



How to Manage Unstructured Data in Distributed IT and Cloud Deployments

Hitachi Edge-to-core Storage Solution

By Hitachi Data Systems

April 2012

Table of Contents

Executive Summary	3
Introduction	4
Challenges with Distributed IT Environments	4
Find Agility with Object Stores	7
Key Terms	8
Hitachi Edge-to-core Storage Solution	8
Welcome to the Edge: Hitachi Data Ingestor	10
Why Not Just Use NAS?	14
At the Core: Hitachi Content Platform	15
Tap into Tangible Benefits	18
The Silver Lining for Cloud Deployments	18
No Longer "Out on a Limb for Branch and Remote Offices	18
Shrink Tape-based Dependencies and Costs	18
Gain Control over Unstructured, Unruly Content	19
Simplify Storage Management and Savings	19
On the Horizon of Change	20

Executive Summary

Among today's most urgent IT challenges are the rampant growth of unstructured content and the need to reduce the costs, complexities and risks associated with it. Particularly tricky is the management and proper handling of this content across distributed IT environments, such as cloud service providers and organizations with branch or remote office sites. Distributed IT environments are often saddled with sprawling storage silos for a multitude of applications and workloads. They suffer from a buildup of backup infrastructure, and they have inadequate IT resources and safeguards at many locations outside the data center.

Hitachi Data Systems is focused on simplifying and accelerating cloud and other distributed IT environments to readily accommodate shifts in business, economic and regulatory demands.

Using a proven and integrated approach to manage distributed IT environments and the flood of storage requirements for unstructured content, Hitachi helps cloud service providers and enterprise customers more easily provide storage services to geographically dispersed users.

This paper examines how the Hitachi edge-to-core storage solution can decipher the challenges associated with managing unstructured content for distributed IT environments. With Hitachi Data Ingestor at the edge and Hitachi Content Platform at the core, this integrated solution addresses storage management, backup, archiving, compliance, content control and other key issues facing distributed organizations and cloud storage service providers.

Introduction

As stored data grows to an unfathomable extent, even in moderately growing organizations, so does the urgency to better control, protect, manage and preserve it. With the amount of both structured and unstructured data reaching colossal proportions, IT leaders are probing for serious solutions to mitigate costs, complexities and risks often inherent in the modern data center. In particular, unstructured data, including documents, files, images and other content that cannot be defined in terms of rows and columns or standard structure, has given rise to a new set of challenges for IT. The proliferation of unstructured data growth and increasing variation in data types continues to accelerate, leading to the spread of unplanned storage, which ultimately drives up operating expenses, capital expenses, management complexity and risk.

Beyond the types of data vying for storage are the matters of extensive scalability and concurrency: How do we keep things coherent while scaling? The reality of manageable data storage remains elusive to many enterprise IT organizations, especially when hundreds of millions of objects must be stored and readily accessible for varying or indefinite periods of time.

Compounded by the emergence of more distributed IT environments, such as remote or branch offices and cloud deployments, the challenges within IT might seem insurmountable. Rather than continually bolting on more capacity, the time has come to change how content is stored and managed.

Hitachi Data Systems has been integrally involved with solving these complex issues on behalf of its enterprise customers. For organizations with branch offices or remote locations and for cloud service providers whose consumers are remote by the nature of the business, integrated Hitachi edge-to-core solutions help to seamlessly and securely move data between distributed locations and a centralized infrastructure. The results are greater cost efficiency and reduced IT footprint at the remote or branch sites, and better control, security and data protection for consumers of cloud services.

Challenges with Distributed IT Environments

For IT, the ability to store and protect company data at one location can be a daunting task. With multiple sites, the challenge of storing, securing and appropriately managing data storage becomes exponentially more difficult (see Figure 1). To better comprehend how to address these data management issues in distributed IT environments, it is wise to first understand why these issues exist. And for purposes of this examination, Hitachi defines distributed IT environments as:

- **Remote or branch offices (ROBOs).** Remote or satellite offices are usually leased offices in a commercial building. They provide professional office space in locations that are near where employees live, to cut down on commuting while still ensuring access to professional-grade Internet, phone service and security. Branch offices, on the other hand, are secondary places of business apart from a company's headquarters. Branch offices vary in staff size and may or may not provide a full suite of services. Some real world organizations that use branch offices include banks, retailers, healthcare groups, manufacturers and government entities.
- **Cloud service deployments.** Cloud is a fresh approach to providing IT services at reduced costs relative to more traditional IT environments. Cloud is delivered as a utility-type service model that charges users based on consumption of services or capacity. Cloud may be deployed by

enterprises doling out services on demand to business units or by cloud service providers. Key tenets of cloud are the ability to:

- **Rapidly provision or de-provision a service**
- **Flexibly scale services up or down**
- **Simply enable self-service access to storage resources**
- **Easily secure direct connections of existing systems to the cloud, without recoding requirements**
- **Quickly employ multitenancy capabilities that segregate and protect the data**

Figure 1. With distributed environments, the challenge of storing, securing and appropriately managing data storage becomes exponentially more difficult.



Specifically, some of main problems IT must tackle when supporting distributed environments include:

Storage Sprawl

When an organization has distributed work environments, more and more critical company data will reside away or downstream from the center or core of the enterprise. This leads to the unplanned, uncontrolled spread of local storage at the edge, which is known as storage sprawl. Storage sprawl can quickly fragment data center resources and complicate how effectively and cost-efficiently data is managed.

Too much to store also means too much to back up and manage. Neither cloud service providers nor enterprises with remote or branch offices can afford to be weighed down with a glut of storage, backup, content control and multifaceted application demands. Traditional approaches of deploying different silos of technology for each workload or geographical location directly adds to the cost, complexity and risk required to manage them.

