HITACHI
Inspire the Next

WHITE PAPER

Hitachi Content Platform Architecture Fundamentals

Secure, Simple and Smart Web-Scale Object Storage Platform Delivers Superior Security, Efficiency and Interoperability

By Hitachi Vantara

June 2020

Contents

Executive Summary

Organizations are swiftly assessing and adopting new technologies and information management practices to defend against and ultimately transcend digital disruptors that are emerging across every industry. Pressure is mounting from internal and external influences alike. IT is in the position to define and lead the digital transformation strategy for the organization. Initiatives such as cloud, big data, mobile, social and data center modernization are all actively being pursued.

With the IT landscape continuing to evolve, it gets even more difficult to ensure the right data is at the right place, at the right time. The scope of sources from which data is being created or accessed is no longer exclusive to traditional applications and workloads. New technologies, third-party applications and mobile devices mean volumes of data are everywhere and constantly changing. The challenge becomes how to retain security, control and visibility of that data, at all times.

Hitachi Content Platform (HCP) is a secure, simple and intelligent web-scale object storage platform that delivers superior scale, performance, security, efficiency and interoperability. It allows any organization to deliver unique, feature-rich, private, hybrid, multicloud, or public cloud storage services at a cost comparable to public cloud. The rich feature set and extensive ecosystem surrounding the platform allow organizations to improve efficiencies and optimize costs. They can choose to move data to on-premises storage tiers, off-site to a choice of public cloud providers or to a combination of both.

HCP serves as the cloud storage platform for a tightly integrated portfolio of offerings built to service a wide range of information management use cases that span traditional and sustaining applications to emergent and disruptive technologies. The portfolio allows customers to execute on their data-centric strategy and make data secure, available, insightful and actionable across end users, edge locations, data centers and clouds. It brings together object storage, file sync and share, cloud storage gateways, and sophisticated search and analytics, to create a tightly integrated, simple and smart cloud storage solution. The HCP portfolio includes Hitachi Content Platform for software-defined object storage; HCP Anywhere for file synchronization and sharing and end-user data protection capabilities; HCP Anywhere Edge for modernizing existing file servers with sync-based data protection and content distribution, HCP Gateway, a cloud file gateway; Hitachi Content Intelligence, for data quality and analytics; and Hitachi Content Monitor for advanced performance monitoring.

Introduction

Organizations are swiftly assessing and adopting new technologies and information management practices to defend against and ultimately transcend digital disruptors that are emerging across every industry. Initiatives such as cloud, big data, mobile and social are no longer just buzz, but imminent.

With the IT landscape continuing to evolve, ensuring the right data is at the right place at the right time is a serious challenge. New technologies, third-party applications and mobile devices mean data is everywhere and constantly changing. The challenge becomes how to retain security, control and visibility of that data, at all times.

A secure, multipurpose and distributed object-based storage system, Hitachi Content Platform (HCP) is designed to support large-scale private and hybrid cloud repositories of unstructured data. The smart web-scale solution enables IT organizations and cloud service providers to store, protect, preserve, retrieve and distribute unstructured content with a single storage platform. HCP supports multiple levels of service and readily evolves with technology and scale changes. With a vast array of data management, data protection and content preservation technologies, the economical system can significantly reduce resource requirements, and even eliminate its own tape-based backups or backups of edge devices connected to the platform.

HCP obviates the need for a siloed approach to storing unstructured content. With massive scale, multiple storage tiers, Hitachi reliability, nondisruptive hardware and software updates, multitenancy and configurable attributes for each tenant, the platform supports a wide range of applications on a single physical HCP instance. By dividing the physical system into multiple, uniquely configured tenants, administrators create “virtual content platforms” that can be further subdivided into namespaces for further organization of content, policies and access. With support for leading APIs, thousands of tenants, tens of thousands of namespaces, petabytes of capacity in one system, and hybrid cloud configurations based on integration with leading public cloud services, HCP is truly cloud-ready.

This white paper describes how the Hitachi Content Platform portfolio provides the ideal ecosystem to support existing content-centric applications and newer cloud use cases and workloads, simultaneously. It also describes new HCP-based functionality and tools that help businesses organize their data: They can extract intelligence and safely share it with a globally dispersed workforce all through a single point of visibility and control.

Hitachi Vantara Content Portfolio Overview

Distinctly unique from the competition, Hitachi Vantara offers an integrated portfolio of the following products to store, protect, manage, access, share and analyze unstructured and semi-structured data:

Hitachi Content Platform: A massively scalable, multitiered, multitenant, multicloud storage solution for small, midsized and enterprise organizations as well as service providers. Hitachi Content Platform transforms existing investments into a cloud storage system, including private clouds, hybrid or multicloud architectures. The rich feature set and extensive ecosystem surrounding the platform allows organizations to improve efficiencies and optimize costs by moving data to appropriate on-premises storage, off-site to one or more public cloud storage providers, or a combination of both. Note that Hitachi Content Platform for cloud scale is based on microservices and is therefore addressed in a separate architecture white paper, although many concepts are shared.

Hitachi Content Platform S series: A highly efficient, highly available, cost-effective storage appliance that supports very large volumes of HCP data in a small footprint. S series systems take advantage of erasure coding to deliver long-term compliance and protection at the lowest cost. Always-on, self-optimization processes also maintain data integrity, availability and durability. S series systems can provide direct-write storage for HCP systems or be used as storage tiering platforms for HCP systems. A single HCP system can seamlessly store data across multiple S series systems, thereby enabling efficient scalability in terms of both capacity and performance.

Hitachi Content Intelligence: A robust solution framework for comprehensive discovery and fast exploration of critical business data and storage operations. Content Intelligence enables organizations to turn multistructured data into valuable business information. It aggregates that data to create a centralized information hub for your workforce to explore, discover and surface actionable business insights quickly.

Hitachi Content Monitor: A robust storage performance monitoring solution for HCP. While HCP provides built-in monitoring capabilities, Hitachi Content Monitor is a tightly integrated, cost-effective add-on that enables comprehensive insights into HCP performance across multiple clusters, to improve capacity planning and simplify troubleshooting. Content Monitor's artificial intelligence features can detect HCP performance anomalies to proactively identify issues before they occur and forecast future HCP performance, based on historical behaviors.

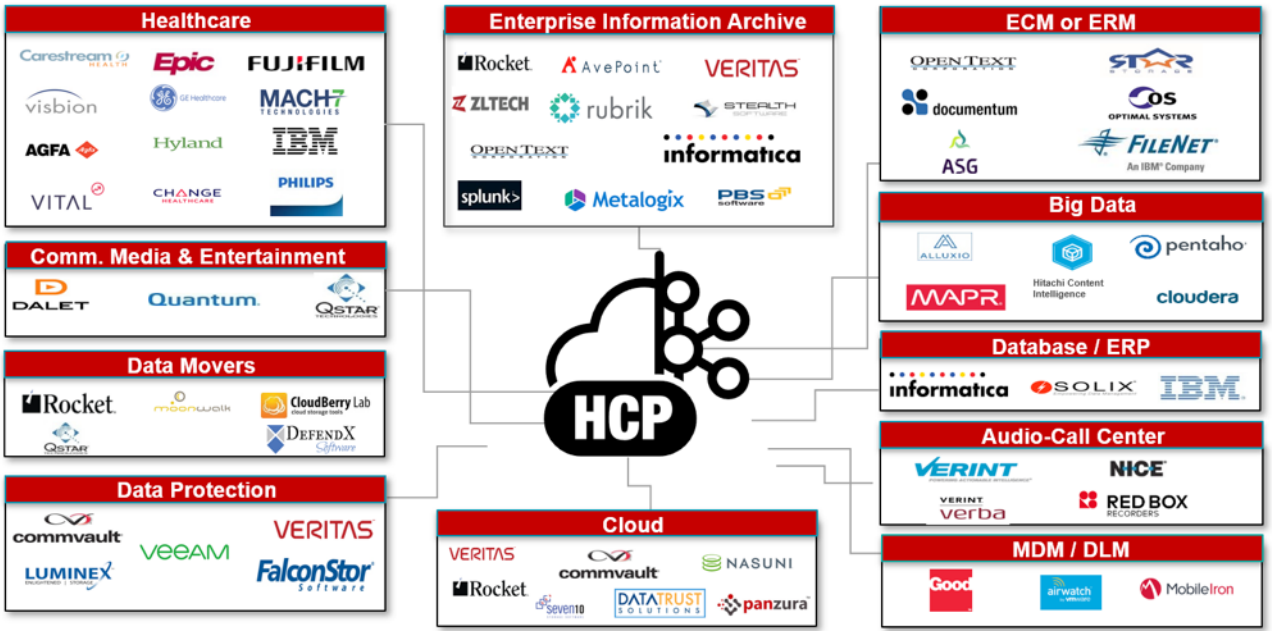
Hitachi Content Platform Anywhere: A secure, enterprise file-sync-and-share solution for protecting, mobilizing and sharing data across your workforce. HCP Anywhere enables a more productive workforce with a corporate-delivered file synchronization and sharing tool that allows for secure access and sharing across mobile devices, tablets and browsers.

Hitachi Content Platform Anywhere Edge: A file server modernization option delivering the innovative sync technology of HCP Anywhere for user devices to remote and branch offices. HCP Anywhere Edge helps right-size and consolidate file servers at the edge by keeping only active data in local storage while allowing easy access to large amounts of data from the private or hybrid HCP cloud in organizations' data centers. HCP Anywhere Edge can be deployed on network-attached storage (NAS) servers to make that data available to HCP Anywhere, where it can be managed from a single location.

Hitachi Content Platform Gateway: A cloud storage gateway that enables users and applications to read and write file data and copies the data to HCP, where it is efficiently stored, well protected and properly governed. Gateway presents the traditional CIFS share or NFS protocols to users and applications using virtual file systems. The virtual file systems are not integrated with the host operating system. Hence, such things as stubs, links, DFS or junction points are not used, and operating system limitations (for example, file count or file system size) are therefore not imposed.

The Hitachi Content Platform portfolio provides the ideal ecosystem to support existing content-centric applications and newer cloud use cases and workloads, simultaneously. It also provides a central way for organizations to securely incorporate hybrid cloud storage on their terms to react faster to change and to optimize costs. The HCP ecosystem includes a number of integrated Hitachi Vantara products and solutions, as well as an expansive set of independent software vendor (ISV) partners and broad protocol support. Examples of HCP ISV and associated industries are depicted in Figure 1.

Figure 1. Hitachi Content Platform Partner and ISV Ecosystem



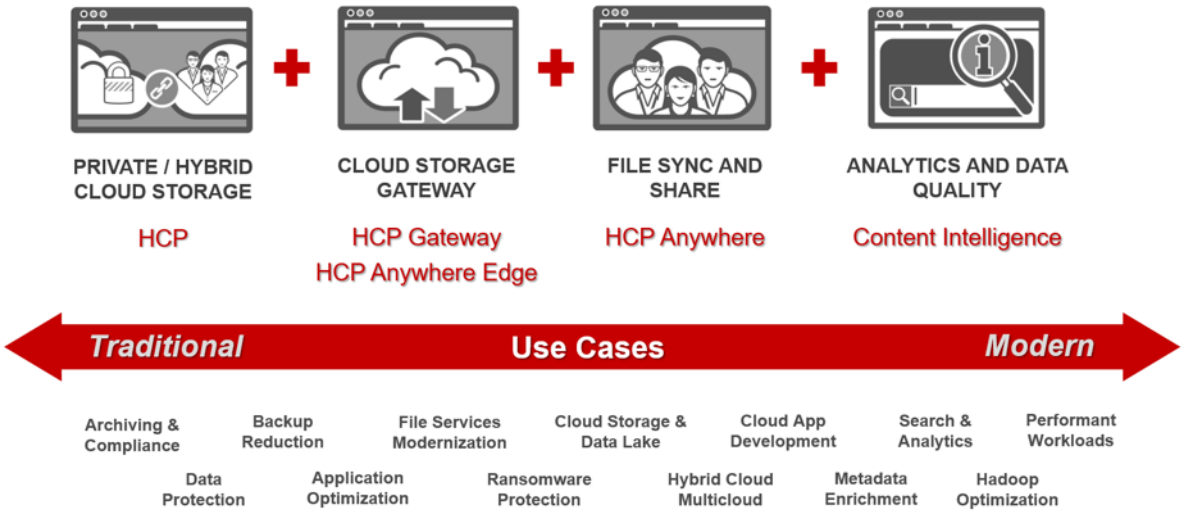
HCP serves as the single foundation for storing data from multiple applications and data sources. Through its flexible architecture, it supports traditional applications, such as a variety of archiving applications (for example, file, email, recordings, database, Microsoft SharePoint and medical images). It also serves as the repository for newer Amazon Simple Storage Service (S3)-enabled cloud, big data, mobile sync and share, remote and branch office, and file and open source application data – all from a single point of management.

With the combination of Hitachi Content Platform, HCP Anywhere, HCP Anywhere Edge, HCP Gateway and Hitachi Content Intelligence, the portfolio of solutions bridges the gaps between traditional IT, cloud and next-generation technologies that extend IT beyond the data center.

