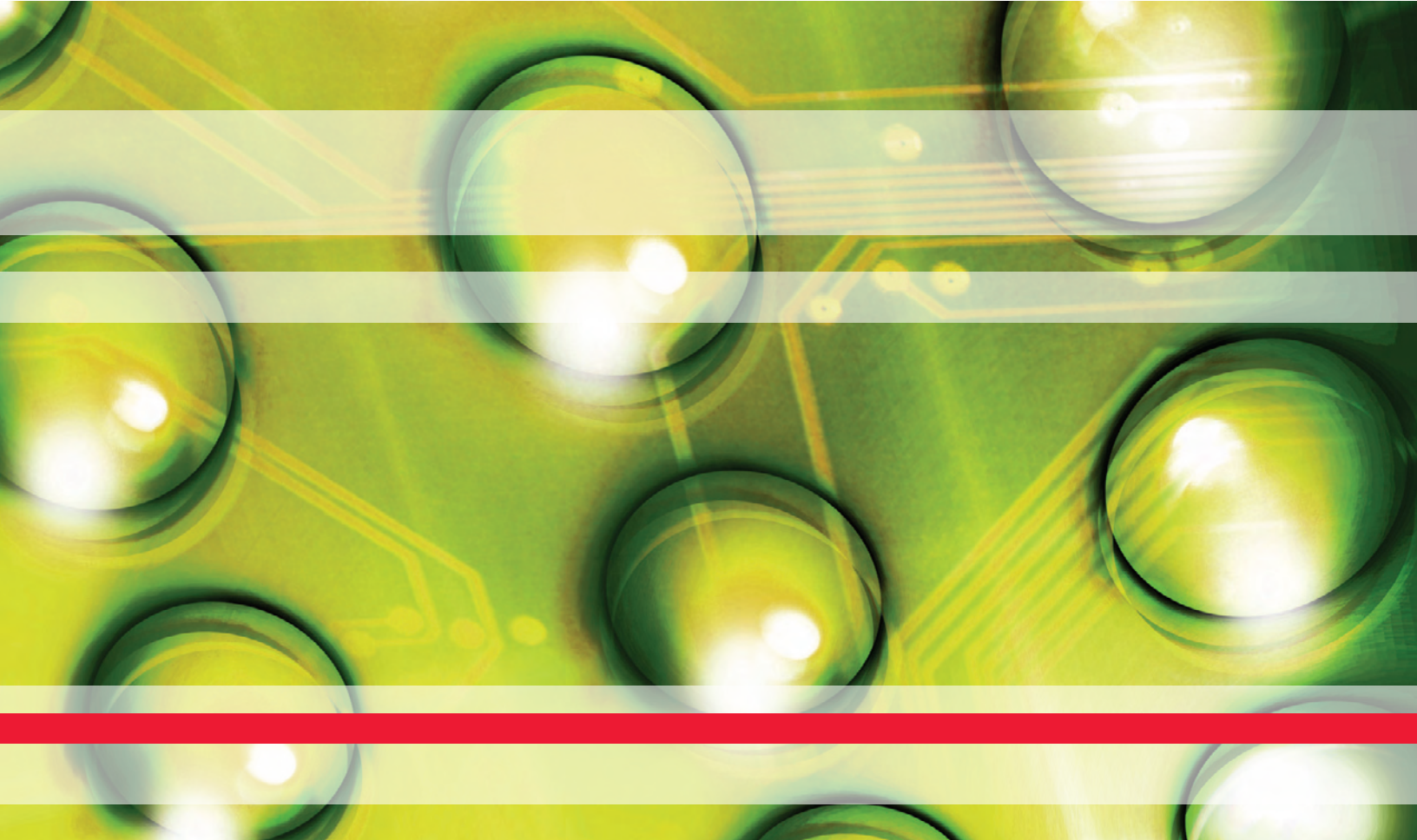


Hitachi Content Archive Platform

An Active Archive Blueprint for Long-term Preservation
of Business Critical Digital Fixed Content and Data



Fixed Content Archive Solutions for a Crowded Digital World

Companies of varying sizes across a broad spectrum of industries face the challenge of effectively managing and preserving the proliferation of business critical, “fixed content” data (data that is not intended to change) in an unalterable online digital format. This is compounded by the increasing pressure for corporations to both meet corporate governance guidelines and reduce the operational impact of internal and external audits related to government regulations regarding the retention and protection of fixed content business records (reports, e-mail records, medical records, digital recordings Web 2.0 content, etc.)

