SOLUTION PROFILE

HITACHI
Inspire the Next

Hitachi Data Ingestor and Hitachi Content Platform

Edge-to-core Storage Solution for Cloud and Distributed IT

Hitachi Data Systems provides distributed consumers of IT, such as remote offices and branch offices (ROBOs), or cloud consumers, with a bottomless, backup-free storage solution. Hitachi Data Systems also delivers fully managed, consumption-based cloud services specific to distributed IT environments. These are achieved through a minimal-footprint or virtual appliance that sends data from the edge to a core infrastructure that employs advanced storage and data management capabilities. Hitachi Data Systems greatly reduces the cost and complexity of providing IT services to geographically dispersed consumers. With Hitachi Content Platform at the core and Hitachi Data Ingestor at the edge, the benefits of cloud are within reach.

Overview

Providing access to, protecting, managing and preserving unstructured data are growing challenges that are made even more complex in distributed IT environments. From organizations with ROBOs to cloud service providers whose consumers are remote by the nature of the business, Hitachi Data Systems integrated solutions can help. We support data movement from distributed locations at the edge to a centralized infrastructure at the core.

These solutions reduce the cost and footprint of IT for ROBOs and help cloud service providers offer their customers greater control, security and protection for their data. At the edge is Hitachi Data Ingestor (HDI), which acts as an "on-ramp" for users and applications. Hitachi Data Ingestor connects the local site to Hitachi Content Platform (HCP) at a core data center. Advanced storage and data management capabilities can be employed at the core data center to store, protect, replicate, preserve, archive, discover, distribute and search content. This eliminates the need to provide these capabilities at each remote location.

This combination helps IT organizations provide storage for ROBOs in a cost-effective manner. It centralizes and automates complex data and storage management practices and eliminates the need for tape-based backups. It also helps cloud service providers deliver the benefits of cloud storage without the need to design and build their own edge-to-core infrastructure.

The Edge: Hitachi Data Ingestor

Hitachi Data Ingestor provides a standard connection, or on-ramp into the core data center without requiring application recoding and without changing the way users interact with storage. Because HDI is essentially a caching device, it provides users and applications with seemingly endless storage and a host of newly available capabilities. HDI employs an intelligent caching algorithm that identifies and retains newer and frequently accessed files. As a result, frequently used files will likely stay on HDI all the time, enabling organizations to deliver IT services to their distributed users and/or cloud consumers without burdening the WAN or sacrificing productivity. All files on HDI are migrated to HCP; however, users may choose to "pin" selected files so that these files will also stay on HDI all the time for quicker and easier access.

How HDI works for distributed environments is simple. HDI at each location is mapped to its designated core infrastructure tenant. Within the tenant are a number of namespaces. Each file system in HDI is mapped to its designated namespace for proper segregation of data and end-to-end access control. This way, HDI systems and their clients can write only to their assigned tenants and namespaces. Each namespace can be, if so desired, shared by multiple file systems on different HDIs, but only the designated one can write to the namespace while others can only read from it. This content-sharing feature enables "edge dispersion." For example, one ROBO could access the content written by another, which is a great capability for any distributed or cloud environment that desires it. HDI is able to adapt cloud infrastructure to users and existing applications, while multiple service levels are managed from a single physical cluster on Hitachi Content Platform. Now, archive, backup, compliance, data lifecycle management, e-discovery and file tiering are all centralized and manageable from a single solution.

For easier and efficient control of distributed IT, HDI comes with a management API that enables integration with the HCP management user interface (UI) and other 3rd-party or home-grown management UIs. Organizations can integrate HDI management into their home-grown management infrastructures for deployment and ongoing management, all of which will accelerate cloud adoption and deployment, as well as improve user experiences.

HDI is also capable of performing "file restore," a practical feature that enables users to retrieve previous versions of a file or even deleted files. HDI does this by creating historical snapshots that users can access.

Furthermore, Hitachi Data Ingestor provides the means for administrators to automatically migrate data from their existing NAS systems into HDI. This allows organizations to maximize the value of their legacy data without going through a disruptive migration process.