Common Use Cases

Hitachi Content Platform is a singular platform that was designed to handle compliance use cases as well as extreme density. Applications written to HCP work seamlessly, regardless of whether HCP is located in the same data center or in the cloud. With HCP, organizations can begin with a very small footprint and grow to have the largest density in the industry. Common HCP portfolio-based use cases are described in Figure 2.

Figure 2. HCP Portfolio Supports Multiple Cloud Storage Use Cases



- ***Archiving:*** Archiving is a common and important use for object storage. Among the primary reasons for its use include providing economical storage that can scale, along with maintaining advanced data protection that removes the need for backup. Nearly 200 different applications have directly integrated with HCP for archiving a variety of data types, such as email, call data records, document management data, healthcare records, medical images, media and entertainment files, and inactive files from file systems.
- ***Regulatory compliance and discovery:*** HCP adds value to archiving with an advanced set of features for retention management, legal hold and automatic data disposition that help organizations meet compliance regulations, such as SEC 17a4, Health Insurance Portability and Accountability Act (HIPAA) and General Data Protection Regulation (GDPR). Combining built-in custom metadata query and content search from applications, such as Hitachi Content Intelligence, HCP allows administrators to identify data subject to litigation and automatically execute legal holds on the data set to prevent deletion.

- ***Backup reduction and optimization:*** HCP's compression, efficient data protection and faster data recall rates provide value beyond backup storage to tape or expensive block deduplication appliances. In particular, HCP provides better reliability and agility than tape while improving recovery time objective (RTO) and recovery point objective (RPO).
- ***Storage for cloud applications:*** Most new applications are being developed leveraging lightweight, web- and cloud-friendly REST and Amazon S3 APIs and targeting private, public or hybrid cloud storage. Object storage, such as HCP, is the most common type of storage presenting these APIs, and optimizes for simple, horizontally scaling storage that can tie a rich set of metadata along with each file.
- ***Unstructured data management/hybrid cloud broker:*** HCP leads in capabilities that support policy-based data management capabilities. Policies can be defined to move data between different classes of storage within an HCP cluster or even external to the cluster, targeting one or more popular public cloud services, such as Amazon, Microsoft Azure and Google. In so doing, HCP can be a broker of these cloud services, moving data to, from and between the services while providing a level of API abstraction for the REST application. HCP provides methods to control geographical placement of data for disaster recovery protection and data distribution. Finally, having the most advanced architecture for custom metadata and built-in query capabilities, HCP is well positioned to act on this information and lead in a new age for data management.
- ***Cloud service enablement:*** Many enterprise IT organizations are revamping their service delivery models to align with growing public cloud service models. Similarly, Tier 2, regional or vertical industry-specific service providers are scrambling to do the same. For these groups, HCP object storage is an excellent storage choice, offering a variety of popular rest APIs and service differentiating features. Furthermore, HCP is the foundation for a tightly integrated portfolio of cloud service applications, including HCP Anywhere for file sync and share, HCP Anywhere Edge for modernizing file servers, and HCP Gateway for file cloud gateway capabilities, and Hitachi Content Intelligence for sophisticated search and analytics. This portfolio provides a quick start to cloud service providing.
- ***Big data storage:*** By its very nature, big data involves massive quantities of mostly unstructured data. Organizations want to unleash strategic value from this data through analysis. HCP safely, securely and efficiently stores this data. HCP's advanced metadata architecture also brings structure to the unstructured data, allowing analytics applications to query for specific subsets of data, hastening analysis and improving results. For example, Hitachi's Lumada Data Optimizer solution can reserve Hadoop for active data while automatically tiering less-frequently-accessed data to HCP while maintaining seamless access via Hadoop Distributed File System (HDFS).
- ***Remote office and branch office (ROBO) file services and content distribution:*** HCP Anywhere Edge and HCP Gateway combine with HCP to deliver elastic and backup-free file services (NFS or CIFS). HCP Anywhere Edge is designed to modernize existing file servers by transforming them into cloud storage gateways. HCP Gateway can be deployed as cloud storage gateway software or as an appliance. When a file is written to Anywhere Edge or Gateway, it is automatically replicated to HCP where it is protected, governed and accessible by other gateways, users and applications. These HCP technologies drastically simplify deployment, provisioning and management by eliminating the need to constantly manage capacity, utilization, protection, recovery and performance of the system.
- ***File-sync-and-share cloud platform:*** Hitachi Content Platform Anywhere provides organizations with a secure file synchronization, sharing and collaboration alternative to consumer-grade or less secure publicly consumed tools. With HCP Anywhere, you can enable a more productive workforce with a file synchronization and sharing tool, delivered from behind the corporate firewall, that allows for secure access across mobile devices, tablets and browsers. End users simply save a file to HCP Anywhere and it synchronizes across their devices. These files and folders can then be shared via hyperlinks. Because HCP Anywhere stores data in HCP, it is protected, compressed, single-instanced, encrypted, replicated and access-controlled. HCP Anywhere also provides enterprise data mobility by enabling mobile access to data in NAS and Microsoft SharePoint storage.

Key HCP Values and Differentiators

Hitachi Content Platform architecture is composed of a comprehensive set of features and capabilities designed to allow organizations to ingest and track information across multiple sources and media types. It simplifies access, search and analysis. It also eases management, improves data protection and lowers costs. These values are enabled through:

- ***Unprecedented capacity scale:*** Start small (4TB) and scale to unlimited capacity. Deploy entirely as software-defined storage (SDS) based on VMware ESXi or kernel-based virtual machines (KVM) or as an appliance cluster. Scale capacity using Ethernet attached storage (HCP S series systems) or Fibre Channel arrays.
- ***Multiprotocol and heterogeneous access:*** Accommodate legacy applications that use NFS, CIFS, SMTP or WebDAV, as well as applications that use modern RESTful APIs including S3 or REST for HCP. Users can write data using any supported protocol and then read data back with another.
- ***Construct hybrid storage pools:*** Using HCP's adaptive cloud tiering (ACT) functionality, manage a single storage pool using any combination of server disks, Ethernet attached HCP S series systems, SAN disks, NFS or a choice of one or more public cloud services, including Amazon S3, Google Cloud Storage, Microsoft Azure, Verizon Cloud, Hitachi Cloud Service for Content Archiving, or any other S3-enabled cloud service.
- ***Multitenancy for application isolation:*** With thin provisioning and capacity quotas, divide your storage resources into thousands of independent tenant and namespace areas, each with independent administration and assigned users.
- ***Powerful service plans:*** Define cradle-to-grave data management plans that govern an object's protection class, access speed and disposition policies.
- ***Compliance storage modes:*** Satisfy regulatory requirements that require immutable, undeletable "write once, read many" (WORM) storage, guaranteed authenticity or proof of chain of custody.
- ***Extensible metadata and search:*** Create and modify custom metadata at any time during an object's life cycle. Multiple authors can have separate sections of custom metadata. Use HCP's API or search console to locate objects for application and analytical use, or to automatically apply legal holds.
- ***Navigate technology transitions:*** With Autonomic Tech Refresh (ATR), HCP boasts a 14-year track record of helping organizations perform online migrations from old to new hardware technology, preserving their application and API investments.
- ***Performance:*** A unique shared-storage architecture that is equally adept at serving small or large objects. Utilize high-performance flash storage to accelerate I/O and deliver consistent responses under heavy loads.
- ***Global access topology:*** Read or write data from any site and control where cached copies reside. Share data, but ensure it is hosted within country or continent boundary.
- ***Portfolio breadth:*** Go beyond the simple archive; choose tightly integrated sync-and-share, file gateway, search and analytics and backup solutions from Hitachi Vantara, or select applications from more than 80 ISV partners.
- ***Data protection and security:*** Capabilities include erasure coding, redundant copy control, RAID-6, AES256 Encryption, 2048-bit SSH service keys, SSL and certificates.
- ***Monitoring:*** The user interface and API provide visibility into hundreds of alerts and event logging, as well as chargeback reporting.
- ***Data durability and efficiency:*** Content verification services, sophisticated self-repair, multisite replication using geo distributed erasure coding, deduplication and compression.

- ***Global support:*** All HCP systems are eligible for free monitoring through the Hitachi Remote Ops (formerly Hi-Track) monitoring system. Hitachi's global service centers are staffed 24/7.

Architecture Overview

An HCP cluster conceptually consists of two required layers (access and storage) and two supplementary layers (load balancing and search). The architecture allows for independent scaling of these layers. HCP clusters may be deployed across multiple locations to form a global access topology with even higher resiliency.

Access Layer

HCP cloud storage software is deployed on virtual appliances (HCP VM) or physical appliances (HCP G series nodes – see Hardware section) that coordinate to provide a highly available access layer. The **access layer** appliances work as a cluster to: process client I/O requests, store the system metadata, manage, virtualize and federate the storage layer. The access nodes provide multiple protocol gateways that allow access to the same stored object with different protocols, which is a feature not commonly found in object storage offerings.

HCP is inherently a distributed system that spreads key functions, such as metadata management, across all access layer appliances. All runtime operations are distributed among the access layer appliances. As such, no single node becomes a bottleneck since each node bears equal responsibility for processing requests, storing data and sustaining the overall health of the system. The nodes work cooperatively to ensure system reliability and performance. Access layer appliances communicate with each other through a private back-end network (see Networking section).

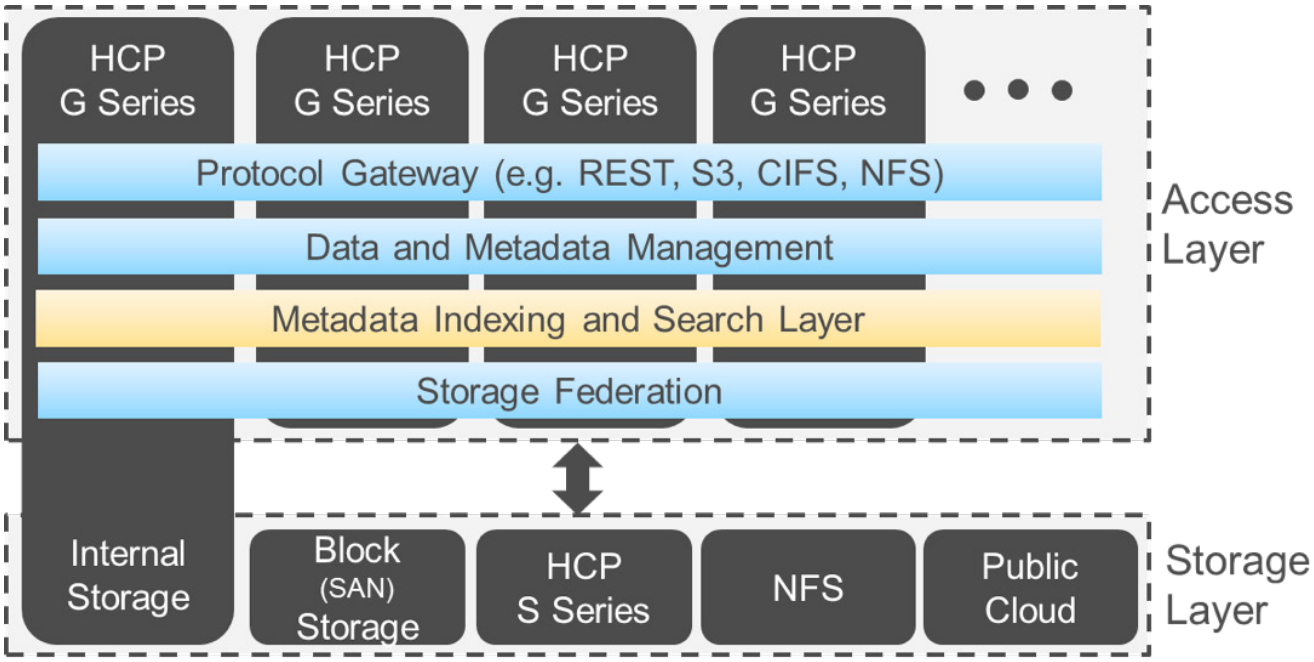
Each appliance in the access layer runs a software stack made up of the appliance operating system and the HCP core software (see Figure 3). All access appliances run an identical software image to ensure maximum reliability and fully symmetrical operation of the system. An HCP access appliance can take over the functions of others in the event of a failure. The load is automatically distributed across the access layer, and traffic control may be enhanced with an optional load balancer.

The HCP distributed processing scheme allows it to scale linearly to accommodate capacity growth or more application clients. When a new access node is added to the HCP system, the system automatically integrates that node into the overall workflow without manual intervention.

These same appliances may be optionally enlisted to store object data; the internal storage components of the access layer appliance may be used as part of the storage layer. Capacity scaling can then be enabled by adding internal storage to the appliance or by adding access layer appliances to the cluster.

The access appliances also provide metadata search capabilities. A Hitachi Content Intelligence solution is optionally available for those who require enhanced search and analytics capabilities.

Figure 3. HCP Access, Storage and Search Layers



Storage Layer

The HCP storage layer provides durable persistence for object data and metadata. It is a virtualized set of storage devices, including erasure coded storage (HCP S series), block (SAN), file (NAS), public cloud, S3-compatible storage, or the HCP G series nodes. HCP ingests new objects to one or more compatible storage devices and may later tier the data to other devices.

