

# Hitachi Content Archive Platform

An Active Archive Platform for the Long Term Preservation,  
Optimization and Discovery of Unstructured Business Critical  
Digital Content

# Fixed Content Archive Solutions for a Crowded Digital World

Companies of all sizes across a broad spectrum of industries face the challenge of effectively managing and preserving the proliferation of fixed content data in a secure, unalterable format. The increasing pressure on corporations to reduce operational costs while complying with regulatory and corporate governance guidelines must be met with innovative solutions that effectively manage the retention and protection of fixed content business records and artifacts such as reports, e-mail records, medical records, scanned documents, digital voice recordings and other digital content.

The amount of digital content created today continues to increase at a near exponential rate. This growth rate alone makes offline archival media management extremely manpower intensive. A ripple effect is often seen as the combined costs of storage, backup and management climb to meet this demand.

An effective archival storage platform cost effectively stores the ever increasing amounts of inactive fixed content data online while delivering scalability of storage capacity without having to “throw more servers or storage nodes at the problem.” An intelligent “active” archival platform provides rapid content search and retrieval of infrequently referenced, business critical, fixed content data. Active archives enable the storage of data in a

secure, unaltered format where the data can be verified and authenticated (under audit or under state and federal evidence codes) as unchanged with a validated “chain of custody.”

## Hitachi Content Archive Platform Answers Fixed Content Challenges

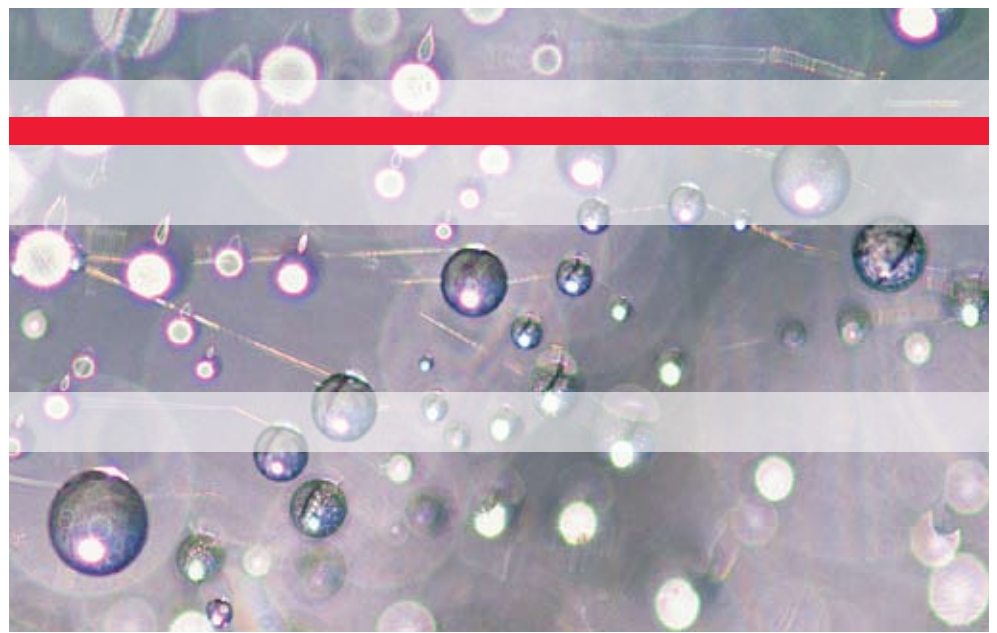
The Hitachi Content Archive Platform is designed to address and fix all of these challenges. This robust active archive preserves and protects fixed content data as verifiably unchanged for long periods of time, optimizes performance and availability to handle content ingestion from a host of application vendors or custom applications and can rapidly search and discover both content and metadata.

To manage the explosive growth of archival data, the Hitachi Content Archive Platform is available in Redundant Array of Independent Nodes (RAIN) appliances [from 2 to 42.5 terabytes (TB)] or via SAN Array of Independent Nodes (SAIN) configurations that cost effectively scale to 40 petabytes (PB) without requiring more storage infrastructure and reconfiguration as storage nodes are added. The Hitachi Content Archive Platform indexes, stores and searches corporate fixed content data (and their associated metadata) using business defined retention and protection policies.