Unstructured Data

Unstructured data continues to grow faster than most IT budgets, leaving IT leaders clamoring for ways to derive more value out of their storage. Data types such as rich media, PAC images, legal documents, presentations and web pages are among the countless content types that impact the data center's ability to maintain control over this unstructured data. The technologies and tools that were developed in support of structured data do not work the same miracles on unstructured data. The ever-increasing growth in unstructured data hampers nearly all the mainstays of storage administration activities, including backup, restoration, performance management and compliance. It is certainly a weight on more emergent "here to stays" such as cloud computing services and ROBO environments.

Scalability

When the enterprise IT organization discusses massive scalability, it is no longer in terms of terabytes but, rather, petabytes. The challenge is how to achieve it without disruption. Because data must reside for indefinite periods in anything from the cloud or content depots to archives and other systems, being able to scale storage flexibly, securely and extensively is essential. Scalability can have a substantial impact on the economic benefits of cloud deployments. To achieve the economies of scale that cloud can offer, storage scalability must be extensive, seamless and on demand.

For distributed IT environments, one of the challenges with scalability is being able to manage the complexities of various storage settings without sacrificing cost efficiency, data agility, performance or security. Organizations are grappling with the need to manage active archives, backup reduction, cloud deployments and other mission-critical facilities. All of these demand extensive scalability, often within a multitenant environment, in order to share that large pool of storage efficiently with diverse users and applications.

Archiving and Compliance

While archives used to be just the place to store static content for the long term, today's IT center is looking for more from its archival systems. They need to comply with corporate governance, government oversight and other content preservation requirements. Also, archiving-as-a-service is cropping up in cloud deployments; organizations want to ensure that digital content is not only preserved for the long haul but also properly managed through its lifecycle and able to be accessed far into the future.

Customarily, companies end up with multiple, and often siloed, archive storage solutions to support a variety of applications and fixed content, subsequently consuming time and resources to maintain. This may generate unintended challenges, such as managing storage utilization and growth; retrieving information for business, e-discovery and litigation requests and compliance audits; and administering changing policies for data retention. The risk comes in the potential loss of data and inability to provide records, which can violate regulations for content preservation and retention, incurring costly fines and penalties. Today's enterprise needs to archive critical data without sacrificing availability or rapid retrieval of data, and ensure file integrity and access along with authentication and retention enforcement.

Tape-based Backup and Data Durability

Many organizations still rely heavily on tape-based backups, which were designed to efficiently write data sequentially but not to rapidly and randomly retrieve that data. As any IT professional can attest, storing data on tape comes with its own set of challenges. These may include everything from media rotation hassles and labor-intensive retrievals to shelf life limitations, media degradation and data loss. Because organizations are often required to lengthen the time that data must remain stored and accessible from months to decades to meet retention mandates, data durability is now a growing issue. Cloud and remote operations demand capabilities beyond what tape backup can provide, including dynamic levels of data protection, constant integrity checking, RAID, data placement, search functionality and fast access.

Business Continuity and Disaster Recovery

Data at distributed IT environments requires but does not always get the same levels of data protection and recovery as data residing on core infrastructure. Challenges with a distributed work model can hinder business continuity and disaster recovery. Some of these challenges include: the lack of automated backup to centralized data storage, insufficient network bandwidth, not enough IT staff to properly manage the local sites, isolated systems and legacy software. In some cases, local files might not always be backed up regularly to the network or upstream infrastructure. This makes them inaccessible to other offices or hinders the ability ensure compliance or quickly and completely recover data. Site-level disasters such as security breaches, catastrophic events and even human error can be minimized when the enterprise deploys a centralized data protection strategy for better business continuity.

Data Reduction

Disparate systems and duplicate copies are common contributors to lengthy backup and restore times, as well as distressing performance and availability of production systems, consequently driving up costs and possibly affecting compliance and legal mandates. Data reduction is the process of minimizing the amount of data that needs to be stored. This can be achieved using several different types of technologies, helping to increase storage efficiency and reduce costs.

Data deduplication and compression are 2 reduction methods that control data size by eliminating unnecessary copies and shrinking the amount of storage used for a given piece of content, respectively. As IT leaders know, traditional techniques of keeping every file forever and backing up every file each week incur too much cost, risk and time. Controlling the overall amount of storage consumed can streamline data retention, reclamation and recovery while mitigating risky economic, regulatory and legal issues.

Find Agility with Object Stores

Hitachi Data Systems looks to object storage solutions as a proven way to address the challenges of ever-growing and long-lived fixed content. At their most basic level, object storage devices are responsible for managing the storage of objects. An object is essentially a data container with at least 2 elements: data and metadata. Data is the file: the PACS image, the email, the presentation or spreadsheet. Metadata is information about that data used to describe the definition, structure and administration of the file. In some cases, the metadata might actually be more important than the data itself, such as a physician's notes about a patient image or a legal reference to a case file.

Intelligent object stores offer an intelligent evolution of data movement, space management and security functions. Unlike block-based storage, which can only move a sequence of bytes or bits as one entity or block, distributed object storage can distinguish and move discrete objects with awareness about the data or metadata. This means having the ability to create and delete objects, write and read bytes to and from individual objects, and determine and acquire the attributes or metadata on objects.

While all storage must consider the same issues, such as data coherency, availability, reliability, durability, replication and recovery, intelligent object stores provide added power and capabilities. IT can now apply retention policies to manage these individual objects as well as provide different replication, versioning and protection settings. In essence, the objects themselves become intelligent with their own unique fingerprint or DNA.

So how does the IT organization support distributed object storage and diverse use cases while simultaneously evolving the data environment efficiently, securely and nimbly? Hitachi Data Systems provides state-of-the-art distributed object storage solutions for short-term return on investments and significant long-term efficiency improvements so that enterprises can meet both the known and unknown IT challenges ahead.