The storage layer can scale independently from the access layer, and it can scale storage performance and capacity separately. By adding storage devices, it is possible to increase storage performance, and by adding storage capacity to existing storage devices it is possible to increase storage capacity.

Much more information about the HCP S series is provided in the Hardware chapter of this document.

Load Balancing Layers

HCP has a variety of built-in load balancing capabilities. Ingest and egress I/O to the access layer based on the REST and S3 protocols is automatically distributed across the available access layer nodes when HCP is accessed using the DNS hostname. This ensures that front-end I/O performance can scale transparently as access nodes are added.

The “ownership” and system metadata of objects are also distributed across access nodes based on the object path for cloud protocols, which is how HCP ensures strong consistency in a cluster. For example, a request for writing a new version of an object may be received by access node A but it is handed off to access node B because B “owns” the object. Even if a single access layer node were to receive all of the requests, the processing of the requests will be distributed across the access layer.

As part of the storage layer, HCP can create one or more virtual storage pools that consist of multiple storage devices. When there are multiple storage devices in a storage pool, HCP allocates writes of objects across the

devices, which ensures that a singular storage device does not receive all of the load. HCP can also rebalance objects across devices to improve performance and avoid hot spots.

Beyond utilizing the built-in load balancing capabilities, HCP can also be deployed behind a load balancer (also known as a traffic manager or application delivery controller). The load balancer can be used to improve performance and resiliency by directing traffic to access nodes within an HCP cluster and to coordinate global traffic to multiple HCP clusters. The load balancer can also block unauthorized access by serving as the security boundary for an HCP cluster. Infrastructure featuring separate physical networks can leverage the load balancer to enable appropriate data access for HCP.

Search Layer

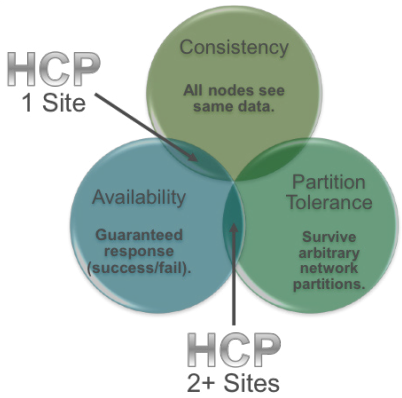
A core principle of object storage is allowing users and applications to find and access objects without requiring a rigid folder hierarchy or descriptive filenames. HCP objects can be enriched with meaningful "custom metadata" that can be used to add relevant information to the object and creates an association with related objects. HCP has an integrated search layer that can be enabled or disabled, and which scales by adding access nodes. Objects written to HCP will have their metadata indexed into a scalable database, which can be searched via an API or a web interface. Further details about HCP's built-in metadata search engine (MQE) are available in the "Object Storage Software Architecture" chapter.

For full content searching of objects, Hitachi Content Intelligence can be used as a comprehensive search engine for HCP objects, among other capabilities. Content Intelligence runs as a software-defined cluster and scales independently of the HCP cluster.

Global Access Topology and Multisite Consistency

HCP installations may or may not span multiple geographical sites (see Figure 4). When considered from a single site perspective, HCP design favors consistency and availability as defined by the Brewers CAP theorem[1] . The theory postulates that a distributed system of nodes can satisfy at most two of these three properties:

Figure 4. Node Properties at HCP Sites



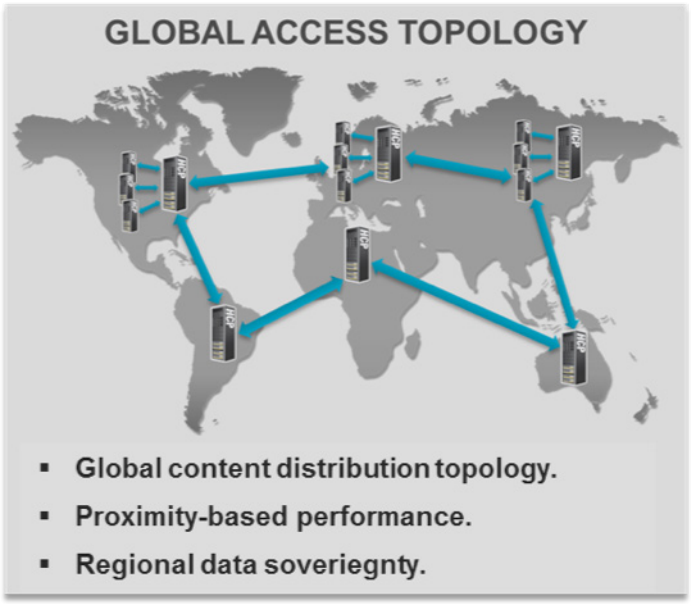
- Consistency. All nodes see the same data at the same time.
- Availability. Every request receives a response about whether it was successful or failed, guaranteed.
- Partition tolerance. The system continues to operate despite arbitrary message loss or failure of part of the system.

[1] CAP Theorem, Eric Brewer, University of California, Berkley

Within a **single site**, HCP will never return stale data, which is critical for applications that require strong data consistency. While HCP can handle many forms of partition failure, it does require that a majority of the HCP access nodes be available and communicating with each other in order to accept write requests. Reads can be processed with as few as one surviving node.

When an HCP deployment spans **two or more sites** and supports an active-active global namespace, HCP favors data availability and partition tolerance over strict consistency, which is also the favored model for public cloud deployments and is referred to as an eventually consistent model. In response to a whole site outage, HCP may deliver data from a surviving site that was not yet consistent with the failing site. This is a result of asynchronous replication but mitigated by HCP's global access topology (see Figure 5), which performs hyper-replication of metadata.

Figure 5. Global Access Topology Hyper-Replication With HCP



Hyper-replication is possible because each HCP system maintains a separate structure for object data versus object metadata. When an application writes an object, metadata is stored in a separate but parallel branch of HCP's internal file system. This physical separation enables many unique capabilities, including better data consistency between sites due to HCP prioritizing metadata replication over replicating the actual object. Intersite consistency is thus less affected by network speed or object size. Participating sites are thus more quickly aware of new or modified objects. A physically separate structure for metadata is also key to HCP search, tiering and fencing capabilities, which are discussed later in this paper. When all sites operate optimally, each HCP can respond to I/O requests with local resources and remain unaffected by the speed or latency of the WAN interconnecting sites.

Deployment Options

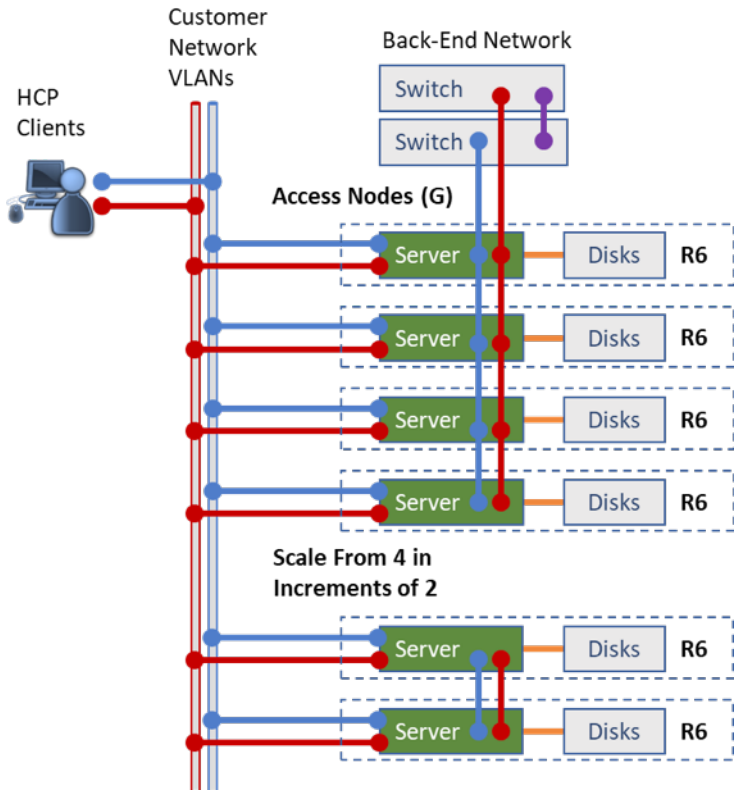
Flexible: Start Small, Scale Big

HCP cloud software offers greater flexibility and choice by deploying as wholly virtual (via hypervisors), wholly appliance, or as a hybrid of both. In all cases, the object repository tolerates node, disk and other physical component failures. The HCP architecture offers deployment flexibility that enables the platform to accommodate small workloads while also being able to scale efficiently to manage larger cloud configurations with ease.

Small Deployment: Shared Nothing Configuration. This deployment is composed entirely of access nodes that can each manage a private pool of internal storage in the 28TB range (see Figure 6). To create the HCP cluster, individual

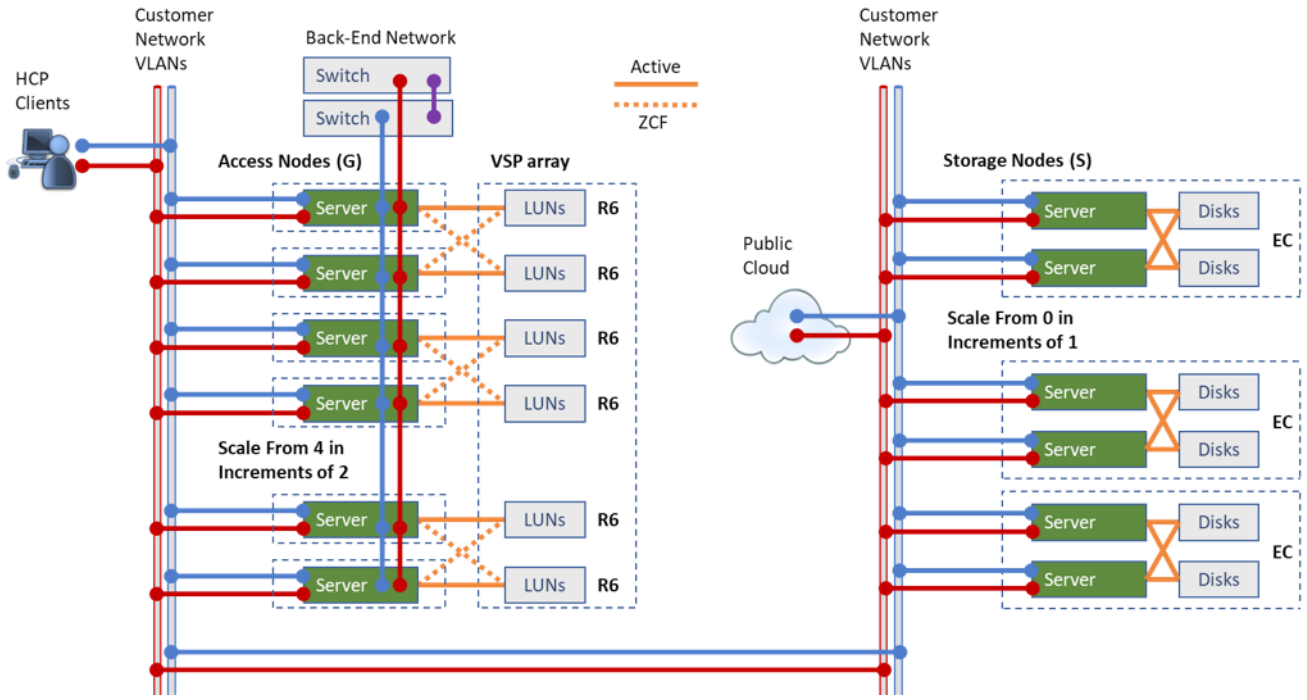
nodes are networked through a set of physical 10Gb Ethernet ports. In this deployment model, capacity scaling is achieved by adding nodes. Although ideal for small, sub-petabyte cloud designs, scaling only with access nodes may needlessly increase the compute capability of the cluster as each increment of storage triggers a significant, potentially underutilized number of CPUs.

Figure 6. HCP Small Deployment



Large Deployment: Shared Storage Configurations. The HCP architecture allows its compute and storage elements to scale independently through the use of shared storage (see Figure 7). Networked storage elements behind an HCP cluster can include any combination of onsite S series systems, Fibre Channel storage arrays [(such as Hitachi Virtual Storage Platform (VSP)], NFS, tape, optical or off-site public cloud storage.

Figure 7. HCP Large Deployment



By incorporating support for off-site public cloud storage targets, HCP encourages adoption of hybrid cloud configurations, which can lower the costs of storing older less-active data. By trading a little performance and latency, organizations gain near instant capacity elasticity while retaining a single point of management for both new and old data.

The Power of Shared Storage

Shared storage lets organizations make hardware investments based on application needs rather than an artifact of architecture design. For example, as the number of clients grows, there is generally a proportional increase on the HCP workload. HCP G series access nodes may be scaled to linearly improve small object performance and large object throughput, or increase CPU power available to HCP search and data services.

Alternatively, an organization may decide to tackle a new application that needs to store larger media or video files. In this case, HCP is not driving a lot of new I/O as much as it is directing many large files. In this case, additional HCP S series systems might be best to quickly add several petabytes to their virtualized storage pool.