The Hitachi Content Archive Platform is ideal for moving inactive data from expensive higher tiered storage to reduce backup costs, dramatically improving discovery time and meeting both corporate governance and regulatory compliance objectives for preservation and retention of critical data. Whether the need is legal discovery, regulatory audits, restoring capacity on tiered storage, improving backup windows or improving customer service levels, the Hitachi Content Archive Platform provides the benefits of a robust active archive for businesses of all sizes at a total cost of ownership (TCO) that is often equal to or better than tape or optical archives.

## Long Term Data Preservation and Protection

Most traditional types of data are transaction-based and have a relatively short period of usability. The applications that record the rise and fall of a stock price or the value of goods in a digital database are subject to constant modification and/or deletion at the click of a mouse. But not all data can be allowed to



change or disappear so easily. For many years now, organizations have been required to preserve transaction records for “dynamically changing digital data” for much longer than the data is useful to the business.

Only in the past several years have regulations emerged that also require the long term preservation of “fixed content,” such as e-mail messages, financial statements, medical images, Microsoft® Office documents, digital audio, video, Web 2.0 content and banking images that do not change over time. As a general trend, this fixed content data is increasingly required to be preserved over specific periods of time.

### Fixed Content Data

Fixed content data is best defined as data that is not intended to change. Fixed content data includes digital records of real world events

that have happened at specific points in time. Examples would include an X-ray image, an e-mail message, a completed digital video, a filing made to a government agency or even the recent influx of Web 2.0 content — such as wiki and blog content. For the fixed content data to remain valuable in the future, it must remain fixed to accurately reflect its original state. For most organizations, fixed content data composes the bulk of all storage needs. Although analyst estimates vary, fixed content data is generally thought to represent between 80 percent and 90 percent of all storage capacity. It is also generally accepted that fixed content data is being created much faster than dynamically changing transaction data. Thus, fixed content digital data growth is driving demand for long term archive storage capacity (see Figure 1).

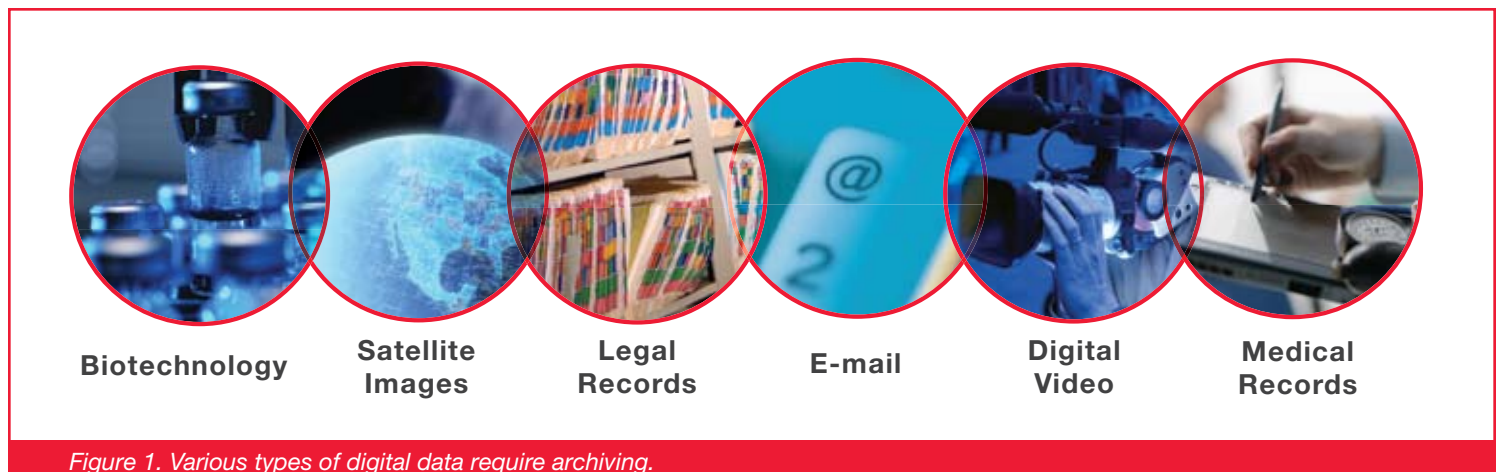


Figure 1. Various types of digital data require archiving.

