Building a Cloud Hosting Solution to Accelerate New Services

Manage more VMs and larger volumes of structured and unstructured data with HCI and object-based storage

Overview

Public cloud hosting is no longer an experiment for organizations attempting to reduce IT costs. Today, it is a key ingredient in transforming businesses and gaining a competitive advantage. Outsourcing crucial IT resources to cloud service providers reduces costs and improves responsiveness to customer needs by increasing IT agility.

Operating a cloud hosting service presents various challenges in serving customers of all shapes, sizes, and needs. In this paper, we examine how you can deploy robust infrastructure services by using hyper converged infrastructure (HCI) and object-based storage. This approach lets you adapt quickly to the dynamics required to support multitenant cloud infrastructures.

ABCs of Storing Data

Traditionally, enterprise IT organizations have maintained two types of application data: structured and unstructured. Structured data applications (for example, databases and other systems of record) require fast data I/O operations in order to satisfy the demands of real-time transactional workloads and provide useful information quickly.

Applications that drive unstructured data, in contrast, prefer a system of folders and directories to help locate unrelated text and media files, as well as other types of distributed content, such as those generated by Internet of Things (IoT) sensors. Unstructured data applications are less concerned with fast I/O and more concerned with ability to find the right needle in the right haystack and do it efficiently at scale. Although transactional high performance isn't required, predictable performance at scale is mandatory.

Of the two, unstructured data constitutes the vast majority of enterprise data. According to many research studies, including IDC studies reported by TechRepublic, unstructured data constitutes 80% or more of all enterprise IT data and is growing at the rate of 55% to 65% per year.

Hyper Converged Infrastructure (HCI)

Service providers are continually looking for new service opportunities that can be brought to market to better serve their current customers and appeal to net-new prospect opportunities. Scale-out HCI technologies have opened the door to new service opportunities, such as on-premises managed private clouds, provider hosted and managed private clouds, and private compute or shared-storage cloud models.

Among HCI solutions, NetApp® HCI enables these new service opportunities by allowing independent addition of compute nodes and storage nodes. Service providers can modularly add I/O performance, data capacity, or both—and charge for them independently.

In addition, the NetApp HCI solution uses differentiated quality of service (QoS) technologies to efficiently manage operations between the hosted storage and host application. The nature of NetApp HCI QoS allows service providers to balance available resources and user requirements, leading to fewer potential "noisy neighbor" performance conflicts, without the requirement of overprovisioning hardware resources. For these reasons, NetApp HCI is a practical solution for hosting structured data and virtualized environments.

Object Storage

Because of the prolific growth of unstructured data, it has become challenging to provide large-scale storage platforms that can support billions (or trillions) of files while facing a future of even more growth. Traditional file-based storage arrays—using NFS and SMB file system protocols—simply were not designed to operate at this scale.

Object storage, which was first proposed in 1995 and delivered in the early 2000s, removes capacity barriers of traditional file-based storage by using a global namespace and universally unique IDs (UUIDs). Data objects are stored in a flat address space that eliminates the hierarchy (and the capacity limitations) of a file system. As a result, object

NetApp®

storage, once relegated to niche archival applications, is increasingly being considered as a mainstream solution for service provider environments. In addition, many enterprise object storage solutions allow you to spread this flat address space across geography to optimize localized access or provide increased service availability.

Because service providers are required to store two distinctly different types of data, and because service providers must aggregate data from many different users with many different priorities, today's service providers are faced with the following quandaries:

- **Structured data.** Can I architect a platform with the agility to always provide high-performance I/O and scale without interruption to my customers?
- **Unstructured data.** Should I consider moving from existing file storage to object storage to address the massive scale needs of emerging applications that my current and prospective customers are developing and wanting to host?

Design Considerations for Hosting Structured and Unstructured Data

NetApp provides multiple data storage solutions for service provider infrastructures, often referred to as infrastructure as a service, or IaaS. Among NetApp IaaS solutions are HCI and object storage systems: NetApp HCI and NetApp StorageGRID®, respectively.

In the remainder of this paper, we discuss several key IaaS criteria that service providers should consider when designing a hosted storage architecture, using the capabilities of NetApp solutions as a case in point.

Quality of Service (QoS)

Traditionally, when multiple applications share infrastructure, IOPS and bandwidth are freely available to all applications across the shared resources. Unfortunately, this can lead to resource contention, which adversely affects latencies and often leads to unhappy customers.

Without a more precise resource allocation, one application or "noisy neighbor" can easily consume an unfair share of the resources, leaving little available for others. This "first-come, first-served" allocation methodology has a huge negative effect on structured data applications in a hosted infrastructure.

Without the ability to granularly carve up application performance and provision to each virtual machine (VM), the only way to confirm a large enough IOPS pool for these VMs is to extensively overprovision an infrastructure. There is no better way to ruin the economics of a shared infrastructure than by being forced to deploy three times as many resources than actually needed.

