

## WHITE PAPER

---

# Scale Out Storage In the Content Driven Enterprise: Unleashing the Value of Information Assets

Sponsored by: Hewlett Packard

---

Richard L. Villars

June 2011

## EXECUTIVE SUMMARY

The continued expansion of file-based, business-critical information within extended enterprises is changing the storage dynamic in a wide range of industries and organizations. As extracting value from these information assets become a core business mandate, they are placing a high priority on boosting the efficiency and reliability of their platforms for collecting, storing, protecting, and delivering file-based information in what IDC calls content depots. The top file-based storage challenges that these organizations face with their existing dispersed file-based storage assets (file servers and standard NAS devices) include:

- ☒ Inefficient use of storage assets (low utilization overall with a few performance hot spots as well as inability to leverage high capacity/low cost storage)
- ☒ Inefficient, unreliable, and non-compliant data backup/retention processes
- ☒ Continuous need for administrator intervention
- ☒ Expensive upgrade/replacement cycles for file servers and NAS systems

The explosion in content depot creation and expansion with its unique use and retention requirements is driving the adoption of new solutions (e.g., scale out NAS platforms like HP's X9000 IBRIX Storage System) that optimize storage utilization and reduce the burdens associated with to data protection and business continuity. In conversations with organizations deploying scale out NAS solutions, the top benefits included:

- ☒ Faster and non disruptive addition of capacity or performance in more cost effective increments (days, not months)
- ☒ Support for automatic data tiering to maximize asset use (reducing the need for administrator intervention by up to 90%)
- ☒ Better use of high capacity/low cost storage (delaying spending on high performance storage and cutting capacity acquisition cost by over 50%)
- ☒ Minimize backup loads (Double or triple effective utilization rate while reducing backup requirements by up to 80%)

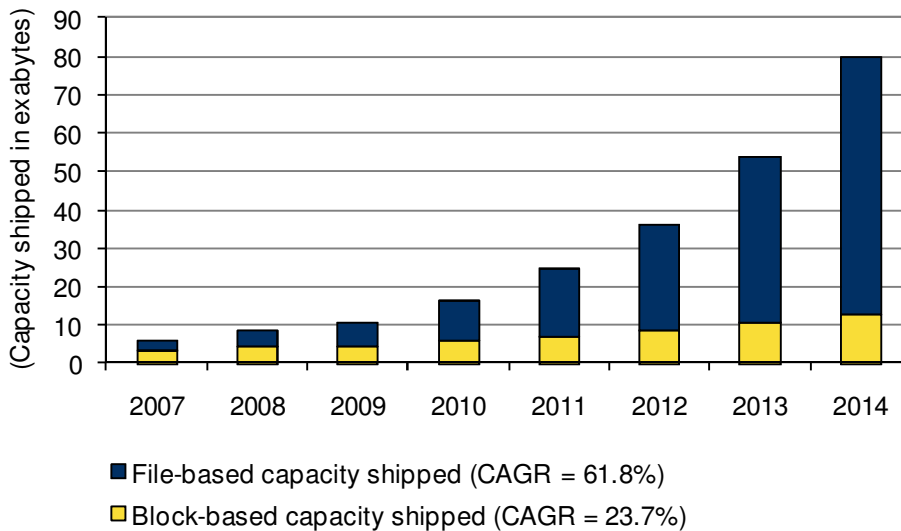
IDC expects companies like HP to continue developing and delivering a progressively growing number of tools and practices to help information's guardians – the CIOs of the world – deal with the here-and-now of their own particular content depots.

## RISE OF THE CONTENT DRIVEN ENTERPRISE

The creation, organization, and distribution of files and rich content are driving most of the growth in organizations' storage investments. In 2011, organizations around the world will spend \$16.4 billion on storage solutions to store the documents, images, videos, and machine generated data records created by their employees, partners, and customers. This level of spending translated into 17,300 Petabytes (PB) of new disk storage capacity deployed in 2011 for file-based data. By comparison, block-based storage capacity shipped to store traditional structure data sets was 7,200 PB (See Figure 1).

**FIGURE 1**

Worldwide Block-Based Vs File-Based Storage Forecast, Capacity Shipments, 2007-2014



Source: IDC's 2010 Enterprise Disk Storage Consumption Model

Given current spending and capacity growth rates, organizations worldwide will be spending \$22.5 billion in 2014 in order to deploy 67,145 Petabytes (67 Exabytes) of file-based storage capacity to store all this information. The main question for IT executives, however, isn't, "Are we deploying too much capacity?" Instead, most will be asking two questions:

- Are we deploying enough capacity?

