

DATA DRIVEN GLOBAL VISION CLOUD PLATFORM STRATEG  
ON POWERFUL RELEVANT PERFORMANCE SOLUTION CLO  
VIRTUAL BIG DATA SOLUTION ROI FLEXIBLE DATA DRIVEN V

WHITE PAPER

## Information Is Everywhere, But How Do You Extract the Knowledge?

### Make NAS Part of Your Big Data Solution

By Hitachi Data Systems

February 2014

## Contents

<b>Executive Summary and Introduction</b>	<b>3</b>
<b>Create a Storage Solution for the Future: Less Is More and Agility Is Key</b>	<b>4</b>
Limitations of Today's Storage Solutions	5
Move Beyond the Status Quo to Deliver Knowledge Efficiently	6
<b>The Ideal Unified Storage Solution to Big Data</b>	<b>6</b>
<b>HDS Unified Storage Solutions Meet the Big Data Challenge</b>	<b>7</b>
Hitachi Unified Storage and Hitachi NAS Platform	7
<b>Summary: Hitachi Helps Extract Knowledge From Big Data and Drive Competitive Advantage</b>	<b>9</b>

# Information Is Everywhere, But How Do You Extract the Knowledge?

## Executive Summary and Introduction

For the last 40 years, it's likely that your main data challenge was using structured data to enable and record faster business transactions. Today, the focus is shifting. Organizations now have to enable larger and faster exchanges of both structured and unstructured data. (Such data may include audio and video files, medical images, movies, gene sequences, data streams, social media and so forth). The amount and variety of this new data flowing into organizations is creating extraordinary opportunities. Organizations can extract more business value in sectors that range from energy, human genomics and healthcare to retail, online search, surveillance and finance. However, with all of this information coming from so many new sources, how do you extract the knowledge?

In this new age of big data, the greatest business challenge is to find better ways to analyze, monetize and capitalize on all your information. This challenge requires improvements to the underlying platforms, from scale-out cloud clusters to unified storage infrastructures. More efficient unified storage system solutions are necessary to ensure that all of your big data is reliably and efficiently ingested, stored, protected, organized, accessed and preserved.



Hitachi Data Systems believes organizations that can make real-time business decisions based on this data at the lowest possible cost will thrive. We also believe that those organizations unable to embrace and use this expanding data will increasingly find themselves at a competitive disadvantage in the market. This disadvantage is particularly evident in industries that are experiencing high rates of business change and aggressive consolidation.

In this paper, we examine issues organizations are experiencing with storing and extracting knowledge from big data, including rising costs and limitations of inefficient storage infrastructures. We also explore the benefits of migrating enterprise data to unified and network-attached storage (NAS) systems. Such migration can greatly enhance performance and scalability, increase capacity efficiency, improve manageability and drive significant cost savings.

## Create a Storage Solution for the Future: Less Is More and Agility Is Key

The archives and data stores that most IT departments have deployed over the past couple of decades have served their purpose. Deploying these siloed solutions made sense when budgets were tight and data growth, most of it structured, was incremental. However, as we move into the era of big data, those same solutions are proving to be far too expensive and inefficient.

Storage efficiency is important and challenging today. Organizations rely on an increasing range of devices, data sources and applications that create a seemingly endless array of data types. Semistructured and unstructured data flows from mobile devices, Web 2.0 applications, financial transactions, audio, video and dozens of other sources into your systems. From barcodes to websites, organizations have more data sources than ever, which means more information to analyze to determine what their customers want and when they want it.



In the face of this onslaught of big data, IT is trying harder than ever to meet its service level agreements (SLAs). IT must help get products to market quicker, provision compute resources faster and deliver more storage services sooner. IT must also enable the business to extract the knowledge it needs from all of that information. That's not easy to do when IT has to answer a litany of questions about where and how to store data from an ever-increasing range of sources:

- Which data has the most value?
- Does the data need to be stored for the short, medium or long term?
- What level of data protection is needed? Is flash included?
- What is the most cost-effective way to store the data?