## Why Retain Fixed Content Data?

Prior to the advent of digital recording media, physical fixed content data was indexed, cataloged and filed in formats such as paper, microfiche, video, audio, film, optical disk and backup tapes to provide a permanent archive that could be accessed when needed. However, certain types of digital fixed content, such as e-mail and instant messages, may not have a physical counterpart since many e-mail and instant messages are rarely printed and are simply left on digital recording media. Since government regulations cover fixed content digital data such as e-mail and instant messages, and since these types of digital data have no physical counterpart, maintaining a permanent record of digital fixed content data is becoming a priority for organizations seeking to achieve regulatory compliance.

The list of reasons for preserving fixed content digital data grows continually. This data provides a historical record not only for the organization, but also for the auditors and regulators who enforce compliance with data retention statutes and work with their clients to ensure best archiving practices for a wide range of industries. Organizations ignoring these requirements may face stiff penalties and legal consequences.

Fact: E-mail is now considered a standard source of evidence in legal proceedings. Searches through e-mail archives are a routine first step in any legal discovery process. Organizations that do not have rapid access to historical e-mail messages are hindered during litigation and saddled with the disruption and expense of manually searching through tape backups to satisfy discovery. On occasion, discovery costs and any associated disruption to operations can effectively force a business to settle a dispute they could easily have won with timely fixed content data production.

## The Challenge: Disk-based "Active Archiving" of Fixed Content

For some industries the growing demand for disk-based active archive digital data storage is forcing IT organizations to rethink conventional storage architectures, since storing digital archival data on top-of-the-line storage systems is usually not cost effective.