Key Terms

To fully appreciate the assessment of the Hitachi Data Systems solution for distributed IT environments and cloud deployments, a few more terms may be helpful.

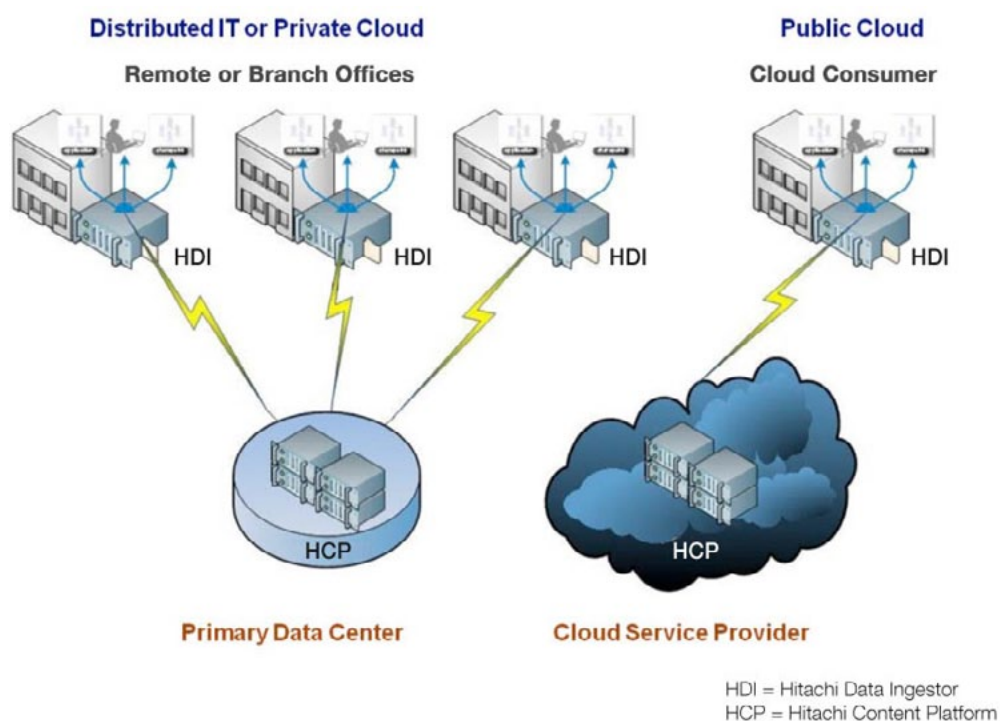
- **Multitenancy.** This architectural model allows multiple end users or cloud service customers to share a single storage infrastructure by partitioning that infrastructure (application, storage pool, network, etc.). The storage pool, for example, is divided into as many as 1000 tenants, which can have a number of namespaces, up to 10,000 in total; these namespaces may be either for separate end users in a hybrid or public cloud, or for business units in private cloud, enterprise or distributed IT environment.
- **Namespaces.** These are used to segregate data, and are uniquely configured for separate service levels, data management policies and data access rules.
- **Representational state transfer (REST).** This is a type of software architecture for client or server communications over the web.
- **On-ramp.** This is a standard connection or entry method that enables applications to connect to storage without disruption of existing infrastructure. For cloud services, on-ramps provide connectivity choices and functionality for populating the cloud from multiple sources.
- **Stub.** This is a small pointer that replaces files at the edge so applications and end users can access the file transparently from the core infrastructure.

Hitachi Edge-to-core Storage Solution

Hitachi Data Systems takes the general concept of object storage a leap forward with the Hitachi edge-to-core storage solution. This solution provides distributed consumers of IT, including remote and branch offices and cloud storage patrons, with seamless, bottomless and backup-free storage

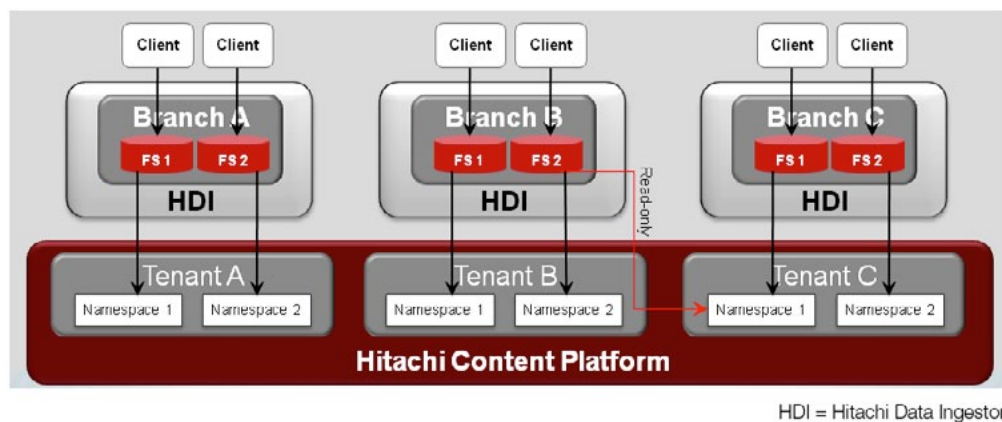
(see Figure 2). The 2 products that make up this offering are Hitachi Data Ingestor (Data Ingestor or HDI), a minimal-footprint or virtual appliance placed at the edge, and Hitachi Content Platform (Content Platform or HCP) as the core infrastructure. Hitachi Content Platform is a multipurpose object store with built-in intelligence, virtualization and massive scalability that eliminates the need for a siloed approach to storing unstructured content.

Figure 2. Hitachi edge-to-core storage solution provides seamless, bottomless and backup-free storage.