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

A robust 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 and storage nodes at the problem.” An intelligent “active” archival platform provides rapid content search and retrieval of infrequently referenced, busi-

ness critical, fixed content data. An “active archive” provides the archiving of data in a secure, unaltered format where the data can be verified (under audit or under state and federal evidence codes) to be unchanged with a validated “chain of custody.”

Hitachi Content Archive Platform Answers Fixed Content Challenges

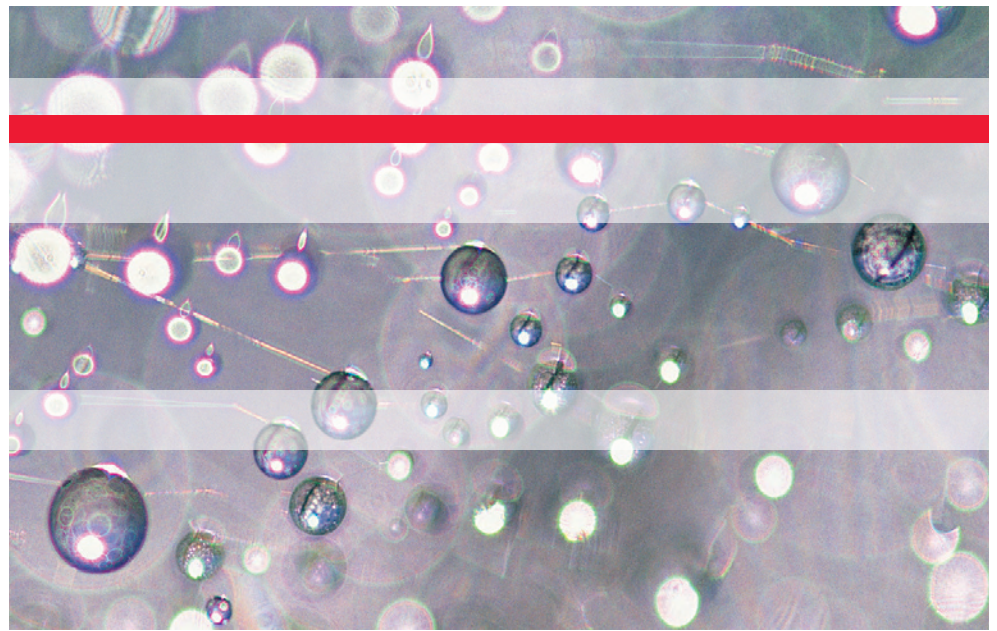
The Hitachi Content Archive Platform is designed to answer all of these challenges. This robust “active” archive both preserves and authenticates fixed content data as being verifiably unchanged for long periods of time, while optimizing performance and availability

to handle the explosive growth of data. The Hitachi Content Archive Platform can be used to move inactive data from expensive higher tiered storage to reduce backup windows and improve response times, as well as to meet both corporate governance and regulatory compliance objectives.

To manage the explosive growth of archival data, the Hitachi Content Archive Platform is available in Redundant Array of independent Nodes (RAIN) models (from 2 to 8 terabytes) or via SAN Array of Independent Nodes (SAIN) models that cost-effectively scale to multiple petabytes without requiring more storage infrastructure and reconfiguration as storage nodes are added. With Hitachi Content Archiver software, this centralized online archive can index, store and search corporate fixed content objects (and their associated metadata) using business defined retention and protection policies. 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 more often equal to or better than tape or optical archives.

Long-term Data Preservation

Most traditional types of data are transaction based, and have a relatively short period of usefulness. The ones and zeroes that record the rise and fall of a stock price or the value of goods in a digital database are subject to constant modification by applications, 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, certain types of organizations have been required to preserve transaction logs for “dynamically changing digital data” for much longer than the data is useful to the business.