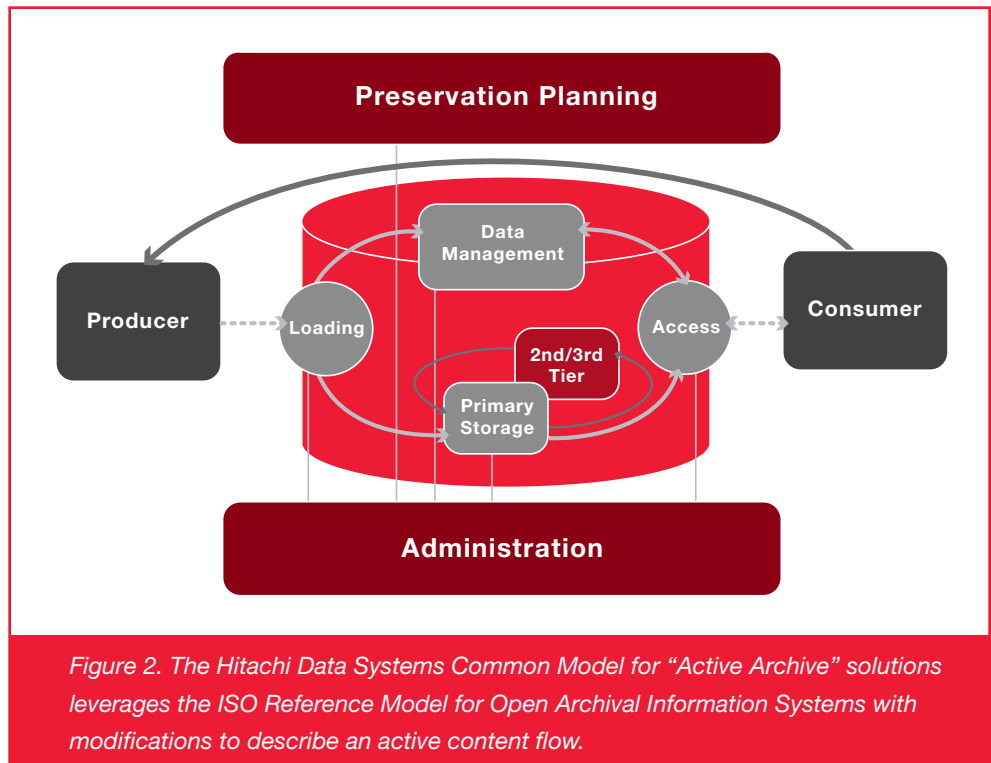


Figure 2. The Hitachi Data Systems Common Model for "Active Archive" solutions leverages the ISO Reference Model for Open Archival Information Systems with modifications to describe an active content flow.

Not all of today's enterprise architectures were designed with long term digital data archives in mind. Information on conventional disk storage systems is usually considered working data and often is characteristically archived to very inexpensive media such as tape or optical media as part of routine IT operations. Unfortunately, tape and optical media do not always satisfy the requirements of today's regulatory environment, since tape and optical media are often manpower intensive to access and slow to search.

Storing archived fixed content information in online RAID protected active archive storage systems offers an alternative to traditional disk with offline backup or offline archival storage using tape, microfiche, optical disk or paper as archival media. An online disk-based active archive can not only improve customer service responsiveness by making infrequently needed data available faster in response to a customer question, but it can also improve online relational database performance by taking the infrequently needed data out of the relational database. An active archive can also reduce the disruptive (business and IT) impact of all types of auditing (including Sarbanes-Oxley) and litigation-based discovery.

## What Would a Modern Disk-based Active Archive Look Like?

In the physical world, the long term preservation of important records and artifacts is the domain of library sciences and archiving. Practitioners in these disciplines have given much thought to the problem of data preservation and have developed several general theories and best practices. Traditional archives provide a repository for records the organization has selected to preserve.

**The archive serves two fundamental goals:**

- Records must be preserved unchanged.
- Records must be easily accessed.

These properties apply equally well to archives containing fixed content. Disk-based active archives are designed to provide long term preservation, authenticity, access and performance over that of alternatives such as standard disk, tape, optical or paper options. The choice archive determines the levels of archival data access (see Figure 2).

## Ingesting Information into the Archive

Fixed content digital data is created by a wide variety of enterprise and custom applications. The digital archive ingestion process allows multiple applications to stream data into the archive simultaneously for long term preservation and

storage. Individual items become archival objects in a metadata software repository for each fixed content record. In this way, the metadata can be quickly searched to retrieve the archived items that are needed for customer service, audit or compliance purposes. The ingest process also passes along a defined retention period with the content. Retention periods ensure that content cannot be deleted before the retention period ends. As an independent software vendor (ISV) pushes data into the archive via one protocol, the data can then be accessed by any of Hitachi Content Archive Platform's supporting protocols.

### **Authenticity and the Chain of Custody**

The contents of a digital archive are only of value if authenticity of each record is guaranteed. In the world of library sciences, the term "provenance" refers to the ability to maintain the origin and chain of custody of an archived record.

An item can only be certified as authentic if the chain of custody is unbroken. This means that the item has been securely managed since it entered the archive. If an item's chain of custody cannot be proven, its reliability as evidence is significantly reduced.

Digital records face the same provenance problems as traditional archived items. For example, an auditor assessing an organization's compliance with the Sarbanes-Oxley Act is unlikely to vouch for the authenticity of archived e-mail records if it cannot be proven that the content was not tampered with.

Archiving applications trigger the content (e-mail, file and/or database instance) to be published as archived content to the archive platform. The archiving application then authenticates the digital records and stores them in "write once, read many" (WORM) format within the archive. This ultimately prevents unauthorized modification or deletion of the archived content.