How it works for distributed work environments is simple, as shown in Figure 3. The Data Ingestor at each location is mapped to its designated core infrastructure tenant. Within the tenant are a number of namespaces. Each file system in that edge appliance is mapped to its designated namespace for proper segregation of data and end-to-end access control. This way, Data Ingestor systems and their clients can write only to their assigned tenants and namespaces.

Figure 3. Hitachi Data Ingestor to Hitachi Content Platform Mapping and Content Sharing

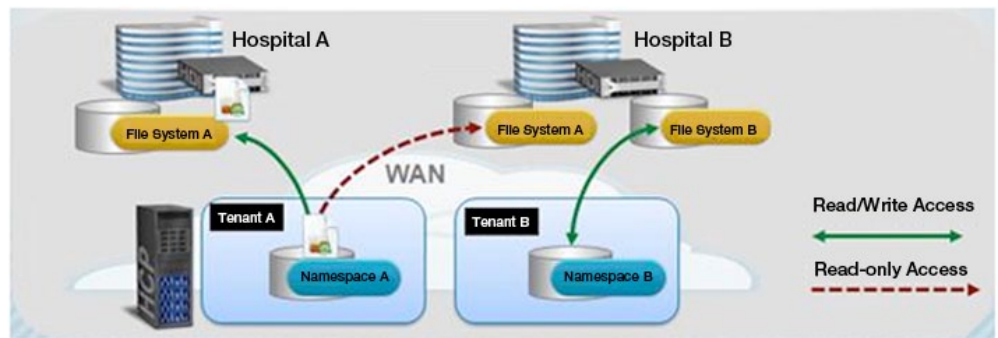


Each namespace can be, if so desired, shared by multiple file systems on different Data Ingestors. However, only the designated one can write to the namespace, while others can only read from it. This content sharing feature enables “edge dispersion,” which allows one branch or remote office to access the content written by another. This is a great capability for any distributed or cloud environment looking to publish and share a consistent set of content among distributed sites, as shown in Figure 4.

Data Ingestor 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.

Figure 4. The content sharing feature enables "edge dispersion."

- Hospital A stores medical files to file system A of Hitachi Data Ingestor (HDI) A.
- The files are migrated to namespace A on Hitachi Content Platform (HCP).
- Hospital B can read the files through file system A, which is mapped to namespace A on HCP.



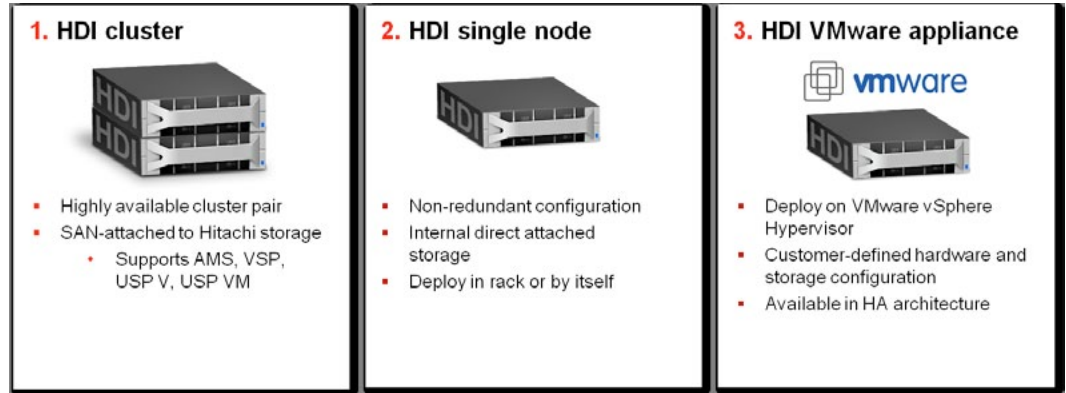
Hitachi Data Ingestor and Hitachi Content Platform work together as a single solution to orchestrate the agility and robust characteristics necessary for rapid cloud adoption and cost-efficient distributed IT deployments. Complex data and storage management practices are cost-effectively centralized and automated and tape-based backup is eliminated. In addition, cloud service providers can deliver the benefits of cloud storage without having to build and maintain their own edge-to-core infrastructure.

Welcome to the Edge: Hitachi Data Ingestor

Operating as a cloud on-ramp for users and applications at the edge is Hitachi Data Ingestor. Data Ingestor is available in 3 configuration options, as shown in Figure 5:

- Dual node configuration with SAN-attached storage
- Single node configuration with internal direct-attached storage
- VMware appliance configuration, supported on the VMware vSphere Hypervisor (ESXi)

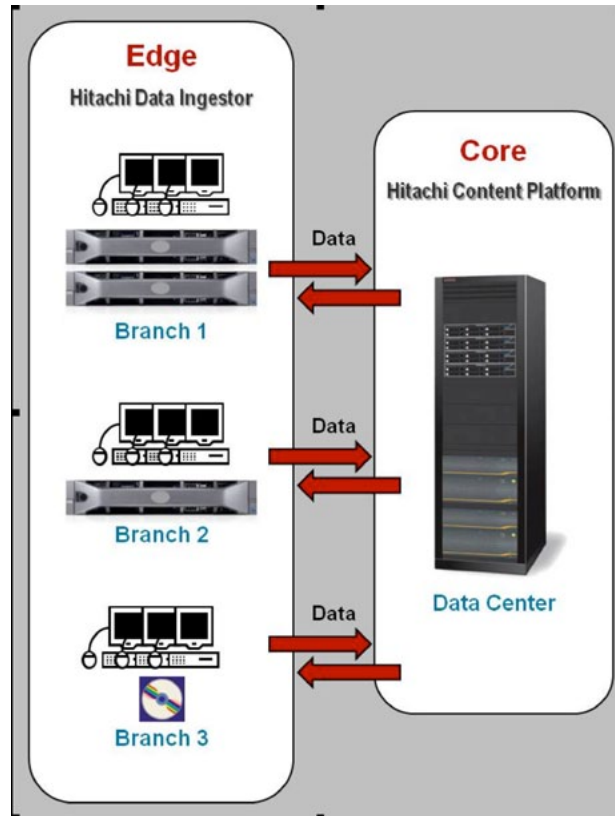
Figure 5. Three Configurations of Hitachi Data Ingestor



HDI = Hitachi Data Ingestor, HA = high availability, AMS = Hitachi Adaptable Modular Storage, VSP = Hitachi Virtual Storage Platform, USP V = Hitachi Universal Storage Platform® V and USP VM = Hitachi Universal Storage Platform VM.