As pressures on IT departments have mounted, the most common response has been to deploy more archives and data stores to retain and manage different types of data. While that traditional approach to enterprise storage has worked up to this point, its limitations are becoming clear.

### IS YOUR STORAGE SOLUTION AS EFFICIENT AS IT COULD BE?

Signs of storage inefficiency include:

- Large amounts of duplicate data.
- Wasted capacity, poor utilization.
- Insufficient data protection and security.
- Unmanaged growth of cold and archival data on primary systems.
- Out-of-control storage costs.
- Limited administrative resources.
- Inconsistent I/O response times.
- Long backups and recoveries.
- Trouble meeting recovery time objectives (RTO).
- Unnecessary energy consumption.
- Increasing RAID rebuild times.
- Inability to deliver high availability.
- Time-consuming storage provisioning.
- Inability to reduce or remove obsolete data.

### Limitations of Today's Storage Solutions

Traditional storage systems involve various disparate silos, or islands, of data in different formats, each dedicated to a single purpose, person or application. These siloed storage systems are notorious for operating far below their optimal utilization. Because they operate separately, one storage system's unused capacity is unavailable to another that's running out of capacity.

IT can do little in response to these utilization issues, except buy more capacity, which leads to an inefficient use of floor space, wasted energy and increasing operational and capacity costs. Further, although adding infrastructure answers immediate needs, it also adds to size, cost and complexity of the overall storage solution. When problems occur with large, inefficient storage infrastructures, the added complexity means longer downtime, which in turn leads to increased costs that can run into the millions of dollars.

Over time, several more problems emerge with traditional storage solutions. The 1st is that they lack a common storage management interface. This problem drives up operational expenses, greatly hinders the speed of provisioning more capacity and makes it nearly impossible to deploy storage on demand. A 2nd challenge is that siloed storage solutions also make it difficult to prioritize data based on business needs, balance workloads and secure different types of data.

A 3rd challenge is that organizations are struggling with ever-tightening backup windows. As the volume and variety of data, particularly unstructured data, increases, backups take longer. Adding to the frustration is that much of the data is file data that is seldom changed, and many files are duplicates. Thus, businesses with siloed storage solutions

end up backing up and protecting the same data again and again. This practice threatens backup windows and adds to spiraling operational and capital expenses.

Finally, traditional storage systems are holding many organizations back, in part because the systems are unable to support the advanced analytics necessary to derive knowledge from big data. That makes it nearly impossible to respond quickly to market trends or make informed decisions based on the real-time information that big data affords.

## Move Beyond the Status Quo to Deliver Knowledge Efficiently

Organizations today must focus on building a storage platform for the future that will enable them to gain business knowledge and insight from big data. To do that, begin by finding an effective way to capture, store and manage the underlying data in a cost-effective manner.

To escape the cycle of declining performance and rising costs, your business cannot continue with the status quo. There is no way to maintain traditional storage solutions in an efficient and cost-effective way while taking advantage of the opportunities made available by big data. Instead, it's time to find more efficient, higher-performing unified storage solutions that improve uptime, protect data better and speed application development, all while lowering costs.

## The Ideal Unified Storage Solution to Big Data

Fortunately, there are proven storage solutions available today that can provide the capacity, security, management functionality and performance you need, even as big data increases exponentially. Consider integration of features such as flash, solid-state disks (SSDs), hard-disk drives (HDDs), automated tiering and migration, and data reduction technologies, such as deduplication. Today's most efficient storage systems can integrate these features to help you manage the rise in big data while limiting complexity, without driving up costs.