In a pure Ethernet deployment, HCP G series nodes and HCP S series systems are networked through a combination of physical 10Gb Ethernet ports and VLANS in a loosely coupled architecture. HCP S series systems are particularly well suited for storage scaling. They can be scaled in approximately 250TB usable storage increments. The flexibility to deploy them in small capacity configurations and scale to nearly one exabyte (EB) per site gives organizations the ability to seamlessly grow as needed and spread capital investments over time.

Management

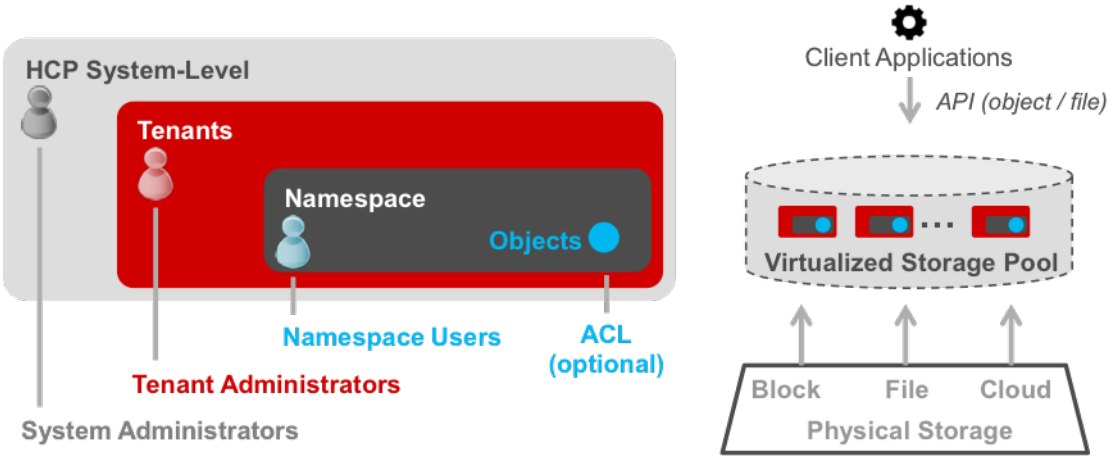
A single HCP administrator can manage exabytes of capacity. HCP supports configurations exceeding 5EB. Harnessing the full potential of HCP's scalable storage capabilities begins with great multitenancy management, delegation and provisioning features (see Figure 8). There is no need to prepurchase or reserve storage for specific

applications. Rather, buy a modest amount upfront, and grow capacity incrementally as demand increases. Manage in general terms, using quotas, and bill users by what they actually consume instead of what they might consume in the future. Offer them service options that appeal to their data usage patterns, such as versioning, compliancy and automated tiering plans to lower the costs of carrying older data.

System-level administration: These management roles cannot read or write data, but they do control how physical storage resources are virtualized and monitored. They design service plans to govern data placement, how it ages and how it is retired. These managers prioritize system services, create **tenants** and delegate control over capacity using a quota system.

Tenants provide management and control isolation at an organizational level but are bounded by policies set forth by the system-level administrator. A tenant typically represents an actual organization such as a company or a department within a company that uses a portion of a repository. A tenant can also correspond to an individual person. An HCP can have many HCP tenants, each of which can own and manage many namespaces.

Figure 8. HCP Capabilities Manage Exabytes of Capacity



Tenant-level administration: There is a separate administrator for each tenant. They create and manage namespaces for application use at a micro level. They control namespace capacity through quotas, define user membership, access protocols and service policies. They further define which users can read, write, delete or search a namespace. The HCP system-level administrator controls the number of namespaces each HCP tenant can create.

A **namespace** is the smallest unit of HCP multitenancy capacity partitioning. Namespaces are thin provisioned and carved from the common virtualized storage pool. Namespaces provide the mechanism for separating the data stored by different applications, business units or customers. Access to one namespace does not grant a user access to any other namespace. Objects stored in one namespace are not visible in any other namespace. Namespaces provide segregation of data, while tenants provide segregation of management.

Applications access HCP namespaces through HCP REST, S3, WebDAV, CIFS (SMB v3.1.1), NFS v3 and SMTP protocols. These protocols can support authenticated and/or anonymous types of access. When applications write a file, HCP conceptually puts it in an **object** container along with associated metadata that describes the data. Although HCP is designed for WORM access of information, namespaces can be enabled with versioning to permit write and rewrite I/O semantic (see software overview).

Management API and Programmatic Monitoring

The management API is a RESTful HTTP interface to administrative functions of an HCP system. Using this API, administrators can manage tenants, namespaces, downstream DNS settings for networks, retention classes, search policy (content classes), system-level user account passwords, tenant-level user and group accounts, replication, erasure coding, HCP licenses and HCP internal logs.

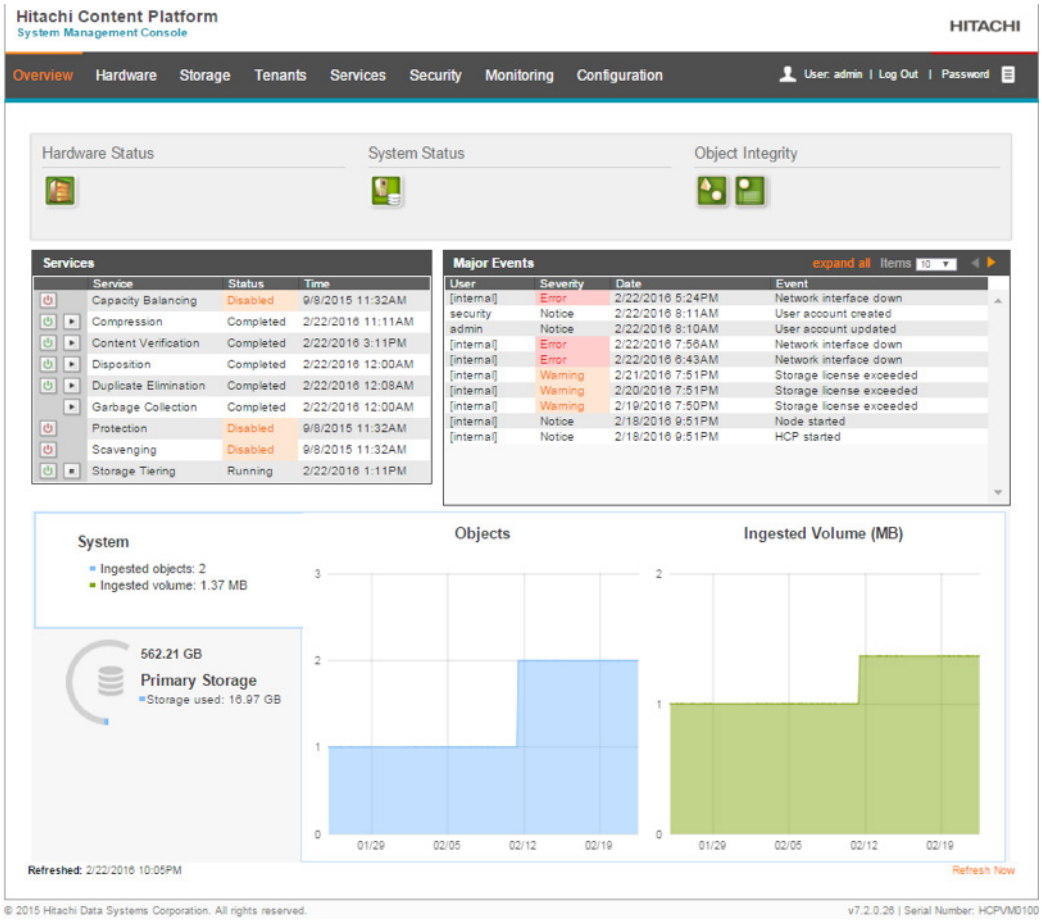
HCP provides a variety of methods to monitor the system, including: System logging (syslog), email notifications, SNMP and monitoring API calls are available. Through these methods, it is possible to programmatically monitor system performance, health, capacity and background services. Customers can also use the optionally available Hitachi Content Monitor to enhanced perform analysis and forecasting.

HCP integrates into the Hitachi Remote Ops (formerly Hi-Track) monitoring system to provide a dashboard of system health and status and enable proactive action by Hitachi Vantara Customer Support.

System-Level Dashboards and Notifications

With the web-based overview dashboard, the system-level administrator can quickly assess HCP cluster status (see Figure 9). The single pane summary displays color-coded health alerts, data services, major events and the total capacity consumed by all tenants and namespaces. Use one-click drill down into any of the 500+ alerts or events and electively choose to enable email notifications, SNMP or system logging (syslog).

Figure 9. HCP Web-Based Overview Dashboard



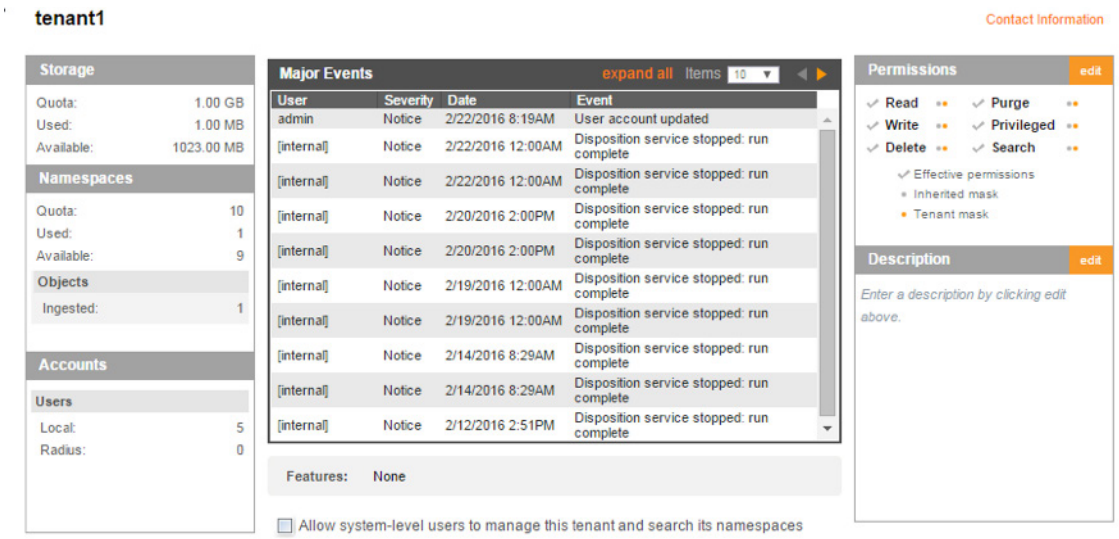
Wizards: HCP provides guided input templates to streamline virtually all major configuration tasks, notably for tenant creation and designing namespace service plans that control data placement and object life cycles.

Scheduling: Use a 7-day, 24-hour calendar to ensure post-process services, such as tiering, deduplication, compression and verification run outside of your peak workload periods.

Tenant-Level Dashboards and Notifications

The overview dashboard for a tenant administrator displays a summary of events, and the sum total capacity consumed by all its defined namespaces (see Figure 10). The panel provides one-click drill down into any events, which are also forwarded to an email address.

Figure 10. HCP Overview Dashboard for Tenant Administrator



Namespace configuration templates: Each tenant administrator is delegated with authority over an allotted capacity. These templates help them create namespaces, configure permitted protocols, set capacity quotas and policies for retention, disposition, indexing and search. Optionally, configuration can be carried out through REST API or Microsoft PowerShell utilities.

Enterprise mode: The tenant administrator is always permitted to create namespaces with an enterprise retention policy. While normal users cannot delete objects under enterprise retention, a tenant-admin can be empowered to preform audit-logged privileged deletes.

Compliance mode: The tenant administrator can be permitted to create namespaces with a compliance retention policy. Objects under compliance retention cannot be deleted through any user or administrative action until their expiry date. Industry-specific regulations sometimes mandate immutable compliance modes to protect electronic business records. Utilize this mode with prudence since even experimenting can create permanent undeletable content.

Chargeback Reporting

Data consumption reports (see Figure 11) are available to both system and tenant administrators as an onscreen display or as a download. At the system level, the report will include a rollup of all tenants and their namespaces, while individual tenant administrators will receive a report limited to the namespace(s) they own.

Figure 11. HCP Data Consumption Report

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	systemName	tenantNa	namespac	startTime	endTime	objectCou	ingestedV	storageCa	bytesIn	bytesOut	reads	writes	deletes	deleted	valid
2	hcp.hcp-demo.com	splunk	folder	2/11/2016 21:25	2/11/2016 23:59	1	384686	385024	384686	0	0	1	0	FALSE	FALSE
3	hcp.hcp-demo.com	splunk	folder	2/12/2016 0:00	2/12/2016 23:59	1	384686	385024	0	0	0	0	0	FALSE	FALSE

Object Storage Software Architecture

Hitachi Content Platform is an object storage platform. Its architecture by nature means it is more efficient, easier to use, and capable of handling much more data than traditional file storage solutions. HCP automates day-to-day IT operations and can readily evolve to changes in scale, scope, applications, storage, server and cloud technologies over the life of data. In IT environments where data grows quickly or must live for years, decades or even indefinitely, these capabilities are invaluable.