Only in the past five years have regulations emerged that also require the long-term preservation of “fixed content digital data,” such as Web pages, e-mail messages, corporate presentations, medical images, Microsoft® Office documents, digital audio and video, and check images that don’t change over time. As a general trend, this fixed content data is increasingly being required to be preserved over long 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 comprises 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.

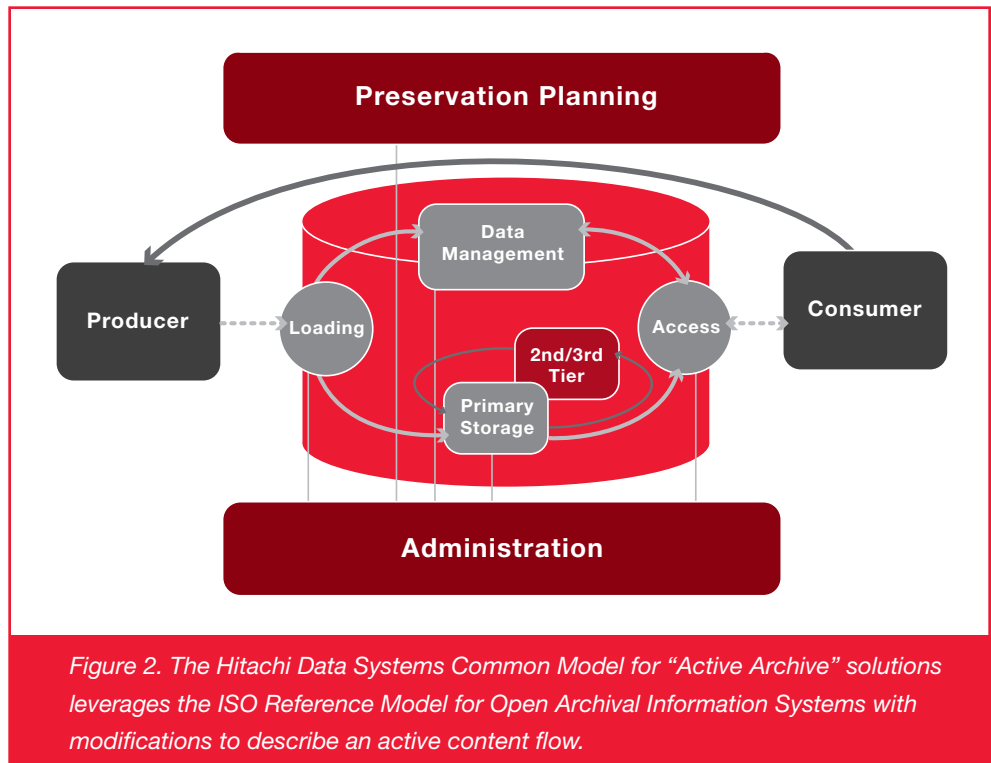
Why Retain Fixed Content Data?

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 formats 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 for example, 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 retaining 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 can face stiff penalties and legal consequences. 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 ready access to historical e-mail messages can be not only handicapped during litigation, but also saddled with the disruption and expense of manually searching through tape backups to satisfy discovery. Sometimes discovery costs and their disruption to operations can effectively force a business to settle a dispute that 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.



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 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?

Rather than reinvent the wheel, IT planners can refer to substantial existing research on the records management process for archiving, whether the archived digital data is to be stored in a passive archive on tape or optical media, or in an active archive on disk. 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, and records must be easily accessed. These properties apply equally well to archives containing fixed content. Of course the choice of digital media selection for the archive will determine the ease of archival data access. These questions can only be answered on an industry-by-industry and enterprise-by-enterprise basis.

Ingesting Information to the Archive

Fixed content digital data is created by a wide variety of enterprise 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 associates a customer defined retention period with the content. Retention periods ensure that content cannot be changed or deleted from the archive until a predetermined period of time has passed. In a traditional archive, the ingestion of records into the archive represents a physical and legal node in the chain of custody of the content.

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 vouch for 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, which means that the item has always been securely managed since it was loaded into 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 has not been tampered with.

As the data is loaded into the platform, business process archiving software triggers the e-mail, file, and/or database instance to be published as archived content to the digital archive. The archiving software then authenticates the digital records and stores them in a "write once, read many" (WORM) file system. This 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

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.