What you need is a storage solution provider that goes beyond consolidating storage. You need a provider that can deliver an extremely efficient, dynamically allocated set of tiered storage pools that support block, file and content data. To get this level of storage proficiency, you'll want to look for unified storage solutions with these essential features:

- **Scalability.** It's no longer cost-effective or efficient to keep adding storage to meet ever-increasing storage needs. Instead, solutions should be highly scalable from the start, both in performance and capacity, to meet your big data needs now and far into the future.
- **Capacity efficiency.** The ideal big data storage solution will use advanced features, such as deduplication, space-efficient snapshots and automated tiering to ensure that you maximize your existing storage capacity.
- **Flexibility.** It's a safe assumption that soon your business will be handling more data of more types, so your storage solution should be purpose-built for diverse, high-volume workloads. Make sure it will work in various environments, including virtualized server environments.
- **Manageability.** For a low total cost of ownership (TCO), you need a storage solution that is easily manageable. Look for a proven unified management software solution and features such as high-speed replications and automated system failover.

With the right big data unified storage solution in place, your organization will enjoy benefits in both the short and long term. In the short term, you'll see greater capacity utilization, lower energy costs, improved availability and much improved performance. Over the long haul, you'll be able to redeploy IT staff from managing the storage infrastructure to other business-forward initiatives. You'll also find yourself deferring anticipated purchases of storage infrastructure, reducing hardware and software maintenance costs, and finding more ways to use big data to fuel business growth.

### KEY STORAGE CAPABILITIES

The ideal big data storage solution allows you to extract knowledge efficiently and features advanced storage capabilities, including:

- Data deduplication.
  - Removes duplicate copies of data blocks automatically, without impacting workloads.
  - Increases storage capacity efficiency.
- Automated file tiering.
  - Dynamically moves data to optimal storage tier based on policies.
  - Lowers capital and operating expenses.
- Space-efficient snapshots.
  - Creates space-efficient file copies.
  - Streamlines access to stored data and speeds data recovery.
- File cloning.
  - Creates space-efficient, writeable file copies.
  - Backup alternative saves space by eliminating need to duplicate data.
- Object-based replication.
  - Reduces amount of data handled and transferred.
  - Enables high-performance replication for disaster recovery at scale.

## HDS Unified Storage Solutions Meet the Big Data Challenge

Hitachi has been developing unified storage solutions to answer the big data challenge since well before "big data" became an industry catchphrase. We bring best-in-industry technology and an earned reputation for performance, scalability and reliability to the challenge. The result is innovative big data unified storage solutions that deliver significant advantages in performance, scalability and economics.

### Hitachi Unified Storage and Hitachi NAS Platform

Hitachi Unified Storage (HUS) and Hitachi NAS Platform (HNAS) provide a unified storage solution for all types of data. The centralized, scalable infrastructure meets the high-performance demands of businesses facing accelerated growth in file-based big data and associated applications.

HUS and HNAS help IT departments support large unstructured data sets and automate the organization of unstructured data, all with a backup-less, active-archive solution. The products offer maximum flexibility, providing help with file sharing, file server consolidation, data protection and application data storage for Oracle, Microsoft® and virtual server environments.

**HNAS  
Power and  
Efficiency**

[READ MORE](#)

### WHAT IS UNIFIED STORAGE AND NETWORK-ATTACHED STORAGE?

Unified storage provides comprehensive and integrated management for block, file and object data. It delivers high-density storage, which reduces data center floor space consumption. It also offers automated data management with dynamic provisioning, dynamic load balancing and autotiering, all while scaling capacity and performance.

Network-attached storage (NAS) is data storage that provides file sharing, file server consolidation, data protection, data management and application data storage. NAS devices can be thought of as dedicated servers used to manage an enterprise's storage infrastructure. Compared to file servers, NAS devices provide dedicated faster data access, easier management and simpler configuration.

Several core capabilities within HUS and HNAS enable businesses to achieve the highest possible utilization and optimal TCO:

- **Consolidate more.** Consolidate file data into fewer systems, which reduces storage sprawl, lowers power and cooling costs, provides simpler data management and enables much greater control over file growth.
- **Scale faster.** Scale up to 8 nodes in a single cluster, add storage without downtime and support up to 32PB under a single namespace.
- **Manage less.** A single, centralized system management unit reduces administrative complexity without compromising performance or scalability.