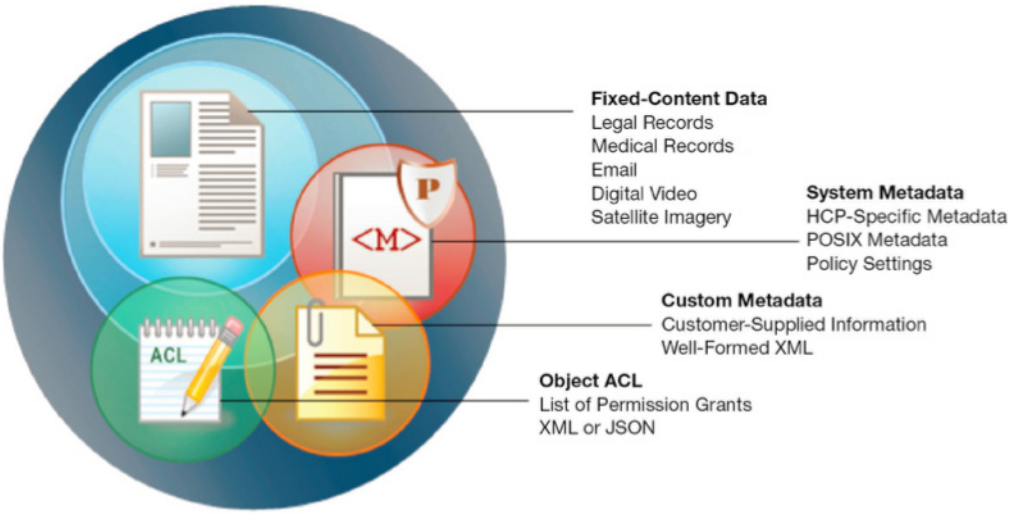
Hitachi Content Platform eliminates the need for a siloed approach to storing unstructured content. HCP software receives unstructured data files via file or REST protocols and stores them as objects. All objects in the repository are distributed across all available storage pools according to policies assigned to the namespace. Externally, HCP presents each object either as a set of files in a standard directory structure or as a uniform resource locator (URL) accessible by users and applications via HTTP or HTTPS. In all cases, the software retains any file directory structure applied by protocols. Once ingested, the software offers a variety of services and tools to protect object integrity, manage object life cycles, search it, and ensure it is always available.

Object Container Structure

An HCP object is composed of fixed-content data (a user's file) and electronic "sticky notes" called metadata. Metadata describes the fixed-content data, including its properties. All the metadata for an object is viewable, but only some of it can be user-modified. The way metadata can be viewed and modified depends on the namespace configuration, the data access protocol and the type of metadata. HCP metadata types include system metadata and, optionally, custom metadata and access control list (ACL). The structure of the object is shown in Figure 12.

Fixed-content data is an exact digital copy of a written file which is "fingerprinted" upon ingest using a hashing algorithm: MD5, SHA-1, SHA-256 (default), SHA384, SHA-512 or RIPMD160. These files become immutable after being successfully stored in a virtual storage pool. If the object is under retention, it cannot be deleted before the expiration of its retention period (see compliance modes). If versioning is enabled, multiple versions of a file can be retained.

Figure 12. HCP Object



System metadata is composed of 28 properties that include the date and time the object was added to the namespace (ingest time), the date and time the object was last changed (change time), the cryptographic hash value of the object along with the namespace hash algorithm used to generate that value, and the protocol through which the object was ingested. It also includes the object's policy settings, such as the number of redundant copies, retention, shredding, indexing and versioning. POSIX metadata includes a user ID and group ID, a POSIX permissions value and POSIX time attributes.

Custom metadata is optional, user-supplied descriptive information about a data object that is usually provided as well-formed XML. It is utilized to add more descriptive details about the object. This metadata can be utilized by future users and applications to understand and repurpose the object content. HCP supports multiple custom metadata fields for each object.

THE IMPORTANCE OF METADATA

Custom metadata brings structure to unstructured content. It enables building of massive unstructured data stores by providing means for faster and more accurate access of content. Custom metadata gives storage managers the meaningful information they need to efficiently and intelligently process data and apply the right object policies to meet all business, compliance and protection requirements. Structured custom metadata (content properties) and multiple custom metadata annotations take this capability to the next level by helping to yield better analytic results and facilitating content sharing among applications. Custom metadata can be divided into multiple partitions, so that multiple users and applications can work with the same object without impacting each other's metadata.

In many cases, the metadata is more valuable than the object itself. An individual X-ray is not that useful beyond the doctor and the patient. But when that image is stored alongside thousands of others, all with well-described metadata, trends can be discovered, connections made and insights revealed (see Search section).

Object ACL is optional, user-supplied metadata containing a set of permissions granted to users or user groups to perform operations on an object. ACLs control data access at an individual object level and are the most granular data access mechanism.

Store Objects

HCP access nodes share responsibility for knowing where content is stored. HCP stores fixed content file data separately from its metadata, placing them in separate parallel data structures. For scaling purposes, HCP nodes also maintain a hash index, which is a sharded database that is distributed among all HCP access nodes. The **hash index** provides the content addressable lookup function to find data. Each node is responsible for tracking a subset of the index called a region, which tells it where to find data and metadata.

Upon receiving a new file, any receiving node is able to write the fixed content file portion to storage it owns, as directed by the assigned service plan. It then computes a hash of the pathname, adds it to the object's system metadata along with the object's location, and forwards it to the node responsible for tracking the hash index region. HCP protects its index with metadata protection level of 2 **(MDPL2)**, which means it will store two copies, saved on different nodes. There is one authoritative copy, and at least one backup copy. A write is not considered complete until all MDPL copies are saved. The actual file is stored in a storage pool defined by the tenant administrator. Storage pools can be constructed with disks inside an HCP access node, HCP S series or SAN storage disks (see Hardware Overview).

Read Objects

Upon receiving a read request file, the HCP node computes a hash using the objects pathname. If it manages the particular hash index region, it can look up the object's location and fulfill the request. If it does not manage the hash region it can query the owner node for the files location (see Networking). In the case of a node failure, it can query the node with the backup hash index. In the case of a whole site failure, DNS can redirect the request to any surviving cluster participating in namespace replication.

Open Protocols

While HCP supports 100% of the S3 API operations needed for CRUD programming (create, read, update and delete), many new cloud applications being developed still favor its native REST protocol for HCP. This protocol

provides capabilities that S3 does not, providing insight into HCP's physical infrastructure, its powerful retention, multitenancy, search and metadata capabilities. HCP also offers proprietary headers for certain S3 API commands so that applications can easily utilize the HCP's compliance features that are not part of the S3 API specification.

For those not quite ready to shed their old file access methods, HCP supports four legacy protocols that include **NFS v3, CIFS (SMB v3.1.1), SMTP** and **WebDAV**. Anything written with these APIs can also be accessed with any of the REST API, with directories and filenames intact. Alternatively, it may be more appropriate to use a separate cloud gateway to provide file services, such as HCP Anywhere Edge or HCP Gateway.

Over 100 ISVs have applications compatible with HCP cloud storage, and enjoy access to Hitachi partner programs, where they can download HCP evaluation software and participate in forum discussions.

HCP Data Services

HCP software implements background services (see Table 1) that work to improve the overall health of the HCP system, optimize efficiency and maintain the integrity and availability of stored object data. Services can run either continuously, periodically (on a specific schedule), or in response to certain events. The system-level administrator can enable, disable, start or stop any service and control the priority or schedule of each service. These controls include running a service for longer periods, running it alone or assigning it a higher priority. Control runtime system loading by limiting the number of threads that the service can spawn, using simple high, medium and low designations.

All scheduled services run concurrently but autonomously to each other, and thus each service may be simultaneously working on different regions of the metadata database. Each service iterates over stored content and eventually examines the metadata of every stored object. On a new HCP system, each service is scheduled to run on certain days during certain hours. If a particular service completes a full scan in the allotted period, the service stops. If it does not finish, the service resumes where it left off at its next scheduled time slot. After completing a scheduled scan interval, the service posts a summary message in the HCP system event log.

Table 1. HCP Background Services

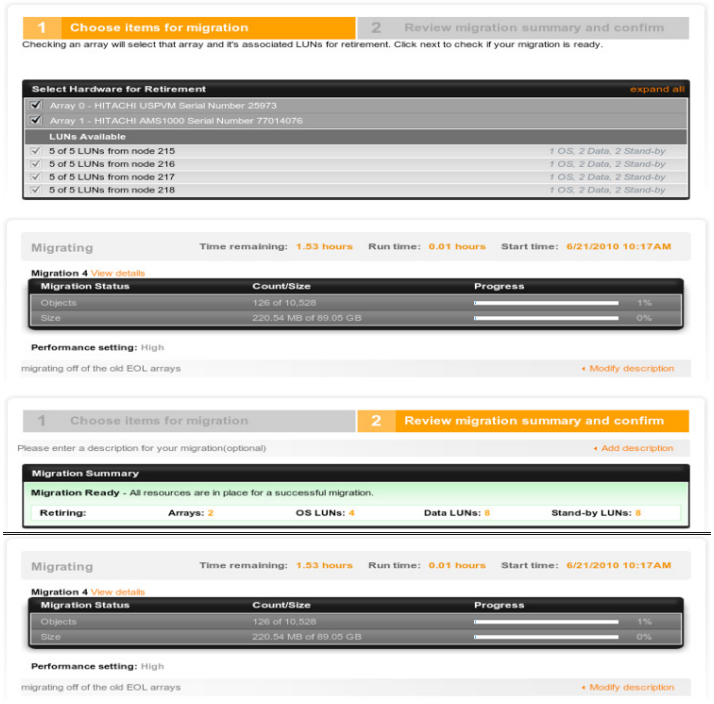
SERVICE	DESCRIPTION
Capacity Balancing	Attempts to keep the usable primary storage capacity balanced (roughly equivalent) across all access nodes in the system. If storage utilization for the nodes differs by a wide margin, the service moves objects around to bring the nodes closer to a balanced state.
Compression	Compresses object data to make more efficient use of physical storage space.
Content Verification	Guarantees data integrity of repository objects by ensuring that a file matches its digital hash signature. HCP repairs the object if the hash does not match. Also detects and repairs metadata discrepancies.
Deduplication	Identifies and eliminates redundant objects in the repository, and merges duplicate data to free space.
Disposition	Automatic cleanup of expired objects. A namespace configuration policy authorizes HCP to automatically delete objects after their retention period expires.
Fast Object Recovery	Ensures that unavailable objects have their status changed to available once they are recovered. Runs in response to an availability event, such as an unavailable node becoming available.
Garbage Collection	Reclaims storage space by purging hidden data and metadata for objects marked for deletion, or data left behind by incomplete transactions (for example, unclosed NFS or CIFS files). May perform disposition as well.
Scavenging	Ensures that all objects in the repository have valid metadata, and reconstructs metadata in case the metadata is lost or corrupted.

S Series Balancing	Attempts to keep the HCP S series storage system capacity utilization percentage balanced (roughly equivalent) within each S series node pool. If the imbalance exceeds a threshold, the service moves objects from the most full S series system to the least full S series system.
Migration	Migrates data off selected nodes or Hitachi storage arrays so they can be retired.
Protection	Enforces data protection level (DPL) policy compliance to ensure the proper number of copies of each object exists in the system.
Replication	Copies one or more tenants from one HCP system to another to ensure data availability and enable disaster recovery.
Replication Verification	Identifies differences and attempts to resolve differences in objects between replicated sites. This is not a scheduled service; it either runs during replication or on demand.
Shredding	Overwrites storage locations where copies of the deleted object were stored in such a way that none of its data or metadata can be reconstructed, for security reasons. Also called secure deletion. The default HCP shredding algorithm uses three passes to overwrite an object, which satisfies the DoD 5220.22-M standard (3 passes). This service is triggered by the Garbage Collection service.
Storage Tiering	Determines which storage tiering strategy applies to an object; evaluates where the copies of the object should reside based on the rules in the applied service plan.
Geodistributed Erasure Coding (Geo-EC)	Geodistributed erasure coding can be applied when HCP spans three or more sites. This technology provides 100% data availability despite whole site level outages. Geo-EC deployments consume 25-40% less storage than systems deployed with simple mirror replication.

Autonomic Tech Refresh (ATR)

Autonomic tech refresh embodies the vision that software and data will outlive the hardware hosting it (see Figure 13). This built-in migration service enables organizations to move a “live” HCP onto new hardware, replace old servers, or siphon content from old storage as a rate-controlled background process and write it to new storage. With ATR, there are no disruptions to customer business as applications continue to run normally. This forethought to maintenance is rather unique and a testament to HCP’s long-term product support strategy.

Figure 13. Autonomic Tech Refresh: Choose, Review and Confirm



HCP Replication Topologies and Content Fencing

Replication Overview

Replication copies one or more tenants from one HCP system to another to ensure data availability and enable disaster recovery. The replication process is object-based and asynchronous. The HCP system in which the objects are initially created is called the primary system. The second system is called the replica. Typically, the primary system and the replica are in separate geographic locations and connected by a high-speed wide area network. The replication service copies one or more tenants or namespaces from one HCP system to another, propagating object creations, object deletions and metadata changes. HCP also replicates tenant and namespace configurations, tenant-level user accounts, compliance and tenant log messages, and retention classes.