### **Preserving Digital Fixed Content Data**

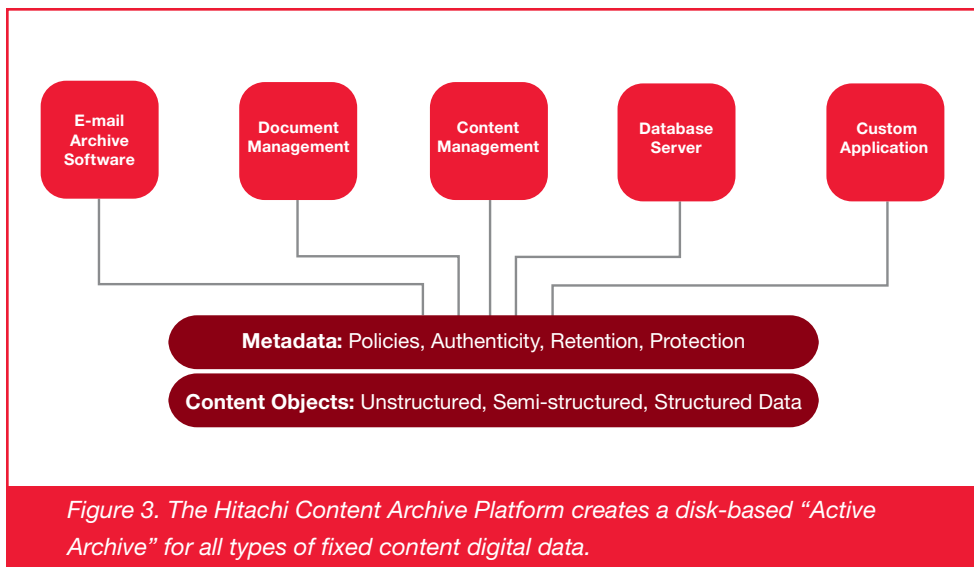
To guarantee the authenticity and accessibility of stored digital records many years after they were originally created, every component of the archive must be capable of being

## Use Case: Tiered Storage and Archiving

The recent escalation in demand for enterprise storage capacity has led many IT organizations to consolidate and centralize storage resources. Storage area network (SAN) and network attached storage (NAS) technologies have produced significant efficiency improvements, allowing administrators to manage much more capacity using a common set of storage management tools and procedures.

In these new consolidated environments, IT planners favor storage solutions that integrate seamlessly with the existing tiered storage infrastructures and with existing enterprise content management (ECM), business process management (BPM) and compliance software platforms. Archival solutions that require a unique management and configuration approach lower administrator productivity, reduce efficiency and raise the overall cost of storage.

Considering a digital archive's potential to consume enormous amounts of storage capacity, it is imperative that the archive integrates seamlessly into an existing IT infrastructure. In this way, IT management can ensure that the archive remains cost effective as it scales up for growth. Support for storage networking and common storage management tools will allow an archive to be treated as simply another tier in a multitiered pool of efficiently managed storage.



The Hitachi Content Archive Platform provides:

- A WORM file system and time-based retention and chain of custody at the file or object level
- Open standards-based interfaces for custom and ISV application ingestion
- Data protection levels that ensure a specified number of replica copies are maintained to tolerate simultaneous points of failure
- Automated object-level remote replication
- Custom metadata support
- Content privacy (optional feature for encryption of data at rest)
- Duplicate data elimination and file level compression for greater storage efficiency and scale
- Authenticated content preservation with a user choice of digital signature or hash algorithms
- Integrated, secure management features that monitor system health and reduce costs by simplifying maintenance activities
- Support for Hitachi Device Manager and Tuning Manager software for reporting of Content Archive Platform metrics

upgraded as new technologies become available. For example, encryption algorithms used today will likely be useless five or ten years from now, as faster computer processors allow existing routines to be cracked.

The open, standards-based interfaces of the digital archive ensure that stored content is available for access using technology that is current at the time the search is conducted. This can be a tremendous challenge when data formats and application versions are continually changing. An essential feature of the digital archive is the ability to upgrade hardware, software and encryption routines without jeopardizing the chain of custody of fixed content data. This also means that before and after copies of each data transformation must be kept as an audit trail so subsequent researchers can determine if any unauthorized changes took place during conversion.