Representing the remote site in distributed IT environments, Data Ingestor connects to Hitachi Content Platform at a core data center (see Figure 6). No application recoding is required for applications to work with Data Ingestor and interoperate with Content Platform. Users work with it like any NFS or CIFS storage. Because Data Ingestor is essentially a caching device, it provides users and applications with seemingly endless storage and a host of newly available capabilities.

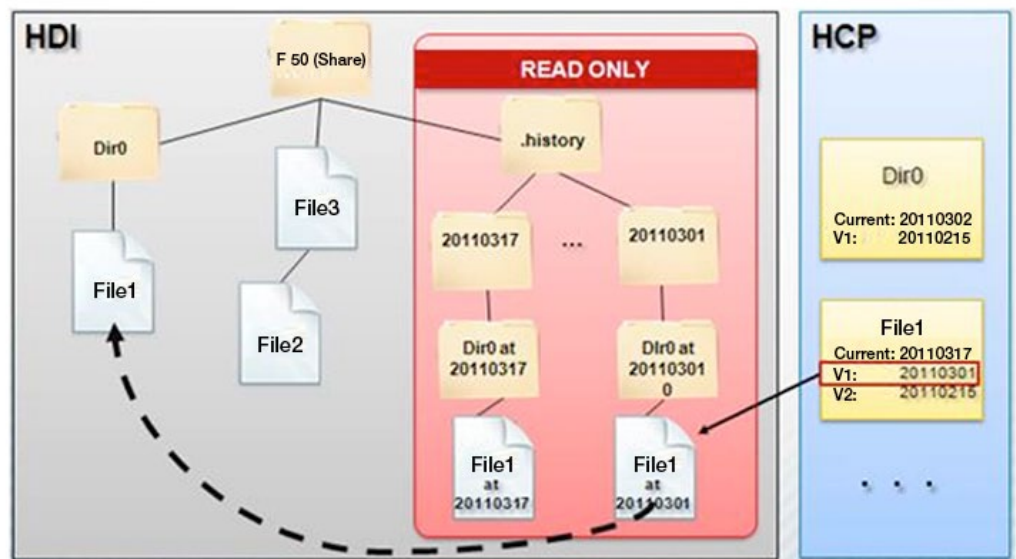
Figure 6. By presenting simplified connectivity at the edge, Hitachi Data Ingestor provides nondisruptive user access to the robust suite of Hitachi Content Platform capabilities.



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

Hitachi Data Ingestor is also capable of performing “file restore,” a practical feature that enables users to retrieve previous versions of a file or even deleted files. Data Ingestor does this by creating historical snapshots that users can access, as shown in Figure 7.

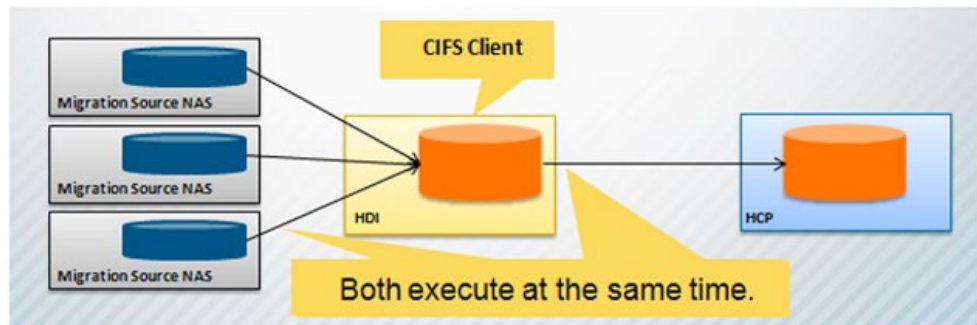
Figure 7. Hitachi Data Ingestor is capable of performing “file restore.”



HDI = Hitachi Data Ingestor, HCP = Hitachi Content Platform

Furthermore, Hitachi Data Ingestor provides the means for administrators to automatically migrate data from existing NAS systems into HDI. This allows organizations to maximize the value of legacy data without going through a disruptive migration process. As shown in Figure 8.

Figure 8. Hitachi Data Ingestor provides “NAS migration.”



HCP = Hitachi Content Platform, HDI = Hitachi Data Ingestor

Through this integrated edge-to-core solution, distributed IT consumers can take advantage of the progressive storage and data management aptitude of Hitachi Content Platform for storing, protecting, replicating, preserving, archiving, discovering, distributing and searching content. As part of a tightly integrated solution, Data Ingestor is able to adapt public and private Content Platform-based clouds for thousands of users and applications by presenting standard protocols and read-write functionality. By presenting simplified connectivity at the edge, Data Ingestor provides nondisruptive user access to the robust suite of capabilities of Content Platform.