HCP offers multisite replication technology called global access topology. With these bidirectional, active-active replication links, globally distributed HCP systems are synchronized in a way that allows users and applications to access data from the closest HCP site for improved collaboration, performance and availability.

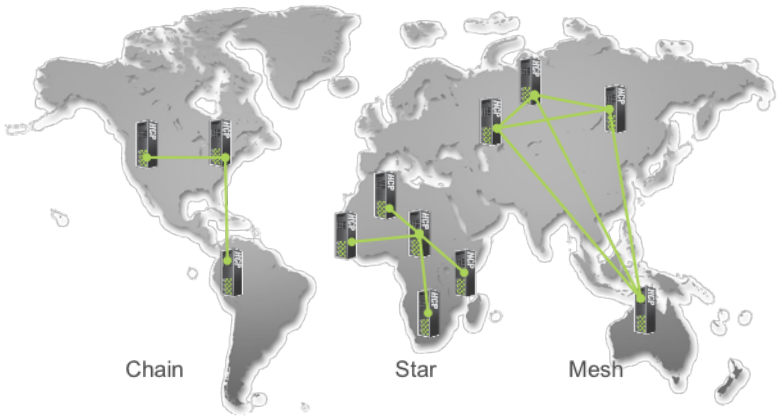
Metadata-Only Replication

Choose to replicate entire objects or just object metadata. A metadata-only strategy allows all clusters to know about all objects, but it controls placing object payload only where needed while saving on WAN costs.

Content Fencing

One practical use case for metadata-only replication is to create data fences, which allow organizations to share data but ensure it stays hosted within a specific country or continent boundary. In this model, HCP replicates metadata but withholds mass movement of data files. Applications at the remote end are able to see files and directory structures, search metadata fields and even write content. In all cases, the final permanent resting place for the object is at the source. Global access topology supports flexible replication topologies that include chain, star and mesh configurations (see Figure 14).

Figure 14. Chain, Star and Mesh Configurations



"Replication Before Tiering" for Economical Regulatory Compliance

Some organizations must maintain two copies of data for regulatory compliance reasons and want to keep one copy at two sites for resiliency. This is a challenge for customers because of the cost of high bandwidth and low latency that would be necessary to synchronously replicate data over a WAN. HCP provides an elegant solution for efficiently maintaining compliance while achieving the resiliency goal.

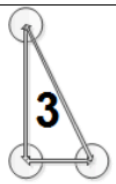
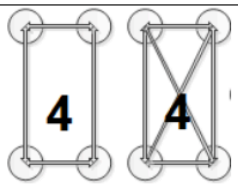
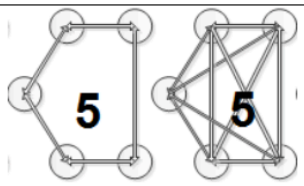
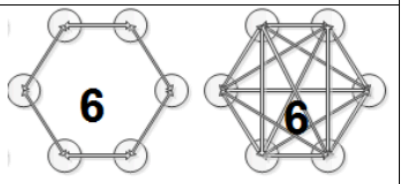
When an object is stored on HCP, it is possible to use a Data Protection Level (DPL) of two: That is, two copies are immediately created at the local cluster to enable and maintain regulatory compliance. HCP can then create another copy at a replica HCP cluster using asynchronous replication, which creates a third copy globally. However, storing three copies is not necessary as only two are required. Therefore, when the replication process completes, HCP's "replication before tiering" feature can reduce the DPL of the original cluster to one, so that there is only one copy stored at the original site and one copy at the

replica site. This capability maintains regulatory compliance throughout the process while avoiding the network costs of synchronous replication as well as costs associated with unnecessary capacity consumption.

Geodistributed Erasure Coding

With replication links established between three or more sites, HCP software offers a geodistributed erasure code service for greater storage efficiencies and cost savings (see Figure 15). The service operates on objects to provide site-level disaster recovery using fewer bytes than a mirrored replica. The service is electively applied to a namespace (bucket) along with settings for activation. Administrators decide when objects should transition to a geodistributed state. New objects can remain as whole objects at all sites for a brief period, providing fast, low latency readback. When this period ends, the object transitions to an efficient geo-EC state that can reduce storage consumption by up to 40%, and keep objects eligible for compression and deduplication.

Figure 15. Cost Savings and Efficiencies From Geodistributed Erasure Code Service

Number of Sites	2	3	4	5	6
Geo-EC Topology					
Capacity Savings	0%	25%	33%	37%	40%

Search

With Hitachi Content Platform, you have access to metadata and content search tools that enable more elegant and automated queries for faster, more accurate results. Through these features you can gain a better understanding of the content of stored files, how content is used and how objects may be related to one another. This understanding can help you to enable more intelligent automation, along with big data analytics based on a best-in-class metadata architecture.

HCP software includes comprehensive built-in search capabilities that enable users to search for objects in namespaces, analyze a namespace based on metadata, and manipulate groups of objects to support e-discovery for audits and litigation. The search engine (Apache Lucene) executes on HCP access nodes and can be enabled at both the tenant and namespace levels. HCP supports two search facilities:

1. A web-based user interface called the search console provides an interactive interface to create and execute search queries with “AND” and “OR” logic. Templates with dropdown input fields prompt users for various selection criteria such as objects stored before a certain date or larger than a specified size. Clickable query results are displayed on screen. From the search console, search users can open objects, perform bulk operations on objects (hold, release, delete, purge, privileged delete and purge, change owner, set ACL), and export search results in standard file formats for use as input to other applications.
2. The metadata query API enables REST clients to search HCP programmatically. As with the search console, the response to a query is metadata for the objects that meet the query criteria, in XML or JSON format.

In either case, two types of queries are supported:

- An object-based query locates objects that currently exist in the repository based on their metadata, including system metadata, custom metadata and ACLs, as well as object location (namespace or directory). Multiple, robust metadata criteria can be specified in object-based queries. Objects must be indexed to support this type of query.

- An operation-based query provides time-based retrieval of object transactions. It searches for objects based on operations performed on the objects during specified time periods. And it retrieves records of object creation, deletion and purge (user-initiated actions), and disposition and pruning (system-initiated actions). Operation-based queries return not only objects currently in the repository but also deleted, disposed, purged or pruned objects.

Multiple Metadata Annotations

Each HCP object supports up to 10 free-form XML metadata annotations, up to 1GB total. This gives separate teams freedom to work and search independently. An analytics team may add annotations specific to their applications, which are different from the billings applications. XML annotation can provide significant advantages over simple key value pairs because the search engine can return more relevant results with XML. Consider

Table 2. Metadata Annotation Example

This example XML record represents a single annotation	Key-value pair
<Record> <Dr>John Smith</Dr> <Patient>John Smith</Patient> <Address>St John Smith Square</Address> </Record>	Dr=John Smith Patient=John Smith Address=St John Smith Square</Address>

Now imagine a search where you want the objects related to doctor named “John Smith.” The XML record allows you to pinpoint search results to this field, whereas the key value pair will produce a much larger set of search hits. As object count grows to millions and billions, key-value searches can quickly become slow and arduous.

Hardware Overview

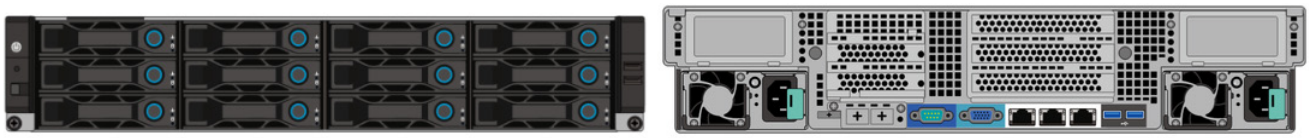
HCP software is deployed on commodity x86 hardware or hypervisors as a distributed storage system. From a hardware perspective, each HCP system consists of the following categories of components:

- HCP G series access nodes (servers) or virtual machine instances.
- Networking components (switches and cabling).
- Infrastructure components (racks and power distribution units).
- Physical storage pools, which may include:
  - Access node internal storage.
  - HCP S series storage (erasure-coded, Ethernet-attached storage).
  - SAN-attached storage.
  - VMware (SAN, RDM, vSAN, NFS datastores)
  - Extended or public cloud storage.

Access Nodes (HCP G Nodes)

If delivered as an appliance, the nodes are constructed from conventional x86 off-the-shelf 2U servers called an HCP G series node (see Figure 16). Each node is configured with multicore CPUs, DRAM, 10Gb networking and up to 12 internal drives for object data protected with RAID-6, and optionally two solid state drives (SSDs) for database acceleration.

Figure 16. HCP G Nodes



The most minimal HCP system will have four access nodes with local disk storage.

- Electively, HCP G series nodes can be purchased with Fibre Channel adapters to utilize Hitachi SAN arrays for bulk HCP object storage.
- Upgrade DRAM, SSD or internal disks anytime, as server space permits.
- Access nodes can be deployed as virtual appliances (see the Virtual Access Nodes section below).

All-Flash Configuration

HCP G nodes can be populated with SSDs for applications that call for higher performance. The SSDs can improve performance for object data, HCP system background processes, and searching.

The SSDs can be used as a flash tier for object data, which provides lower latency and higher throughput. Administrators can specify certain applications (that is, namespaces) to use the SSD tier as storage, and other applications to use lower-cost storage options. Policies can be defined to move the objects from the flash drives to lower-cost storage devices, like HCP S series, which frees up capacity on the SSDs. Policies can also be used to move objects from other storage tiers to the flash tier in case higher performance is desired.

The performance of HCP system processes is also improved in an all-flash configuration. The background services are able to process more objects in less time and provide consistent performance when there is a heavy I/O load.

The metadata searching capability of HCP operates with higher performance when running in an all-flash configuration, leading to faster query results and faster indexing of object metadata.

Flash-Optimized Acceleration Option

To improve the performance and increase the maximum number of objects, hard-drive-based HCP G series nodes may be populated with a pair of SSD, configured in a RAID-1 mirror. These drives are not used as a cache, and do not contain user object data. Rather, they are used for the singular purpose of accelerating database performance related to operating HCP's content-addressable hash index. The judicious and focused use of SSDs is known to improve read and write performance, especially when the number of objects managed per node grows large (>100M objects).

Virtual Access Nodes

This deployment model lets administrators install virtual instances of HCP access nodes on hardware they supply. HCP access nodes (both virtual and physical) can run with as little as 1TB of licensed storage. At present, HCP supports running within KVM or VMware hypervisors. Best practice installations would ensure that virtual machines (VMs) reside on separate bare metal nodes to ensure the cluster operates with maximum availability. Capacity expansion can be accomplished with HCP S series storage nodes or customer-owned Fibre Channel storage.

Customers can use HCP virtual access nodes to take advantage of the VMware vSAN software-defined storage platform by using hyperconverged (HCI) hardware, such as Hitachi Unified Compute Platform (UCP).

HCP S Series Storage

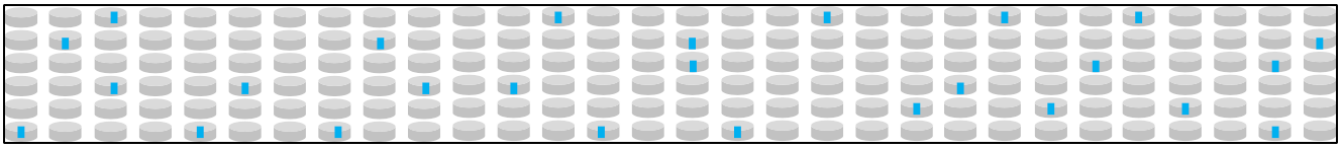
Ethernet attached storage (HCP S series) provide erasure code (EC) protected storage pools that are managed by HCP access nodes. These appliances allow an HCP to scale storage capacity independently from its compute capacity. All HCP S series systems feature two independent active-active x86 controllers, with dual path access to all SAS enterprise disk drives in attached trays. The control and data path between access nodes and the HCP S series storage occurs over bonded and redundant 10Gb Ethernet connections (up to eight ports, in total). Each of the two

HCP S series controllers provide erasure coded data protection as well as load balancing capabilities that enable seamless, automated adaptability as the storage volume increases:

- **HCP S11:** One 4U tray supporting up to 94 disks and 6 SSD, with two integrated x86 single CPU controllers; expandable with one 4U enclosure for an additional 106 disks. Disk configurations range from 300TB (230TB usable) up to 3,200TB (2,461TB usable).
- **HCP S31:** One 4U tray with up to 94 disks and 6 SSD, with two integrated x86 dual CPU controllers; expandable with up to eight 4U enclosures for an additional 848 disks. Disk configurations range from 300TB (230TB usable) up to 15,072TB (11,593TB usable).

Protection: Reed-Solomon 20+6 Erasure Coding (see Figure 17). Immediately following insertion into the storage pool, new disks are divided into 64MB chunks called extents. The extents are grouped into an EC-protected extent group by choosing and then writing to 26 extents, each on a different disk. As objects are ingested, they are efficiently stored within these extent groups. The extent group forming constraint guarantees that any disk loss will only affect at most one extent from a particular extent group. With 20+6 erasure coding, objects within an extent group remain readable despite the simultaneous loss of six drives containing the data of that extent group.

