool for public (AWS, Azure, GCP) and private clouds (NetApp HCI and on-premises infrastructure) that helps lower cost and reduce down time. Its dashboard displays over fifty types of information, providing end-to-end analytics, advanced anomaly detection and troubleshooting.

Data Services

NetApp Data Fabric and Cloud Volumes

Data Fabric is NetApp's multi-endpoint data architecture based on the core Cluster ONTAP operating system. It connects NetApp HCI private cloud to other ONTAP instances in NetApp systems on-premises or remotely, or to ONTAP in the public cloud.

Cloud Volumes for AWS or Google Cloud Services is a cloud-native file service that provides NAS volumes over NFS and SMB/CIFS with all-flash performance. This service enables any workload, including legacy applications, to run in the AWS cloud or Google cloud, and is also available on Azure as “Azure NetApp Files”. It does run on ONTAP OS, but is sold completely as a service, on a per-GB, per-month basis.

Cloud Volumes (CV) ONTAP is a software-only storage appliance built on the Clustered Data ONTAP OS with support for NFS, SMB/CIFS and iSCSI protocols. CV ONTAP is sold as a license and deployed in the public cloud, but managed and updated by the user, offering the same features and data services as on-premises deloyments of ONTAP.

NetApp also offers Cloud Volumes ONTAP running on NetApp HCI as a service, deployed by NetApp and sold by capacity consumed. This bundle includes NKS (see “NetApp Kubernetes Service”) and can be managed by the user or by NetApp. Again, CV ONTAP support both block and file protocols, where Cloud Volumes are file services only.

Netapp ONTAP Select

ONATP Select is NetApp's NFS, SMB/CIFS and iSCSI services in a software-defined storage application that runs as a VM on VMware or KVM. ONTAP Select software can be installed on a server or HA pair or on an HCI cluster to convert an available storage volume into an ONTAP storage platform. This storage volume can come from a traditional storage array, from a server's internal drives or from HCI shared storage. Each node can support up to 400TB of raw capacity.

ONTAP Select is a simple way to deploy file and block storage capacity on the HCI cluster, that incorporates the features set of NetApp's ONTAP platform and is managed with a common set of tools. Like CV ONTAP, ONTAP Select also provides NetApp Data Fabric connectivity and support for most of the features available on a NetApp storage system, including SnapMirror, SnapVault, FlexClone, SnapRestore and NetAp Volume Encryption. Additionally, SnapLock Enterprise, FabricPool, FlexCache, SyncMirror, MetroCluster SDS and NetApp Data Availability services are offered with the purchase of a separate license.

Data Protection – Volume Snapshots and Replication

Element OS has snapshot-based backup and restore that enables any volume to be rolled back to a previous state, but they can't be mounted or written to. However, snapshots can be used for replicating copies locally or to a remote location as a backup, without third-party backup and restore software products. The target for this replication can be any object store or device that is S3- or SWIFT API-compatible. This can make for faster and more efficient backup and restore operations with no limit to the number of hosts and/or applications. The snapshot limit is 32 snapshots per volume. Asynchronous, bi-directional replication allows clusters to be paired with up to four other clusters and the volumes paired 1:1 for replication.

Encryption

NetApp HCI has data-at-rest encryption and is compliant with the Advaned Encription Standard (AES) 256 encryption algorithm. If the SSD media is removed from a node the data will be unusable. This makes it possible to return and/or replace failed parts without data being compromised and prevents unauthorized access to data in the cluster. This type of encryption fills one of the security requirements in healthcare, banking, government agencies, and other industries that deal with sensitive data.

NetApp Kubernetes Service

NetApp Kubernetes Service (NKS) is a SaaS-based Kubernetes orchestration platform for deploying, managing and upgrading Kubernetes clusters in pubic clouds. From the acquisition of StackPoint.io, NKS runs on Azure, AWS and GCP and is designed to provide cloud-native environment with multiple application solutions available, including: Istio, Trident, Prometheus, HELM, fabric8, autoscaler, linked, calico, sys did and GitLab.

Evaluator Group Comments

NetApp was somewhat late to the HCI market but joined this important product segment with some compelling technology. Its disaggregated architecture addresses one of the fundamental drawbacks to hyperconverged products, inefficient resource allocation, by enabling NetApp HCI to scale storage capacity and compute power independently. Along with the intrinsic performance advantages of Element OS, this architecture generates better storage performance, which reduces the number of nodes required to support a given set of workloads. Element OS storage layer designed for all-flash storage and advanced QoS features also provide consistent, guaranteed performance in the mixed workload environments common to HCIs.

While HCIs have historically been a popular choice for small and mid-sized companies, NetApp's HCI is clearly targeted at the enterprise, with a 6-node minimum cluster (4 storage + 2 compute). The Helix architecture provides redundancy without the delays associated with RAID rebuilds, but Element OS does not support stretched clustering. This product only supports the VMware hypervisor but does provide a Kubernetes orchestration platform that runs AWS, GCS and Azure.

Cloud Volumes ONTAP allows companies to set up ONTAP instances in the public cloud that use the NetApp Data Fabric to connect with ONTAP Select running on NetApp HCI in the data center. Together, these technologies enable the creation of a hybrid cloud, the infrastructure deployment model that companies need to embrace. This is all based on the NetApp operating system that leverages the range of ONTAP data services for snapshots, data protection, availability, etc.

More detailed information is available at http://evaluatorgroup.com