HDI is offered in 2 deployment options to support either local or remote HCP access. HDI offers a high availability (HA) dual-node cluster with external storage, a single node with internal storage, and a software-only configuration running on VMware vSphere Hypervisor (Virtual HDI) that also supports HA architecture. Virtual HDI offers a provisioning infrastructure that enables configuration and deployment of custom virtual HDI appliances via VMware vSphere. In all configurations, HDI acts as a tiering solution, moving its resident files to HCP, and providing links to those files for on-demand recall.

HDI supports leading WAN acceleration solutions to further enhance WAN performance, including distance WAN quality and bandwidth; it delivers the best cost efficiency and operating efficiency over the WAN. WAN optimization solutions assist in breaking down the WAN bottlenecks to complement HDI and HCP in the distributed environments. Additionally, the solutions provide WAN acceleration by compressing and deduplicating data while traversing the network.

Figure 1. The Hitachi Data Ingestor and Hitachi Content Platform Solution

The Core: Hitachi Content Platform

The core infrastructure deployed at primary data centers is Hitachi Content Platform. This massively scalable, multitenant distributed object store can be divided into virtual content platforms, each configured for the desired level of service.

HCP treats file data, file metadata and custom metadata as a single object that is tracked and stored among a variety of storage tiers. With multitenancy and configurable attributes for each namespace within a tenant, the object store can be divided into a number of smaller virtual object stores. Each object store has the appropriate attributes to provide different service levels that can support various users, applications and even organizations. This allows HCP to support a wide range of workloads, such as archive, data protection and content depot from a single physical infrastructure.

In addition, HCP gracefully handles changes in scale and storage technology, minimizing disruption and enabling the environment to live for decades or even centuries. This also allows the solution to scale smoothly from tens of terabytes to an amazing 40PB of capacity in a single cluster. Add compression, deduplication and even optional deletion capabilities to a single platform that provides massive and incremental scalability with minimal downtime. The result? Storage sprawl in support of unstructured data becomes a thing of the past.

The Solution: Integrated Edge-to-core Technologies

The on-ramp adapts public and private clouds, with Hitachi Content Platform at the core, to users and applications by presenting standard protocols and read and write capabilities. Behind the scenes, the Hitachi Data Ingestor

sends data to HCP, where advanced data and storage management capabilities can be used to store, protect, preserve, distribute, retain, govern and search content (see Figure 1). Since HDI is tightly integrated with HCP, users can access the robust HCP capabilities without disrupting existing users or processes.

Hitachi Cloud Services: Storage as a Service in the Content Cloud

Hitachi Data Systems delivers fully managed, consumption-based, onsite cloud services, helping users reduce capital and operational expenditure (CAPEX and OPEX). At the same time, these services provide the agility to grow and shrink capacity to meet the ever-changing demands of the business.

Hitachi Cloud Service for File Serving addresses specific challenges of "distributed IT" in a scalable, easily provisioned, pay-per-use fashion. By leveraging HDI and HCP in a private cloud delivery model, organizations can centralize their distributed data into the content cloud environment located within their own firewalls. This way they pay only for the storage that they consume when they consume it. Capacity is requested and provisioned from the Hitachi self-service portal. This allows organizations to provision additional storage as needed, offer a variety of service levels and avoid managing the private cloud environment themselves.

Benefits

Hitachi Data Systems provides organizations with significant cost savings with this integrated edge-to-core storage solution and its related Cloud Services. These savings benefits range from reduced complexity of overall IT infrastructure to management overhead at distributed location, as well as to capital expenses for distributed storage and backup infrastructure. Specifically, the solution can help organizations with the following:

Stop Storage Sprawl

Organizations with distributed consumers of IT services, such as cloud services, groups with ROBOs, issues such as local storage growth, content control and regular backups add significant cost and complexity. Different applications have different storage requirements, work with different protocols, need varying levels of protection and performance, and so forth. The traditional approach has been to deploy different silos of technology for each workload and each geographical location. With Hitachi Data Ingestor and Hitachi Content Platform, a wide range of workloads and distributed users can be serviced by a single, shared infrastructure. This enables IT service providers to deploy HDI at distributed locations. They can leverage the scale and functionality of a centralized HCP to significantly reduce the IT footprint at distributed sites while improving service levels, reducing complexity and driving down costs.