Figure 17. HCP S Series System With 90 drives: Sample EC Extent Group Allocation



Common HCP S series system properties:

- On-demand initialization (no long formatting).
- Self-optimizing algorithms for extent group repair and auto rebalance.
- Rebuild duration is a function of “bytes used” rather than “bytes available.” For example, if a 6TB disk fails and it contained 1TB of written extents, the rebuild algorithm needs to rewrite only 1TB of extents.
- No “hot spares.” Extent groups have flexibility in terms of placement, thus all drives are used all the time. When a drive fails, erasure coding recreates just the missing data using the free capacity on remaining drives. It therefore returns to optimal status without requiring the addition of new disks.
- Extend service intervals. Because failed disks no longer imply a degraded platform, it’s possible to defer the disk replacement service until the number of disks with free capacity runs low. The system automatically notifies the administrator when that point is approaching.
- Unlike RAID, rebuild I/O is not throttled by the write capabilities of one drive. When a drive fails, the S series software repairs and redistributes erasure-coded extents to all active drives.
- Storage efficiency and rebuilds are file-size agnostic because HCP S series utilizes 64MB extents. These systems can store literally hundreds of small files in a single extent. Thus, one write I/O to repair one extent member (64MB chunk) effectively repairs hundreds of files. In this way, HCP S series systems are optimized for both large and small files.

Extended Storage (NFS and Public Cloud)

With **adaptive cloud tiering (ACT)** technology, HCP access nodes can leverage storage from third-party sources to construct storage pools built with:

- On-site NFS or S3 compatible storage.
- Off-site public cloud storage sources, including Amazon S3, Microsoft Azure, and Google Cloud Storage.

ACT makes it possible to construct hybrid and multicloud HCP configurations that share resources between public and private clouds. With simple edits of an existing service plan, the system-level administrator can easily shift cluster composition:

- Easily scale HCP storage capacity up or down as needed. With elastic capacity, make room on premises for a sudden capacity need by encrypting and tiering older objects to third-party public cloud storage services.
- Gain cloud broker capabilities. For example, easily switch public cloud storage allegiances towards the best long-term storage rates, or better reliability.
- Encrypt content that you're not quite ready to part with and direct it to a public cloud storage service(s) for the remainder of its retention period.

By waiting for replication to complete before triggering tiering to a different DPL, customers can achieve higher data durability with a minimal increase in consumed capacity.

Service plan adjustments are transparent to client applications from a protocol perspective. However, policy changes may temporarily increase the total I/O workload if the changes trigger a background data migration. The impact can be mitigated by setting tiering-related priority levels.

Networking

All HCP systems feature at least two physically separate Ethernet networks referenced as the private back-end network and the public front-end network. The front-end and back-end networks are 10Gb capable, and are constructed with a bonded port pair, which connect to independent switches. Administrators may segregate system management traffic to a third, physically separate network.

Private back-end network: The two switches comprising the isolated back-end network carry traffic vital to internode communication; they are provided as part of an HCP appliance and do not provide outside access. To ensure maximum network reliability and availability, the network design requires two unstacked switches with options for optical or copper media. While a choice of 1Gb or 10Gb switches is offered, a pair of 1Gb switches provides ample communication bandwidth for many configurations.

Front-end network: This VLAN-capable network connects to a customer-supplied switch infrastructure. This network carries application read/write traffic, management traffic and HCP S node traffic (if present) using VLANs. The recommended front-end setup would include two independent switches that support 1Gb or 10Gb Copper (RJ45), or optical 10Gb SFP+ connections.

External communication with HCP is typically managed via DNS, which round-robins client requests across all nodes to ensure maximum system throughput and availability. With DNS, clients reference a domain name rather than a specific node or IP address. Oftentimes, a subdomain or delegate is defined within the corporate DNS and all nodes in the HCP cluster are listed as hosts. HCP uses load balancing internally and manages all inbound read/write requests, passing them to the appropriate node for execution. This network can be configured in native IPv4, native IPv6, or dual IPv4 and IPv6 modes, where each virtual network will support either or both IP versions. IPv6 is mandated by many government agencies, and it is necessary to support very large-scale networks.

Management network (optional): If enabled, a management network can segregate system administration, tenant administration, management API, SNMP, syslog, outgoing SMTP and SSH traffic from the front-end and back-end networks. Each physical access node (HCP G series node) includes a port that may be exclusively used by the management network. A virtual access node can use a virtual NIC exclusively created for the management network.

VLANs (virtual LANs): HCP supports virtual networking for the front-end network through which clients communicate with the system and different HCP systems communicate with each other. HCP allows virtual network separation of replication, the storage layer, data access and system management. Typical use of a virtual network would include one VLAN for replication data, and one VLAN for traffic between HCP access nodes and storage nodes. And for each tenant there would be a VLAN for data access and a VLAN for management access.

Configurations Using Only HCP G Nodes

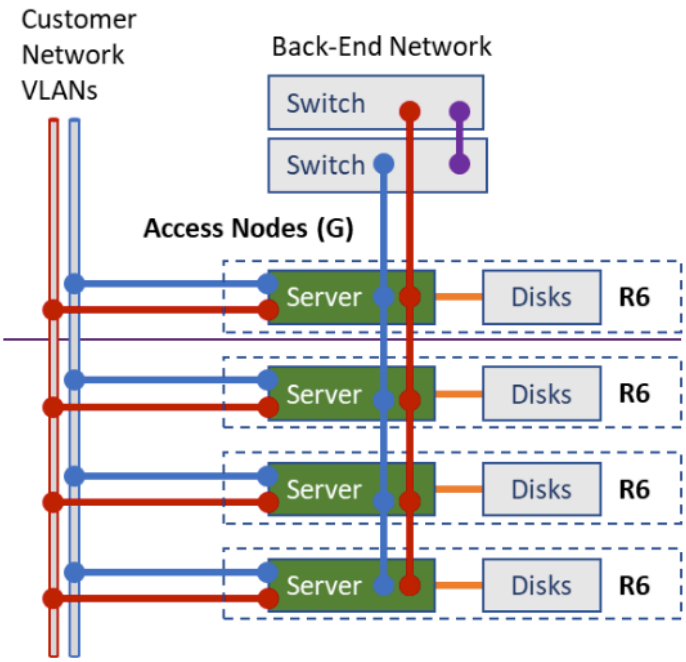
HCP systems with server-only disk configurations use their internal disks to store files, metadata and the appliance software (see Figure 18). A minimal cluster consists of at least four HCP G series nodes populated with at least six disks to form one

RAID-6 group (4D+2P). At any time, the server capacity can be expanded with a second set of six disks to create another RAID-6 group. All-flash HCP G series nodes use 12 SSDs in a 10D+2P configuration.

In a RAID-6 based configuration, the cluster is normally operating with a data protection level of 2 (DPL2), which means each file object (and associated custom metadata) is stored twice using separate protection sets on different nodes. Note: DPL is distinct and unrelated to MDPL, which controls the number of system-metadata copies. By default, the MDPL is also set at level 2. All of this ensures the cluster will remain fully functional in the event of a node loss.

For example, a single site, four-node HCP G11 series node cluster with 12 x 4TB hard drives provides 56TB of usable capacity when storing objects with DPL2, and 112TB if storing with DPL1 (DPL1 configurations using only HCP G series nodes should only be considered if deploying HCP across two sites using HCP replication). The HCP licensing policy permits operation with as little as 4TB. Later, the capacity license can be increased to accommodate new capacity needs. A license can also exceed the capacity of physically attached storage.

Figure 18. Configuration Using HCP G Series Nodes (Access Nodes)



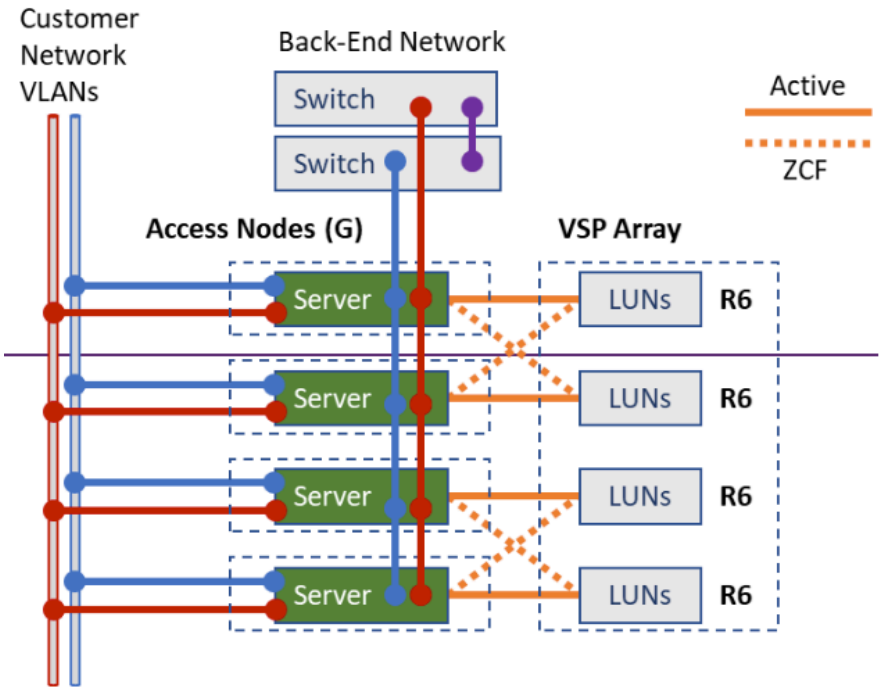
The capacity of HCP clusters configured with server-only storage can be expanded in two-node increments. Scaling capacity in this fashion may needlessly increase the compute capability of the cluster, since each increment of storage pulls in another server pair with two CPUs between them.

Situations that require significant capacity scaling should consider HCP S series systems because they almost always deliver better economics in the form of lower cost per gigabyte.

HCP SAN-Attached Configurations

An HCP system with SAN-attached storage uses Fibre Channel arrays from Hitachi to store fixed-content file objects and custom metadata, but saves portions of system metadata on its internal disks (see Figure 19). A minimal cluster consists of at least four HCP G series nodes, each with a two-port Fibre Channel card and each with six disks to form one RAID-6 group (4D+2P). The internal disks are used exclusively for the appliance operating system and HCP's system metadata database.

Figure 19. HCP SAN-Attached Configurations



Fibre Channel arrays can supply each access node with up to 1PB of storage capacity, providing much better storage density than server-only storage configurations. Moreover, these configurations can be operated with a data protection level of 1 **(DPL1)** while maintaining high data availability based on the shared storage capabilities of a SAN. This is feasible because HCP software employs a feature called zero copy failover (ZCF). It works by pairing two nodes that are enabled with multipathing such that they can see and access one another's Fibre Channel storage LUNs, which is a process called cross-mapping. When the cluster is optimally configured, the LUNs managed by a node during normal operation are considered primary LUNs. The LUNs visible on the other node are considered standby LUNs. In the event of a node failure, cross-mapping of LUNs allows a surviving node to assume ownership of the standby LUNs. This shared storage configuration enables capacity scaling without increasing access node count. If scaling access nodes, they must be added in groups of two to facilitate ZCF pairing.

MDPL2 along with ZCF ensure the data stored in the cluster will remain fully functional in the event of a node loss.

Often, these configurations will not require network fabric because many Hitachi arrays provide enough ports to directly connect up to 16 access nodes. Configurations exceeding these limits may be purchased with either Brocade or Cisco Switches (not shown in the diagram).

It merits mention that HCP S series systems can also be used for capacity expansion, and offer very similar scaling and failover advantages of Fibre Channel.

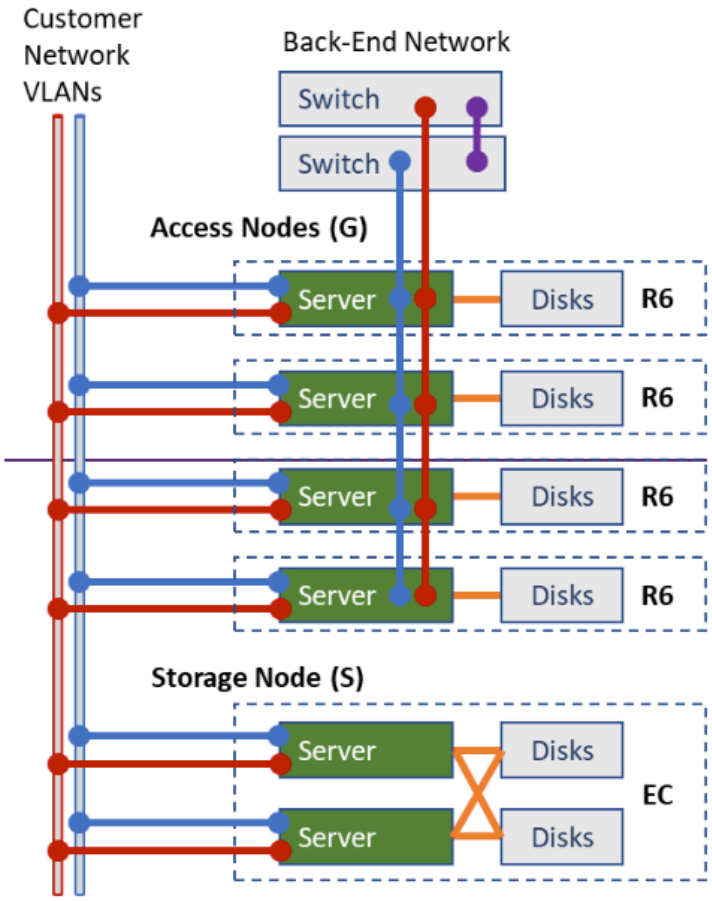
Capacity Scaling With HCP S Node

Ethernet-attached HCP S series systems provide capacity expansion options for any HCP configuration (see Figure 20). These appliances combine enterprise HCP software and commodity hardware to achieve greater scale at a lower cost. HCP S series systems easily rival and sometimes exceed the storage density offered by Fibre Channel arrays. Moreover, as an Ethernet-connected storage resource, they offer shared access properties similar to a SAN. As such, these configurations can be safely operated with a data protection level of 1 **(DPL1)**. To ensure cluster reliability, these clusters use MDPL2. MDPL2 along with HCP S series storage pools ensure the cluster will remain available and fully functional despite the loss of an access node.