- ☒ Are we using that capacity effectively?
  - ☒ Are we managing/administrating that fast growing capacity efficiently?
- 

## **Forces Driving the Information Explosion**

Organizations ranging from small and medium-sized businesses (SMBs) to the largest enterprises and service providers are experiencing an explosion in the volume of content generated, accessed and stored. Some of the forces driving this explosion apply equally across all sectors, while others pertain to specific industry transformations.

### ***Diversifying Information Assets and Uses***

Organizations ranging from small and medium-sized businesses (SMBs) to large enterprise datacenters are:

- ☒ Relying on email, collaboration tools, and Web sites to communicate and conduct business with customers and business partners.
- ☒ Collecting, storing, and analyzing more information about products, customers, and transactions.
- ☒ Digitizing records, design documents, and other types of unstructured data to boost efficiency, offer new services, and comply with evolving government regulations.

This increasing dependence on unstructured data types (e.g., documents, records, images, and video) for competitive differentiation and more effective operations is spurring an explosion in demand for file-based storage capacity versus more traditional bloc-based storage capacity. In turn, the expanded deployment of file-based storage systems is posing a number of challenges for IT and storage administrators:

- ☒ Inefficient use of storage assets due to uneven capacity and performance demands (low utilization overall with a few performance hot spots as well as inability to leverage high capacity/low cost storage)
- ☒ Inefficient, unreliable, and non-compliant data backup/retention processes
- ☒ Continuous need for administrator intervention to rebalance and reconfigure file pools based on load and size
- ☒ Long and expensive upgrade/replacement cycles for file servers and NAS systems

### ***Long Term Value of Information (Archiving)***

Organizations are increasingly creating, storing, and accessing more information for long periods of time. Historically, the primary, if not the only motivation for firms that are consciously archiving information has been the need to stay compliant with

regulations that demanded information retention for extended periods of time. Over time, firms have come under increasing pressure to not only store and retain large amounts of data in archives but also to quickly retrieve this information when needed.

For a growing number of organizations, however, access to archived information is now also critical for ongoing operations (e.g. doctor access to patient medical records), business analysis (e.g., mining of customer buying behaviors), and new business development (e.g., repurposing of movie archives, for new media launches).

Firms are mining their historical data to analyze and extract data for market intelligence, product planning, and inventory planning. In R&D environments, reuse of historical information can yield vast savings in time and effort, which in turn saves money and in some cases provide competitive advantage by shrinking the time required to bring products to the market.

In this new universe of active archiving, simply moving content to tape archives is ineffective and/or impractical. IT organizations need to provide continuous access to active archive information to exploit new business opportunities while reducing the costs of data retrieval. Their customers and governmental organizations will demand storage solutions that quickly deliver the right information to the right person in a timely fashion. They will also want that information to be secured from inappropriate use from permanent loss or extended unavailability, and yet easily accessible for reuse and analysis.

### ***Intelligent Industries and Industry Digitization***

The most profound and long lasting driver of the content explosion is current and still expanding "digitization" of entire industries: These include:

- ☒ Media/entertainment (e.g., HD conversion, digital production, digital delivery)
- ☒ Healthcare (e.g, digital x-rays, electronic medical records)
- ☒ Life sciences (genomic sequencing, protein sequencing)
- ☒ Legal services (eDiscovery and review)
- ☒ Mobile communications (apps for smart phones and tablets)
- ☒ Utilities (smart grid)
- ☒ Public Cloud Services Providers (delivering millions of virtual desktop and server images and operations)

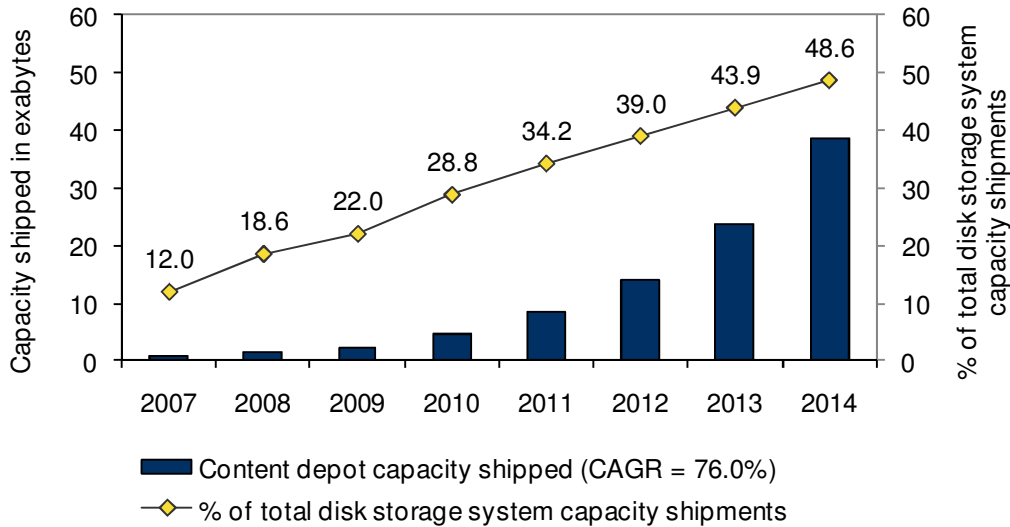
These developments are driving an IT investment wave by "intelligent industries" and cloud services providers. All of these organizations' businesses are almost entirely based on the storage, organization, long-term retention, timely retrieval, and near continuous analysis of data and archived unstructured content.