NetApp HCI contains sophisticated QoS settings that eliminate resource contention and variable application performance caused by noisy neighbors. Each user can be configured with minimum, maximum, and burst settings, providing predictable performance for each application without incurring the capacity sprawl and low utilization that are common with other hyper converged infrastructures.

For unstructured data, NetApp StorageGRID features a dynamic policy engine that lets you configure data management policies to optimize resource availability, latency, data retention requirements, geo-location, and network costs. With this information, service providers can determine where objects should be placed geographically, the number of copies that should be made, and the data retention policy that should be applied, among other things.

Scalability

Designed to be easy to deploy and to manage, NetApp HCI compute and storage resources are combined into a single shared resource pool using hypervisor technology. This simple, efficient design enables service providers to manage aggregated resources across individual nodes as a single federated platform. NetApp HCI consists of modular building blocks of compute and storage nodes that can be scaled independently by nondisruptively adding nodes to the infrastructure. In this way, service providers can provide the performance their users need to operate structured data applications, even as they scale.

NetApp StorageGRID makes it simple to expand unstructured data repositories to stunning levels: thousands of applications, petabytes of storage capacity, billions of objects. The StorageGRID architecture supports up to 16 geo-dispersed regions—all connected in a single global namespace.

With single node capacities of up to 720TB, multiple nodes combine to create a distributed grid capacity of hundreds of petabytes. Unlike other solutions, you do not need to reserve substantial storage capacity or performance headroom when adding nodes to an existing grid, because StorageGRID does not require massive rebalancing of stored data when new nodes are added.

Data Protection

NetApp HCI delivers self-healing resiliency, continuously accessible data, and a range of backup, restore, and disaster recovery options that are crucial in structured data environments. NetApp HCI provides native protection through Helix data redundancy, NetApp SnapMirror® replication across a Data Fabric, and third-party trusted integrations to keep data protected and safe.

When dealing with vast volumes of unstructured data repositories, it is paramount to protect this data on multiple levels. StorageGRID provides two integrated mechanisms

for protecting large-scale object data from loss: cross-site replication and layered erasure coding.

StorageGRID operates over WAN links, giving the platform the capability of complete site loss protection. Copies are made and distributed throughout the platform so that objects are continuously available. In infrastructures with multiple sites, this distribution of copies means that even if access to an entire site is lost, its data is not lost; clients are able to seamlessly retrieve the data from other sites.

Layered erasure coding is the second method used by StorageGRID to protect object data. When StorageGRID is configured to create erasure-coded copies, it slices object data into data fragments, computes additional parity fragments, and stores each fragment on a different storage node. When an object is accessed, it is reassembled using the stored fragments. If a data or parity fragment is corrupted or lost, the layered erasure-coding algorithm can re-create that fragment by using a subset of the remaining data and parity fragments. This approach optimizes the re-creation at node level first, only rebuilding across geography as required.

Storage Efficiencies

The always-on data efficiencies of NetApp HCI maximize data capacity without performance penalty. Combining all-SSD (solid-state drive) performance with highly efficient data distribution and management, NetApp HCI efficiencies include embedded granular thin provisioning, multilayer data compression, and global data deduplication. These techniques solve traditional storage deficiencies, making flash at scale an economic reality for service providers, while delivering the superior performance necessary for structured data applications.

NetApp StorageGRID derives efficiency chiefly through erasure coding, which consumes 50% less disk space when compared to traditional replication or RAID mirroring, while providing superior levels of durability and availability.

Important to service providers, StorageGRID nodes can also be flexibly deployed through Docker containers across bare metal, across VMs running on third-party servers, or as an enterprise engineered hardware appliance. Service providers can use any combination of node types to rapidly deploy petabytes of storage. These flexible deployment options allow service providers to take advantage of low-cost (or no-cost) resources already available within their infrastructure.

Summary

Supporting two distinctly different types of data, with diverse characteristics, is causing cloud service providers to rethink their approach to building hosted architectures. Hyper converged infrastructure, with flash-based performance and modular construction, is well suited for structured data applications. Object-based storage, with the ability to manage billions of files within a flat address space, is best suited for unstructured data applications.

NetApp offers both HCI and object-based storage to meet the needs of today's cloud service providers. Built on the principle of a common Data Fabric, NetApp products reduce service provider costs and help build IaaS revenue with a scale-out architecture that grows easily—without downtime.

For more information:
NetApp IaaS solutions
NetApp HCI
NetApp object-based storage

About NetApp

NetApp is the data authority for hybrid cloud. We provide a full range of hybrid cloud data services that simplify management of applications and data across cloud and on-premises environments to accelerate digital transformation. Together with our partners, we empower global organizations to unleash the full potential of their data to expand customer touchpoints, foster greater innovation, and optimize their operations. For more information, visit www.netapp.com. #DataDriven