For optimal enterprise architecture, a tiered storage approach to archival data can be as important to consider as a tiered storage approach to online operational data. Disk-based fixed content digital data archiving and tape and/or optical fixed content digital data archiving, therefore, are examples of various archival storage tiers. The data to be placed on each are determined by what makes the most financial sense for the business. (See Figure 2.)

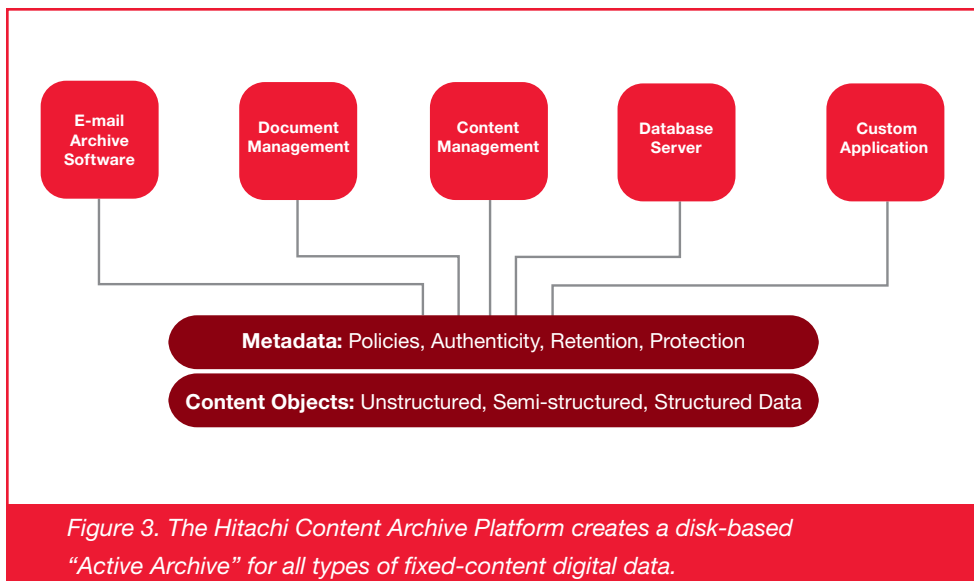


Figure 3. The Hitachi Content Archive Platform creates a disk-based “Active Archive” for all types of fixed-content digital data.

were originally created, every component of the archive must be capable of being 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, data format 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 that subsequent researchers can determine if any unauthorized changes took place during conversion.

Search 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 many 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 Platforms use world class Hitachi Content Archiver software to provide scalability, availability and performance, satisfying the growing demand for long-term fixed content storage management. With built-in authentication, protection and retention capabilities, the highly available platform guarantees archived content will be continually available for access for years to come.

The Hitachi Content Archive Platform provides:

- A WORM file system and time-base retention at the object level
- Automated object-level remote replication
- Custom metadata support
- Content privacy (encryption of data at rest)
- Standards-based interfaces, including NFS, CIFS/SMB, HTTP, https, WebDAV, SMTP, NDMP
- Duplicate data elimination
- Authenticated content preservation with a user choice of digital signature or hash algorithms
- Support for Hitachi Device Manager and Tuning Manager software for reporting of Content Archive Platform metrics

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 solution is available in affordable configurations from 2TB to 8TB 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 both existing storage management and business continuity processes from Hitachi Data Systems.

- Highly scalable with up to 32 billion objects (such as files or e-mails), and up to 20 petabytes 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 (reliable SAN storage with an array of independent nodes)
- Optional 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 Platform delivers clear business benefits:

- **Migrate content from tiered storage.** Move static data to an active archive tier and provide a high availability environment, long-term file integrity and access, authentication and retention enforcement.
- **Protect your business.** Reduce risk of fines and penalties for not meeting regulatory compliance or providing timely responses to audit and legal discovery requests.
- **Meet regulatory and governance requirements.** Ensure compliance with requirements for content preservation and retention.
- **Save time and money.** Reduce cost and complexity via a single active archive repository; reduce duplicate copies of the same data.
- **Expedite content retrieval.** Rapidly retrieve authenticated content through Hitachi Content Archive Platform integrated search.*
- **Protect and secure content for long-term preservation.** Content is continually checked throughout its retention period for integrity.
- **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 multitier, heterogeneous storage infrastructure supported by application, content, data and storage services. **Services Oriented Storage Solutions** enable the storage infrastructure to respond to the specific performance, availability, functionality and cost requirements of