The primary job of IT teams in these industries is building, managing, and continually expanding large content depots. As an aggregate community, these content depots accounted for 28.8% of all storage capacity shipped in 2010, and by 2014, they will

account for 48.6% of all data center storage capacity shipped into organizations (See Figure 2).

**FIGURE 2**

Worldwide Content Depot and Public Cloud Storage Forecast, Capacity and Percent of Market, 2007-2014



Source: IDC's 2010 Enterprise Disk Storage Consumption Model

While specific requirements will vary by industry (e.g., media/entertainment companies that need to concurrently stream a song or movie to millions of mobile users versus a utility that is ingesting sensor data from hundreds of thousands of meters every 5 minutes), they all need storage solutions that can:

- ☒ Achieve economies of scale, operational efficiency, comprehensive disaster recovery and business continuity, and information governance far beyond that achievable with dispersed file servers and standard NAS systems
- ☒ Comply with best practices and government regulations by properly retaining and storing file type information over predetermined periods of time
- ☒ Reduce management complexity and eliminate performance and capacity management constraints to align storage infrastructure investments with the demands of the business
- ☒ Enhance (aka monetize) the value of unstructured data through more sophisticated business intelligence and analytics solutions

## Limitations of Existing Storage Solutions For Large Content Environments

The continued expansion in information creation, consumption, and retention will have a dramatic impact on enterprises across a wide range of industries and geographies in the next five years. In the very near future, the management, organization, and continuous mining of large content pools will become the primary task for many storage administrators in corporate datacenters.

This shift is already having a significant impact (and it will only grow over time) on how companies assess storage solutions in terms of systems' performance, operational efficiency, and systems intelligence. Areas where many existing storage solutions are part of the problem, not part of the solution, include:

- ☒ Dispersed storage assets (primarily file servers and traditional NAS platforms) within data centers and across the extended enterprise
- ☒ Inability to intelligently optimize/tier information assets based on either performance or capacity requirements
- ☒ Difficulties in classifying, reorganizing, and rapidly analyzing original content and metadata associated with newly created and archival files
- ☒ Limited (and expensive options) to set up business continuity, disaster recovery, and data migration processes and policies for very large (multi-petabyte) content pools

Organizations dealing with the explosion of content as well as the need to more effectively store and use that content are demanding more targeted, role-based storage solutions designed to meet their ever more diverse storage and information management needs.

## PICKING THE RIGHT STORAGE SOLUTION FOR CONTENT

While not all companies, regardless of size, are in industries experiencing rapid file-base data growth, the scale of content growth across all companies is significant. In the very near future, your organization will need to consider new approaches for dealing with data growth and active access to archive data.

For some companies, especially small and mid-size organizations, this transition will involve greater use of cloud-based storage services. For many others (especially organizations in content-driven industries as well as cloud-based storage service providers themselves), it will require deployment of new storage solutions optimized for content ingest, storage, and access. Despite the differences in size of environment, there are a set of fundamental, common elements that you need to consider when evaluating a scale-out NAS solution:

- ☒ **Multidimensional Scalability.** Solutions that offer elastic scalability will arm customers with the ability to manage fast expanding and varying loads. Many

content environments start with modest beginnings but can quickly overwhelm the initial environment. Scalability isn't just about expanding hardware capacity, however. In content intensive environments, scalability must also encompass throughput, file size, and file volume. The best solutions will allow you to scale each dimension independently to ensure the optimal configuration for your specific needs.

- ☒ **Storage efficiency.** Solutions that enable easy adoption of a continually expanding range of storage asset optimization technologies. Over time, these capabilities will include data reduction technologies (e.g., data deduplication, compression, thin provisioning) and data optimization technologies (e.g., automated data tiering with SSDs and capacity optimized HDDs). The best solutions will allow you to introduce or recalibrate these capabilities without major disruptions or system migrations.
- ☒ **Intelligent Information Management.** Solutions that allow organizations to continuously track data usage changes, analyze usage patterns, and reorganize information pools in near real-time. The best solutions will provide tight integration with advanced metadata, index, and analytics solutions (developed internally or provided by a third party developer) that enable more intelligent information management as volumes of data sets grow and new use cases proliferate.