In the example below, a single HCP S series storage system is added to the front-end network using HTTPS and VLANs. HCP G series access nodes then register the HCP S series system as a storage component and add it to the storage pool. In this configuration, HCP access nodes use their internal disks to store metadata, while file objects are

stored on the HCP S series storage. The additional capacity from the S series system is then available to all nodes for new REST writes, or as a tiering target for existing objects stored on internal server disks.

Figure 20. Capacity Scaling With HCP S Series

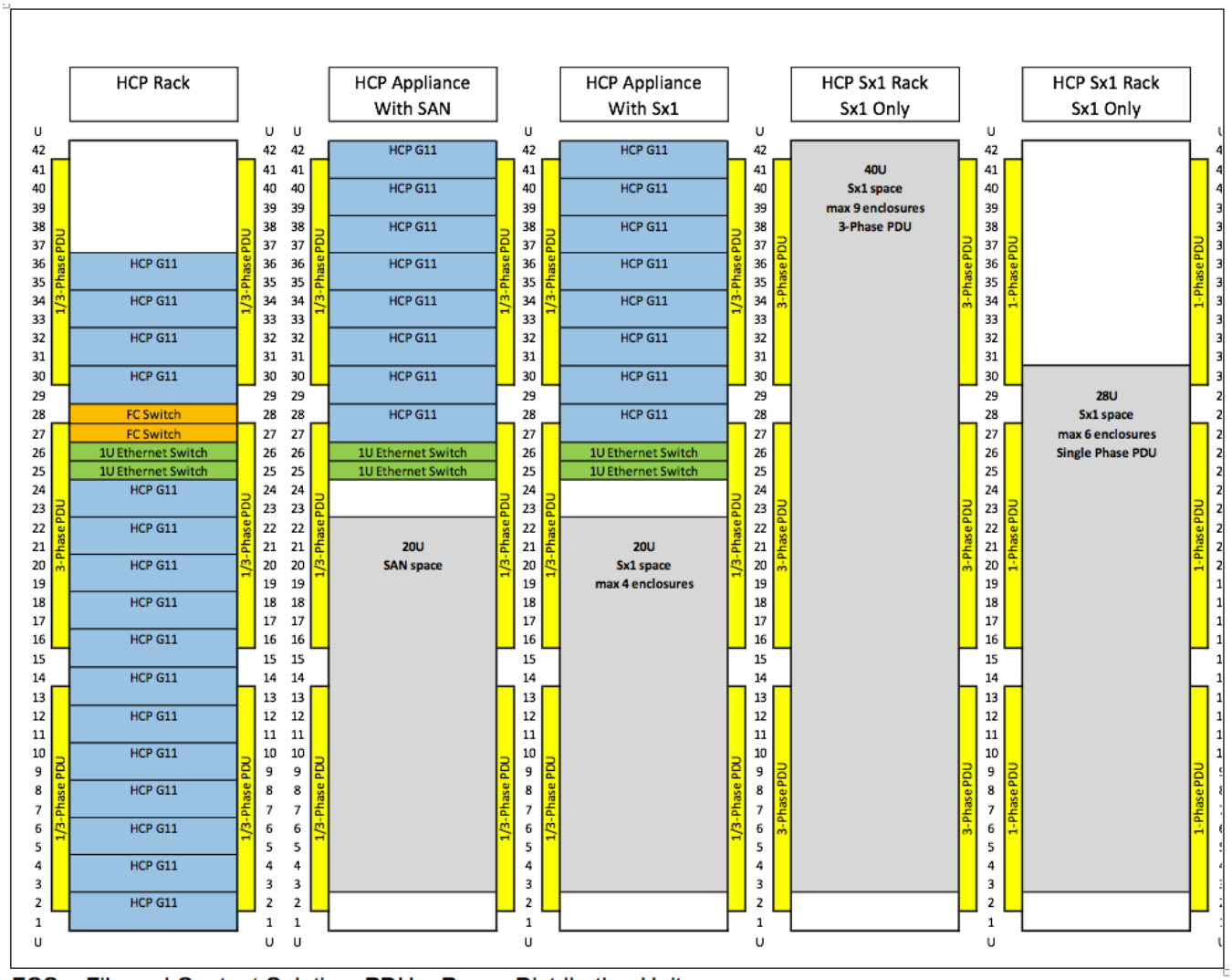


HCP Racking Options

There are numerous racking options that combine the HCP G11 node with HCP S series systems (see Figure 21).

The “HCP Appliance with Sx1” configuration (third from the left) combines up to eight HCP G series nodes and four HCP S series enclosures. If an existing SAN-attached array must be supported, only two HCP S series enclosures can be installed in the HCP appliance rack. For larger clusters, HCP G series nodes (left) and HCP Sx1 systems (the two shown furthest to the right in the table) are normally deployed in different racks.

Figure 21. HCP Racking Options With VSP and/or HCP S Series (S11/S31)



FCS = File and Content Solution, PDU = Power Distribution Unit

Security

With any cloud deployment, whether it's public, private or hybrid, data security is paramount. Hitachi Content Platform is readily able to store and protect data at scale, which is crucial to safeguarding all data all of the time. It is an extremely secure and efficient object storage solution with a comprehensive suite of data protection and security must-haves. HCP. Some of these features include:

Ports to support web protocols and administration: HCP cloud storage software requires certain ports to support web protocols and administration. For example, HCP needs TCP/IP ports 80 (HTTP) and 443 (HTTPS, also known as HTTP over TLS), and port 8000 (admin) to conduct both appliance management and data storage tasks.

Host-based firewalls: HCP follows security best practices and disables all external ports and processes that are not required by the software. Moreover, each HCP node runs a firewall that is designed to block all ports not associated with an active HCP service.

Secure remote service: All remote service is performed using SSH, with a 2048-bit key that is exclusively used by Hitachi Vantara's support organization. Organizations are encouraged to disable this SSH access unless service is required, at which time SSH access can be enabled by the system-level administrator.

SSL server certificates: HCP requires one SSL server certificate (self-generated or uploaded PKCS12) for each defined domain to prove authenticity to clients.

Encryption: If enabled, HCP utilizes an AES-256 block cipher with a key (or block) length of 256 bits. This cipher is required for FIPS 140 compliance. Cipher keys are protected with the Shamir Shared Secret mechanism: The key is built only into volatile memory of an HCP system, and can only be rebuilt with a quorum of nodes.

User authentication: In addition to local user accounts, HCP supports enterprise identity services: Microsoft Active Directory and RADIUS.

Per object ACL (access control list): Utilize ACLs to control data access at an individual object level. ACLs provide more granular data access controls that limit the permissions granted to users or user groups, as well as the operations they can use.

Dedicated management network: Administrative tasks can be isolated on VLANs or physically separate Ethernet ports available on HCP servers.

For more detailed security information about HCP, please review the white paper, "Hitachi Content Platform (HCP): An Overview of Server Security and Protection".

System Availability and Data Integrity

A globally distributed Hitachi Content Platform using HCP S series storage nodes provides 10 nines of accessibility with 15 nines of durability , which is equivalent to three milliseconds of annual downtime and one lost object in 100 trillion years per thousand objects. These levels of availability are appropriate for a backup-less cloud platform and are achieved through self-healing and self-monitoring functions that operate both locally and globally. Allowing the addition of new hardware, retiring old hardware, and upgrading software levels without disrupting service also eliminates planned downtime.

In the event that a site or HCP system is unavailable, users and applications can transparently access another HCP system at another site, ensuring continuity of service with 10 nines accessibility. HCP can rebuild content using fragments dispersed across clusters, an approach which consumes up to 40% less storage than mirroring the entire set of data. An HCP system includes at least four access nodes and can continue full operation despite the failure of a node, resulting in five nines accessibility at the individual cluster level. An access node is the server that serves as a building block for an HCP cluster, and it includes multiple redundant components and 99% node-level accessibility.

Each object stored on an HCP S series storage system is dispersed across its drives, and each fragment can survive the simultaneous loss of six drives, thus providing a durability of 15 nines. Besides orders of magnitude higher resiliency than RAID, the HCP S systems offer Reed-Solomon Erasure Coding, which delivers higher capacity efficiency than RAID-6 and RAID-5. HCP's data-protection-level policy ensures multiple copies of each object are maintained and self-healing occurs when a copy becomes unavailable. When data is stored on HCP's access nodes, it leverages both the data protection level and underlying RAID-6 scheme to provide at least 13 nines durability. Background services periodically run on access nodes to ensure the integrity of content and metadata.

For a more thorough description and analysis of HCP availability and durability please review the white paper, "The Path to 10 Nines Availability with Hitachi Content Platform".

Conclusion

The Hitachi Content Platform object storage system offers an intelligent solution for the dilemma of unbridled unstructured data growth. It eliminates the limitations of traditional storage systems and can efficiently scale to virtually unlimited storage. Hitachi Content Platform allows IT to perform all essential data management tasks from a single system. It treats file data, file metadata and custom metadata as a single object whose life cycle is managed as a series of planned storage tier transitions. With secure multitenancy, HCP helps organizations manage huge pools of capacity by dividing them into smaller virtual object stores and delegating partial control of these to owners, called tenants. Within bounds, the tenant owners are given authority to assign capacity quotas, set configurable attributes and choose service levels that support their application needs. This approach allows the object store to support one or more workloads concurrently, including content preservation, archiving, governance and compliance, data protection, cloud storage, hybrid cloud workflows, content distribution and storage for cloud application development and delivery — all from a single physical infrastructure.

Perhaps most exciting, we see markets rapidly adopting object storage as a technology that can be used for any and all storage workloads; it is no longer limited to large and infrequently accessed datasets. HCP cloud storage is the foundational part of a larger portfolio of solutions that includes Hitachi Content Intelligence to transform data into insight, Hitachi Content Platform Anywhere for file synchronization and sharing, HCP Anywhere Edge for modernizing existing file servers, HCP Gateway as a cloud storage gateway, and Hitachi Content Monitor for visualizing and analyzing HCP performance. One infrastructure is far easier to manage than disparate silos of technology for each application or set of users. By integrating many key technologies in a single storage platform, Hitachi Vantara's object storage solutions provide a path to short-term return on investment and significant long-term productivity and efficiency improvements. They uniquely protect existing investments while helping IT evolve to meet new challenges, stay agile over the long term, and address future change and growth.

Additional Resources

- IDC MarketScape: Worldwide Object-Based Storage (OBS) 2019: https://www.hitachivantara.com/en-us/pdf/analyst-content/worldwide-object-based-storage-2019-vendor-assessment-idc-marketscape.pdf
- IDC Hitachi Content Platform: End-to-End Portfolio for the 3rd Platform https://www.hitachivantara.com/en-us/pdf/analyst-content/hitachi-idc-technology-assessment-hcp.pdf
- Datasheet: Hitachi Content Platform https://www.hitachivantara.com/en-us/pdf/datasheet/hitachi-datasheet-content-platform.pdf
- White Paper: The Path to 10 Nines Availability With Hitachi Content Platform https://www.hitachivantara.com/en-us/pdfd/white-paper/path-to-10-nines-availability-with-hcp-whitepaper.pdf
- White Paper: Hitachi Content Platform — An Overview of Server Security and Protection https://www.hitachivantara.com/en-us/pdf/white-paper/content-platform-whitepaper.pdf
- Third-Party Compatibility Guide: Hitachi Content Platform, Hitachi Content Platform Anywhere and Hitachi Data Ingestor https://www.hitachivantara.com/en-us/pdf/best-practices/hitachi-interoperability-guide-hcp-hcp-anywhere-hdi.pdf

Hitachi Vantara

Corporate Headquarters
2535 Augustine Drive
Santa Clara, CA 95054 USA
hitachivantara.com | community.hitachivantara.com

Contact Information
USA: 1-800-446-0744
Global: 1-858-547-4526
hitachivantara.com/contact

WP-553-D BTD June 2020