At the core, HCP avoids the limitations of traditional file systems by intelligently storing content in far larger quantities and in a much more efficient manner. This enables HCP to handle the explosion of unstructured data and its growing importance to organizations, their partners, their customers, their governments and their shareholders.

Reduce the Dependence on Tape-based Backups

Because data written to Hitachi Data Ingestor is stored in Hitachi Content Platform, it can take advantage of advanced HCP data protection and content preservation capabilities. This allows organizations to address data protection and disaster recovery requirements at their distributed sites without the financial and management burden of tape-based infrastructure at each location. With the ability to control replication and other capabilities at the level of individual namespaces (a subset of a tenant) users can select the appropriate level of protection needed for individual applications or datasets. From a cloud service provider perspective, they can avoid the cost of tape-based infrastructure for the core. They also can pass those savings and more on to their customers by offering solutions that eliminate their customers' needs for tape backup. In either case, tape-based backups at the edge can be eliminated in favor of centralized, advanced replication capabilities at the core.

Once the data is inside HCP, the dependence on tape-based backups can be minimized even further by leveraging HCP data reduction and protection capabilities.

Data deduplication and compression are used to control data size by eliminating unnecessary copies and shrinking the amount of storage used for a given piece of content. As new objects are written to the object store, the content is compared with similar objects and unnecessary, duplicate data is eliminated or compressed to save space. If the core data center is replicated for disaster recovery purposes, selective replication can be used to reduce the amount of data at replica sites and conserve precious replication bandwidth. Controlling the overall amount of storage consumed streamlines failover to secondary systems and recovery of primary systems once the failure is repaired.

Another way HCP can help with data reduction is through its data retention and disposition services. These services automatically keep content for the prescribed duration and, barring a retention hold, automatically delete expired content so the capacity can be reclaimed and recycled back into available storage. These deletions can be logged and annotated to provide an audit trail of what content was removed, when, by whom and why. These technologies are key as the traditional methods of keeping every file forever and backing up all files every week are too costly and risky in today's economic, regulatory and legal climates.

From a data protection standpoint, HCP ensures

KEY BENEFITS

Reduces Cost

- Eliminates backups at the edge by providing a highly available on-ramp into a centralized storage solution and taking advantage of robust storage management and content governance capabilities
- Reduces the overall IT footprint and delivers a variety of storage services from a single infrastructure
- Improves efficiency and utilization by consolidating distributed silos

Reduces Complexity

- Simplifies and speeds the process of bringing new users or applications on board and time-to-market for cloud services
- Easily adapts to changes in scale, adopts new storage technologies and mitigates the impact of infrastructure changes
- Monitors, reports and audits usage with built-in tools

Reduces Risk

- Automates management and enforcement of regulatory, governance and data lifecycle policies
- Secures and protects content with encryption, WORM, replication, RAID, data integrity checks and more
- Supports full integration with Active Directory and LDAP

Streamlines Cloud Deployments and Adoption

- Supports thousands of edge devices with a single core system
- Enables the sharing of content between edge sites to aid in content collaboration and distribution
- Improves performance and access times with bandwidth optimization and content pinning

KEY ATTRIBUTES

Scalability

- Up to 400 million files and thousands of users per Hitachi Data Ingestor
- Up to 40PB and billions of objects in a single Hitachi Content Platform cluster
- Up to 100 file systems across the environment

Performance

- Metadata at native operating system (OS) speed
- Cached file content at native OS speed
- Reads at HTTP speed
- Writes at native throughput speed

Supported Technologies

- Supports NFS, CIFS, REST
- Supports full integration with Microsoft Active Directory and LDAP authentication
- Supports Hitachi Virtual Storage Platform and Hitachi Adaptable Modular Storage 2000 family

data integrity with "write once, read many" (WORM), RAID-6, encryption and more. By adding services such as dynamic data protection levels, advanced replication, version awareness and the ability to browse the environment, the object store ensures objects are well protected and easily recoverable. As the data is onsite and on disk that can be easily browsed, content can be recovered quickly, on demand, at a particular point in time and in a self-service manner. This reduces help desk costs and avoids the hassle of recovering from tape.