### Discovery and Retrieval

The ability to successfully search and retrieve stored records is a key function of an archive. If records cannot be located they are effectively lost, and all attempts to preserve the content are wasted. A digital archive is almost certain to contain millions of records; therefore, efficient, high-performance search and retrieval of information is critical.

Whether an archive is conventional or digital, the curators and managers of the information face a common problem: it is almost impossible to anticipate how future users of

the archive will want to query the information. Some users will know exactly what information they are looking for and will be able to use specific indexes to locate items. Other users will need to browse the archive looking for relevant information. Digital archives can also support data mining technology, allowing researchers to gain insights into archived content unavailable through other search and retrieval mechanisms.

## Hitachi Content Archive Platform

The Hitachi Content Archive Platform is a robust active archive that enables effective long term, fixed content data preservation for businesses of all sizes. Designed to seamlessly integrate into an existing enterprise storage infrastructure, the high-performance, high availability, highly scalable archiving solution satisfies an organization's regulatory compliance requirement by ensuring the secure, long term preservation and fast search and retrieval of valuable business records (see Figure 3).

The Hitachi Content Archive Platform is an integrated software and hardware solution that provides extreme scalability, high availability and high performance, and satisfies the growing demand for long term fixed content storage management. With built-in authentication, protection and retention capabilities, this enterprise-class platform guarantees archived content will be continually accessible for years to come.

### Hitachi Content Archive Platform 300

Hitachi Content Archive Platform 300 is a highly available RAIN architecture with embedded storage and RAID protection on each node with no single point of failure. This affordable archive solution is available in configurations from 2TB to 42.5TB that meet mid-sized enterprise or departmental archive requirements.

### Hitachi Content Archive Platform 500

Hitachi Content Archive Platform 500 is an enterprise-scalable active archive solution built around the Hitachi SAIN architecture. The solution leverages Hitachi SAN storage solutions, existing storage management and business continuity processes from Hitachi Data Systems.

- Highly scalable — up to 40PB of total data archiving.
- Support for Hitachi Universal Storage Platform™ V, Hitachi Universal Storage Platform VM, Hitachi Universal Storage Platform, Hitachi Network Storage

Controller, Hitachi Adaptable Modular Storage and Hitachi Workgroup Modular Storage

- High availability based on SAIN architecture
- Optional Hitachi Content Archive Platform Search provides embedded full text index, search and retrieval for content discovery
- Configurable content privacy (encryption of data at rest utilizing patent pending “secret sharing” technology)

## Business Benefits

The Hitachi Content Archive Platform, when coupled with a custom or ISV application, delivers clear business benefits:

- Migrate content from existing or overloaded high cost storage and file servers. Move inactive data from expensive higher storage tiers, reduce backup costs, dramatically improve discovery time and meet both corporate governance and regulatory compliance objectives for preservation and retention of critical data.
- Preserve and protect critical data via a high availability environment, long term file integrity and access, authentication and retention enforcement.
- Meet regulatory and governance requirements. Ensure compliance with requirements for content preservation and retention while protecting your business and reducing risk.
- Save time and money. Reduce cost and complexity via a single active archive repository; reduce duplicate copies of the same data.
- Rapidly discover and retrieve authenticated content with integrated Hitachi Content Archive Platform Search functionality.\*
- Grow with your business. Scale horizontally to support multiple applications and content types; scale vertically to support continued data growth.