Advanced features enable HUS and HNAS to excel in a wide range of environments, from enterprise data centers to medium organizations and remote data centers. These features include file tiering, deduplication and various data protection features:

- **Intelligent file tiering.** Automatically place data on the appropriate tier of storage for optimal storage capacity utilization.
- **Primary storage deduplication.** Extend the life of existing storage assets by eliminating redundant data, resulting in up to 90% capacity savings.
- **Data protection.** Synchronous disaster recovery support, high-speed snapshots and high-speed replication of objects combine to simplify and enhance data protection efforts.

In addition to supporting Microsoft Windows® consolidations and Oracle environments, HUS and HNAS are uniquely designed and optimized for virtual workloads and environments. Supporting features include file cloning and simplified management through Hitachi NAS Virtual Infrastructure Integrator. Storage systems within HNAS offload common virtualization workloads. This action frees up critical I/O and reduces overhead on the back-end storage system.



### BENEFITS OF HITACHI UNIFIED STORAGE AND HITACHI NAS PLATFORM

- Powerful performance.
  - Offers highest usable capacity and performance per node in the industry.
  - Provides industry-leading scalability to meet needs of large, mixed, big data workload environments.
  - Speeds time to test and deploy virtual applications.
  - Improves productivity and application performance.
- Enhanced efficiency.
  - Reduces complexity and simplifies management.
  - Increases utilization, with greater usable NAS capacity.
  - Features fast, nondisruptive deduplication.
  - Minimizes interruption of data access with advanced data protection and fast recovery.
- Superior TCO.
  - Lowers TCO by 30% or more.
  - Eliminates redundant data and defers need to purchase more capacity.
  - Reduces required storage capacity by up to 90% with deduplication.

## Summary: Hitachi Helps Extract Knowledge From Big Data and Drive Competitive Advantage

To turn the structured and unstructured data you collect daily into knowledge that can help your business grow, you need more than traditional storage systems can provide. High costs, poor manageability, lack of scalability and other limitations of siloed solutions hinder business performance. These challenges can prevent you from taking full advantage of the enormous opportunities of big data.

Hitachi helps you extract knowledge from your data by starting with a highly scalable unified storage platform that can handle multiple data types. Then, we add object-based metadata to help with automated data classification, distributed management and customization. Finally, our fast search tool works in an open, heterogeneous, multiserver environment designed to index and search files and metadata at petabyte scale. Hitachi is the only vendor to provide all 3 tools integrated within a single portfolio and a common management interface.

With our Hitachi Unified Storage and Hitachi NAS Platform, Hitachi Data Discovery Suite search engine, and Hitachi Content Platform, Hitachi Data Systems can help you leverage big data for competitive differentiation. Our future-ready big data storage solutions span from midrange to high end, featuring an industry-leading object-based file system operating environment and a full suite of intelligent management tools. Our complete storage ecosystem is designed to ensure storage efficiency. It enables IT to provision faster, allocate resources when needed and provide better overall service, while lowering the TCO of the entire storage infrastructure.

For organizations with large volumes of unstructured data, HUS and HNAS, combined with our experience and leadership in the storage industry, make Hitachi Data Systems an ideal partner. We help you turn big data into knowledge efficiently. We help you drive competitive advantage that will spur business growth and innovation for years to come.

**@Hitachi Data Systems**



**Corporate Headquarters**

2845 Lafayette Street  
Santa Clara, CA 95050-2639 USA  
[www.HDS.com](http://www.HDS.com) [community.HDS.com](http://community.HDS.com)

**Regional Contact Information**

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

© Hitachi Data Systems Corporation 2014. All rights reserved. HITACHI is a trademark or registered trademark of Hitachi, Ltd. Microsoft and Windows are a trademarks or registered trademarks 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, expressed or implied, concerning any equipment or service offered or to be offered by Hitachi Data Systems Corporation.

WP-475-A DG February 2014