Hitachi Data Ingestor maintains a local cache for frequently accessed content from Hitachi Content Platform and provides system recovery in the event of failure. Data Ingestor is offered in 2 deployment options to support either local or remote Hitachi Content Platform access. Although the 2 deployments offer the same features, they require distinct configurations optimized for the use case in which they are being deployed. Data Ingestor offers a highly available dual-node cluster with external storage, a single node with internal storage, and a software-only configuration (virtual HDI) running on VMware vSphere Hypervisor. The virtual HDI offers a provisioning infrastructure that enables configuration and deployment of custom virtual HDI appliances via VMware vSphere. In all configurations, Data Ingestor acts as a tiering solution, moving its resident files to Content Platform, and providing links to those files for on-demand recall.

For applications that are local to Content Platform, Data Ingestor can be configured as a "local" system, which supports mainstream Content Platform application use cases.

For distributed IT environments and storage-as-a-service architectures, Data Ingestor is configured as an appliance for locations that are "remote" from Content Platform. Data Ingestor communicates with applications and end users over the LAN, and communicates with Content Platform over a WAN, the Internet or a virtual private network (VPN). All configurations of Data Ingestor include the following important features:

Native File System Access

Hitachi Data Ingestor provides native file system access to Hitachi Content Platform, including common Internet file system (CIFS) and network file system (NFS) access. These standard network protocols allow file system clients remote access and sharing between nodes on the network. Native file system access also preserves Active Directory investments and LDAP authentication policies.

Migration

Hitachi Data Ingestor automatically replicates content to Hitachi Content Platform. This migration is determined by automated policies, and will move all changes once daily. If needed, the policy can be adjusted by an administrator to match specific recovery point objectives (RPOs). All content on the edge appliance is migrated to a central Content Platform and maintains a local link to the migrated content. This is done via a persistent data structure, which always preserves the previous version of itself when modified, making data and its metadata immutable. Changes to files are written as new object versions while the migrated content is saved unmodified in its original form and may be indexed by search.

Data Ingestor supports wide-area migration with HTTPS or HTTP, and replication-aware failover. An application can write a file to Data Ingestor, which will then replicate the file to Content Platform. All files on Data Ingestor are migrated to Content Platform; however, users have the capability of pin-

ning files of their choices so that these files will also stay on HDI all the time for quicker and easier access.

Once capacity reaches 90%, Data Ingestor will then delete any "non-pinned" file above the 90% threshold and create a 4KB stub to replace it. These stubs are transparent to clients and are backed up just like a complete file. File recall is done simply by reading the stub.

Retention and Compliance

Hitachi Data Ingestor includes a comprehensive set of compliance features to allow clients to specify retention data and "write once, read many" (WORM) restrictions that best meet governance policies and regulatory requirements. Data Ingestor supports full read-write access to files, making retention optional.

Similar to the compliance mode on Hitachi Content Platform, retention rules for the Data Ingestor state that once a file's retention date is specified, it may be extended but cannot be shortened. These retention values are managed by an internal clock and are immediately enforced and transferred with file content to Content Platform. Data Ingestor prohibits deletion of WORM files until file system retention expires, and it prohibits renaming of WORM files and directories containing WORM files.

High Availability and Advanced Data Protection

Both Hitachi Data Ingestor and Hitachi Content Platform have high availability architectures with no single point of failure in the solution. Both have replication and failover capabilities, and the edge devices can rebuild themselves using data that is protected, regularly checked and always available at the core. By combining advanced data protection with robust high availability, IT organizations can ensure data availability at both the edge and core with far less cost and complexity than traditional solutions. This is accomplished by centralizing data protection and disaster recovery where more advanced technology can be applied to care for the data.

WAN Optimization

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

Why Not Just Use NAS?

Unlike network attached storage (NAS) solutions, Hitachi Data Ingestor is a seemingly bottomless, backup-free on-ramp appliance that requires little administrative overhead. This solution is intended for organizations looking to reduce their storage footprint and administrative overhead at the edge. For example, many distributed organizations, such as retail, banking, insurance and healthcare, do not have the IT staff or expertise at their distributed locations to properly manage, protect and back up the storage residing at the edge. Data Ingestor is purpose built for Hitachi Content Platform and functions as a caching device, offering NFS/CIFS ingestion and performance acceleration. Data Ingestor has automated policies that replicate all resident files to Content Platform. Since all of the

data is replicated to Content Platform, Data Ingestor can recover all of its files in the event of a failure. Due to its role as a caching device for Content Platform, Data Ingestor has a limited amount of local storage, which cannot be expanded, and requires Content Platform as its data migration target to expand its capacity. This makes Data Ingestor a nearly bottomless, backup-free solution, reducing management overhead and freeing IT from constant capacity planning. Overall, Data Ingestor as a remote appliance placed at the edge in distributed IT environments helps lower operating expenses and, ultimately, the total cost of ownership. No application recoding is required for applications to work with Data Ingestor and interoperate with Content Platform.

At the Core: Hitachi Content Platform

At the core of this integrated solution is Hitachi Content Platform, an intelligent, multitenant, multi-tiered and massively scalable intelligent object store that is virtualized, reliable and efficient. It can be segregated and individually configured into virtual content platforms, or tenants, and does not require tape backup. Hitachi Content Platform helps facilitate one platform for all data, allowing IT to manage a broad range of applications, data types, workloads and users on a single physical cluster, while saving time and money (see Figure 9).