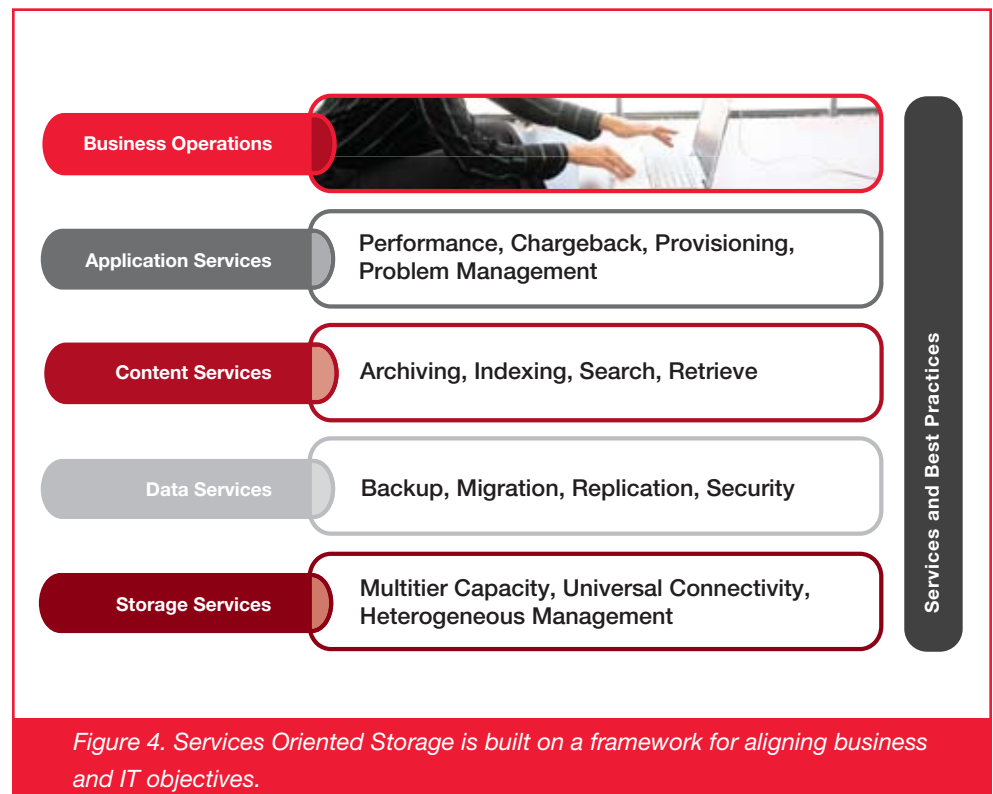
## Services Oriented Storage Solutions Align Business and IT Objectives

**Services Oriented Storage Solutions** from Hitachi Data Systems provide an integrated approach to developing a storage infrastructure based on business requirements rather than technology features. The storage services can be applied to data as needed in a multi-tier, heterogeneous storage infrastructure supported by application, content, data and storage services. **Services Oriented Storage** enables the storage infrastructure to respond to the specific performance, availability, functionality and cost requirements of each application (see Figure 4). By delivering capacity and services based on the specific needs of the business, **Services Oriented Storage** can substantially lower both operating expense (OpEx) and capital expense (CapEx) for the enterprise.

Since **Services Oriented Storage** can be applied on an “as needed” basis to various storage tiers, the Hitachi Content Archive Platform can seamlessly be integrated into almost any existing tiered storage infrastructure or enterprise software architecture.

## Implementing the Hitachi Content Archive Platform

The Hitachi Data Systems Global Solution Services (GSS) team can help organizations of all sizes design and implement the optimal Hitachi Content Archive Platform configuration to meet specific business and application requirements.



\*Available on Hitachi Content Archive Platform 500

**Corporate Headquarters** 750 Central Expressway, Santa Clara, California 95050-2627 USA  
Contact Information: + 1 408 970 1000 [www.hds.com](http://www.hds.com) / [info@hds.com](mailto:info@hds.com)

**Asia Pacific and Americas** 750 Central Expressway, Santa Clara, California 95050-2627 USA  
Contact Information: + 1 408 970 1000 [www.hds.com](http://www.hds.com) / [info@hds.com](mailto:info@hds.com)

**Europe Headquarters** Sefton Park, Stoke Poges, Buckinghamshire SL2 4HD United Kingdom  
Contact Information: + 44 (0) 1753 618000 [www.hds.com](http://www.hds.com) / [info.emea@hds.com](mailto:info.emea@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 on this Web site 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. This document describes some capabilities that are conditioned on a maintenance contract with Hitachi Data Systems being in effect and that may be configuration dependent, and features that may not be currently available. Contact your local Hitachi Data Systems sales office for information on feature and product availability.