Improve Content Control

Hitachi Content Platform offers replication capabilities that not only protect content, but also aid in the control and distribution of content among geographically dispersed storage resources.

These features enable the consumers of storage to ensure that their data is placed in the appropriate physical location and that it can be accessed by the appropriate users. A file written to Hitachi Data Ingestor has the desired retention attributes assigned and is then copied to HCP, where it becomes an object. A stub is left behind at the local site to provide access and recovery of current as well as older versions of the object stored at the core.

An organization's needs may include distributing content for local access, ensuring that location-specific data does not leave a given site, and complying with regulations and requirements around where data can be physically located. HCP provides the tools to ensure that data is handled properly over its life. HCP also makes sure the data is disposed of properly when it is no longer of value to the organization, or is no longer a candidate for storage in the cloud.

Extract the Value of Data

Hitachi Cloud Services provide intelligence, content awareness, metadata mining tools, application independence, and automation needed for "big data" situations, making information easier to find, share, manage and repurpose.

For Distributed IT

For organizations with ROBOs, storage growth, content control and regular backups add significant cost and complexity. In sectors such as healthcare, banking, insurance, retail and others where branch and local offices are the norm, providing for the storage, management, protection and compliant handling of data is expensive, time-consuming and complex. Hitachi Data Systems helps reduce the IT footprint needed at those remote locations by providing an edge device with which users and applications interact just like in traditional storage. Unlike traditional storage, though, the edge device requires only a small amount of local storage. This amount serves as a local cache for data that needs to be copied to or retrieved from the core, where the data ultimately resides. Once at the core, advanced data and storage management features can be applied to the data to ensure protection, preservation, governance and security.

For Cloud Services Providers

For service providers that seek to offer cloud storage services to external customers, there are the added issues of security, performance, data protection, chargeback and more. Even adapting users and applications to the use of cloud-based storage poses a significant challenge. Hitachi Data Systems offers a unique approach that delivers seemingly endless, backup-free storage at the edge that acts as an on-ramp to advanced cloud storage and data management capabilities at the core. This combination helps cloud service providers deliver the benefits of cloud storage without the need to design and build their own edge-to-core infrastructure. Hitachi Data Systems offers fast time to market with a solution that adapts to how users and applications work today. This integration of cloud on-ramp and cloud infrastructure enables service providers to offer more reliable, trustworthy services with a wide range of features and options.

Summary

IT organizations serving distributed consumers benefit from the cloud on-ramp powered by Hitachi Data Ingestor. HDI coupled with Hitachi Content Platform creates an edge-to-core storage solution that is ideally suited for the challenges of supporting the needs of unstructured data in distributed IT models, such as ROBOs and cloud service providers. With Hitachi Data Ingestor at the edge sending data to the core infrastructure in the data center, distributed IT environments can reduce their dependence on local storage at the edge. In doing so, they reduce the cost and complexity of backup, archiving, compliance, content distribution and more. At core IT data centers, Hitachi Content Platform ensures protection, security, reliable access, preservation, compliance, adherence to policies, and a host of other capabilities that simplify management and automate complex processes. Combined, Hitachi Data Ingestor and Hitachi Content Platform propel the general concept of object storage forward, help rein in the cost of distributed IT and bring cloud within reach.

Hitachi Data Systems

Corporate Headquarters
750 Central Expressway
Santa Clara, California 95050-2627 USA
www.HDS.com

Regional Contact Information
Americas: +1 408 970 1000 or info@hds.com
Europe, Middle East and Africa: +44 (0) 1753 618000 or info.emea@hds.com
Asia Pacific: +852 3189 7900 or hds.marketing.apac@hds.com