Particularly useful for distributed IT environments, Content Platform shaves away complexity while supporting changes in technology and business requirements, and it presents a common distributed object view to both applications and users. Because the intelligence is built into the core infrastructure, the cost and management overhead required at distributed locations are alleviated, as is the need for distributed storage and backup hardware. Content Platform provides a single point of management and supports diverse use cases, including:

- **Cloud** — with built-in chargeback capabilities, functionality to uniquely segregate and facilitate multitenancy in a secure manner, advanced metadata map management and dynamic expansion to simplify and automate scaling, and a cost-efficient storage platform for managing unstructured content across cloud services and deployments
 - **Content Depot** — with smooth scaling from tens of terabytes up to forty petabytes, long-term storage and file preservation, content indexing, rapid retrieval of content and support for a wide range of content sources for providing a vast, searchable repository of information
 - **Archives** — with data integrity checks, encryption, access control, audit logs and accommodation of infrastructure changes without disruption or compromise of content for ensuring compliance with regulations, internal mandates and legal requirements
 - **Backup Reduction** — with highly reliable replication and dynamic data protection, single instancing (deduplication), data compression and direct user access to multiple versions of content for faster and more granular recovery, reduced backup times and data volumes, and less dependence on tape-based infrastructure
-

Figure 9. Hitachi Content Platform helps facilitate one platform for all data, saving time and money.



Hitachi Content Platform connects to a wide range of applications, via HTTP, REST, NFS and CIFS to name a few, so that organizations can seamlessly share and move data without compromising security or completely rewriting critical applications. Combining an assortment of protocols with multitenancy and support for many storage types, including Hitachi Virtual Storage Platform, Hitachi Adaptable Modular Storage and 3rd-party storage, means that multiple applications can run on a single cluster with Content Platform. And Content Platform ensures high availability and fault tolerance with node and path failover, plus no single point of failure, while offering multiple data protection formats to meet IT specifications.

By avoiding constraints found in traditional file systems, Content Platform elegantly handles the avalanche of unstructured data and its growing importance to the business with near-bottomless scalability. For organizations that will actively archive data or manage multitenant environments, Content Platform touts 40PB of capacity and billions of objects per cluster. It also offers state-of-the-art data reduction technologies, allowing data to comfortably and securely reside for decades or forever.

Content Platform helps enterprise IT mitigate much of the risk and costs associated with controlling data growth and managing content for decades or even centuries. Designed to preserve and protect content while easing the burden of unstructured data, Content Platform cost-efficiently scales with the business as well as compulsory regulations. A closer look at the capabilities of Content Platform reveals an overall value to IT organizations in need of reliability, flexibility and efficiency.

Worry-free Replication and Data Integrity

Hitachi Content Platform helps IT organizations shore up business continuity with a wealth of data protection qualities, including RAID-6, replication, data integrity checks, encryption, WORM, audit logging and multiple versions of objects, which is known as version awareness. These contribute to the enterprise's ability to manage and enforce policies for retention, disposal, shredding and other compliance and lifecycle management functions. With extensive tools and more granular control to ensure data is properly administered, Content Platform safeguards a wide variety of important tasks, such as distributing content for local access, ensuring that location-specific data stays put and complying with data placement regulations. These tools enable the IT organization to cost-efficiently grow business horizontally to support multiple applications and content types while scaling vertically to support continued data growth.

Because Hitachi Content Platform is remarkably well protected and always available, the need to perform traditional tape backup for unstructured data simply goes away. Content Platform uses advanced replication to protect content and also assist with control and sharing of content among distributed IT environments. So consumers of cloud and users at branch or remote offices can be confident that their data is placed in the right physical location and can be instantly accessed by the appropriate users. In the edge-to-core solution, any file written to Hitachi Data Ingestor is first assigned its retention attributes and is then copied to Content Platform, where it becomes an object. A stub is anchored at the local site to ensure access and recovery of the most recent version of the stored object. The data now becomes easily recoverable by the user, in a self-service, on-demand way, which also helps shrink help desk or IT costs and avoid lengthy tape-based retrievals and recovery.

Flexibility with Multitenancy and Namespaces

Hitachi Content Platform simplifies how IT will deal with data and its parts; file data, file metadata and any custom metadata are collectively treated as a single object. As a result, objects can be easily tracked and stored across a variety of storage tiers. The cost of object storage can also be economically spread across numerous use cases and workloads as needed to leverage cloud or other distributed environments. A single Content Platform supports 1000 tenants, with each "virtual content platform," or tenant, able to be uniquely configured for its specific level of service and access control without a buildup of storage silos. The tenants can then be subdivided into namespaces, up to 10,000 in total, for more granular organization of content, policies and access. This is accomplished through multitenancy and the configurable attributes (e.g., retention, indexing and replication) for each namespace within a tenant.

Storage Efficiency with Compression and Deduplication

For greater storage efficiency, Hitachi Content Platform uses data deduplication and compression to control data size by eliminating copies and decreasing the amount of storage needed for content. As new objects are written to Content Platform, the content is weighed against similar objects to avoid duplication. Unnecessary duplicate data is then eliminated or compressed to save space.

If the core infrastructure is replicated for disaster recovery purposes, selective replication can be used to minimize the amount of data at replica sites and conserve valuable replication bandwidth. Controlling overall storage consumption then streamlines failover to secondary systems and the recovery of primary systems once the failure is fixed.

Tap into Tangible Benefits

The Hitachi edge-to-core storage solution offers a surplus of benefits for distributed IT environments, especially for cloud deployments and ROBO environments. From the enterprise out, the ability to centralize control over security policies, availability, migration and all other key factors associated with managing content across a dispersed organization, is the crux of cost-effectively achieving business objectives. For many enterprises, and certainly all cloud service providers, time is money. In other words, attaining financial viability and economies of scale requires the organization to adequately plan for evolutionary transitions and efficiently execute them. Evolving from siloed storage and cumbersome IT practices at the edge, to an elegant, economically minded, all-encompassing integration with the Hitachi edge-to-core storage solution accomplishes just that.