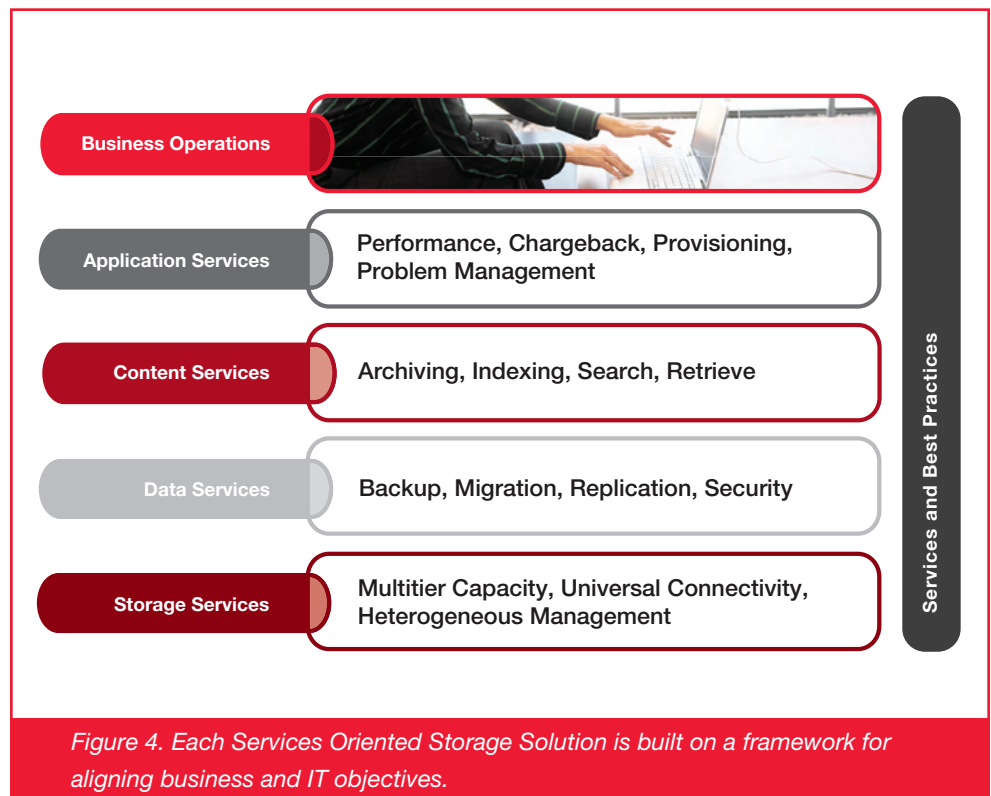
*Available on Hitachi Content Archive Platform 500

each application (see Figure 4). By delivering capacity and services based on the specific needs of the business, **Services Oriented Storage Solutions** can substantially lower both operating expense (OpEx) and capital expense (CapEx) for the enterprise.

Since **Services Oriented Storage Solutions** 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.



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

Europe Headquarters Sefton Park, Stoke Poges, Buckinghamshire SL2 4HD United Kingdom
Contact Information: + 44 (0) 1753 618000 www.hds.com / info.uk@hds.com

Hitachi is a registered trademark of Hitachi, Ltd., and/or its affiliates 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.

Universal Storage Platform is a trademark of Hitachi Data Systems Corporation.

Microsoft is a registered trademark of Microsoft Corporation.

All other trademarks, service marks and company names are properties of their respective owners.

Notice: This document is for informational purposes only, and does not set forth any warranty, express 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.

Hitachi Data Systems sells and licenses its products subject to certain terms and conditions, including limited warranties. To see a copy of these terms and conditions prior to purchase or license, please go to <http://www.hds.com/corporate/legal/index.html> or call your local sales representative to obtain a printed copy. If you purchase or license the product, you are deemed to have accepted these terms and conditions.

© Hitachi Data Systems Corporation 2008. All Rights Reserved.
SB-007-A DG July 2008