The Silver Lining for Cloud Deployments

Faster time to market, faster time to value and simpler ways to bring users and applications on board are what cloud service providers need from their intelligent object store solution. Hitachi Data Systems delivers all of these in one unified package. Being able to rapidly alleviate the issues of security, performance, data protection, reporting and chargeback accounting helps service providers more readily offer cloud services to external customers. By providing practically endless and backup-free storage at the edge through a physical or virtual on-ramp appliance, Hitachi edge-to-core storage solution helps connect users and applications to advanced cloud storage and data management capabilities. As a result, IT organizations can deliver a wider berth of reliable, trustworthy storage services without having to construct their own edge-to-core infrastructure.

No Longer "Out on a Limb" for Branch and Remote Offices

No IT leader wants to dole out precious resources without an advantageous return on assets or investments. For distributed IT environments, particularly branch offices or remote sites, being able to successfully control content, backups and storage growth for edge locations can add significant strain to cost, complexity and risk. Unlike traditional edge storage that can become isolated, sprawled or not fully protected, the Hitachi edge-to-core storage solution bolsters scalability, preservation and security of local storage without the need to continually grow the IT footprint at the edge. Hitachi Data Ingestor uses only a small amount of local storage in order to serve as local cache for data that needs to be copied or retrieved from the core (where data actually resides). By simplifying and reducing the IT footprint at the edge with Data Ingestor, users and applications can communicate impeccably with centralized storage infrastructure to instantly access or share content, and even recover data on demand without help from IT.

Shrink Tape-based Dependencies and Costs

For both cloud and ROBO environments, the ability to withdraw from requisite tape backup can dramatically reduce the financial and liability burdens usually present with tape media. Data being generated at distributed locations is written to Hitachi Data Ingestor, which is then immediately and safely replicated and stored within the confines of Hitachi Content Platform. This integrated and highly scalable process eliminates the need to store or maintain tape-based equipment at each distributed locale. Furthermore, with the ability to control replication and other attributes granularly, at the namespace level within the core infrastructure, users can determine protection levels needed for

individual applications or particular data sets. For cloud service providers, the ability to avoid costly, labor-intensive tape media requirements can quickly translate into savings that are passed directly to their cloud consumers.

Following the cost savings upstream, once data resides on Hitachi Content Platform, the reliance on tape-based backup and related infrastructure can be further diminished. By leveraging its data retention and disposition services to automatically keep and dispose of content as prescribed, Content Platform is able to reclaim or recycle capacity back into viable storage. In sharp contrast to conventional methods of keeping every file forever, Content Platform creates a thorough audit trail of these processes to meet regulatory, legal and economic requirements while trimming the amount of data storage required.

Gain Control over Unstructured, Unruly Content

Agile data centers and IT facilities are crucial to effectively harness the wild growth of unstructured content. The matchless combination of Hitachi Data Ingestor and Hitachi Content Platform fosters that agility for unstructured content environments. Hitachi edge-to-core storage employs a single optimized storage platform that is capable of buffering back-end storage complexities and standardizing how storage is managed for multifaceted applications and workloads. IT administrators are finally able to adapt to changes in scale, technologies and infrastructure without time-consuming or insurmountable problems to solve. Instead of trolling among dispersed storage islands to migrate, manage or maneuver content, IT staff can centrally control how and where content is placed within the infrastructure. Hitachi edge-to-core storage automates supervision and enforcement of data lifecycle and governance policies. It safeguards content with an arsenal of data quality tools such as encryption, WORM, replication, RAID-6, data integrity checks and advanced protocols, including REST.

Simplify Storage Management and Savings

"Simplify and Save" is likely a mantra for almost any organization, distributed or not. Managing backup and infrastructure at numerous locations outside of headquarters can escalate costs and complexity. And with a growing assortment of application requirements and protocols, as well as protection levels, many IT departments will opt for segregating technology for each workload or geographical location. Storage sprawl quickly reaches a crescendo that is difficult to properly manage.

When it comes to addressing storage creep and costly silos, Hitachi edge-to-core storage is very cost-effective. Integrating Hitachi Data Ingestor and Hitachi Content Platform allows for support of a wide range of workloads and distributed users through a unified shared infrastructure. In this way, IT administrators gain greater visibility across the distributed environment. This facilitates a cohesive look at what data and metadata are actively stored, more cost-effective tiering and the ability to better uphold policies. By deploying Data Ingestor at the edge and capitalizing on the scale and functionality of Content Platform as the centralized core, organizations lower the IT footprint at distributed locations. This enhances service levels, simplifies storage management and reduces the total costs of ownership.

On the Horizon of Change

With the Hitachi edge-to-core storage solution, the flanks of once beleaguered IT organizations are now well equipped to battle the known and unknown issues of unstructured content growth. The integrated combination of Hitachi Data Ingestor and Hitachi Content Platform provides one solution for many purposes. It helps to solve convoluted matters, such as archiving and compliance, business continuity and replication, cloud computing and content depots, and backup reduction and disaster recovery. With massive, nondisruptive scalability, multiple storage tiers, Hitachi reliability, secure multitenancy and configurable attributes for thousands of tenants, IT administrators can support a wide range of applications on a single cluster. Particularly suited for cloud services and branch and remote office environments, Hitachi edge-to-core storage delivers on the promise of mitigating costs, complexities and risks often inherent in these distributed IT deployments.

For more information on Hitachi edge-to-core storage solution, please visit <http://www.hds.com/products/file-and-content/data-ingestor.html>

 **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

Hitachi is a registered trademark of Hitachi, Ltd., in the United States and other countries. Hitachi Data Systems is a registered trademark and service mark of Hitachi, Ltd., in the United States and other countries.

All other trademarks, service marks and company names in this document or website are properties of their respective owners.

Notice: This document is for informational purposes only, and does not set forth any warranty, expressed or implied, concerning any equipment or service offered or to be offered by Hitachi Data Systems Corporation.

© Hitachi Data Systems Corporation 2012. All Rights Reserved. WP-392-F VA April 2012