The benefits to expect from effective deployment of such a scale out NAS solution include:

- ☒ Faster and non-disruptive addition of capacity or performance in more cost effective increments (days, not months)
- ☒ Support for automatic data tiering to maximize asset use (reducing the need for administrator intervention by up to 90%)
- ☒ Better leverage of high capacity/ low cost disk storage (reducing or delaying spending on high performance storage capacity and cutting average storage capacity acquisition cost by over 50%)
- ☒ Reduced backup loads (Double or triple effective utilization rate of backup assets while reducing backup time and resource requirements by up to 80%)

The remainder of this white paper examines the scale out storage solutions from Hewlett Packard (The HP X9000 IBRIX family of scale out storage systems) and assesses how well they address organizations' growing need for storage systems that address rapid and increasingly complex content creation, access, and archiving needs.

## HP CONVERGED STORAGE

Hewlett Packard (HP) is a worldwide leader in the development and delivery of IT products (e.g., servers, storage, and networks), software and professional services. HP's Storage Division is an innovator in the delivery of converged storage solutions

designed to address organizations' fast evolving data center business environments. HP defines Converged Storage as an architecture and portfolio derived from scale-out storage software delivered on industry standard blade server hardware and managed with converged storage, server, and networking orchestration software. Along with its business partners and HP's own services organization, HP provides storage solutions optimized to address:

- ☒ Virtual and converged IT infrastructure for data center transformation and private clouds
- ☒ Data deduplication and disk-based data protection for more efficient backup/recover and business continuity
- ☒ High volume, high velocity transactions and business analytics (Big Data)
- ☒ Scale-out storage and archiving for content driven organizations and large content depots

In addition, HP's Storage team works with vertically focused teams that address the full range of information management needs in industries such as media/entertainment, healthcare/life sciences, financial services, government, and public cloud services.

### **HP X9000 IBRIX Storage System: Delivering Scale Out NAS in Multiple Dimensions**

The foundation of HP's scale out storage solution is the IBRIX Storage System family. The X9000 IBRIX Storage System is a tightly integrated hardware and software solution designed to be simple, efficient and adaptable. HP's goal is to provide content depots, enterprise customers, and other large organizations with a better solution for managing explosive data growth and file sprawl.

The X9000 IBRIX Storage System provides massive capacity and performance scalability in an easy to manage platform. Today, all HP X9000 systems (V5.6) leverage the latest processors (G7), network links (10 GbE), and HDDs (6 Gbps SAS), build on a modular converged platform. HP currently delivers three different specialized hardware bundles:

- ☒ X9300 Gateway: This is a gateway solution that lets users easily leverage traditional and scale-out storage block-based storage (SAN) within the NAS scale out architecture. It includes out-of-the-box support for HP storage arrays, including HP 3PAR Utility Storage. When combined with 3PAR Storage Systems this Converged Utility solution inherits the thin provisioning and multi-tenant features included in 3Par Utility Storage which are required by many cloud-based data service providers.
- ☒ X9320: A Modular configuration that allows companies to start small and independently add capacity and performance as required. Leverages 10 GbE for high throughput and supports both high performance and high capacity SAS drives in conjunction with automated "in-the-box" file tiering (14 T to 192 TB capacity range)



- ☒ X9720: Also referred to as “extreme storage,” the X9720 features a very dense architecture with up to 1.3 PB (raw) per system in a single cabinet (customers can deploy a single system across multiple cabinets). It uses 1 or 2 TB disk drives and is enhanced for performance

HP's goal with the HP X9000 IBRIX Storage Systems is to allow users to create a virtual file storage environment where they can match storage price/performance with application requirements, yet manage a single, highly scalable system. All three appliances models, or multiple instances of them, can coexist within a single namespace. This gives users the ability to manage various types of storage with different performance profiles as storage pools in a single file system that scales to 16 PB and to 1024 nodes, providing shared storage resources that can grow, contract, and be optimized with uninterrupted data access

The key to meeting this goal is the X9000 IBRIX scale out file system software. In combination with these tuned hardware nodes, this software allows organizations to scale independently across multiple dimensions (capacity, I/O performance, throughput), while managing the entire content pool as a single entity. The advantages to this approach include:

- ☒ Reduced management staff overhead: organizations can manage petabytes of storage capacity with a single FTE administrator
- ☒ Highly tunable and scalable capacity and throughput levels: Organizations can deploy up to 16 petabytes in a single namespace, across 1000+ nodes.
- ☒ Rapid system expansion and upgrades: Administrators can non-disruptively add capacity and/or performance resources within minutes, so they no longer need to buy/overprovision capacity months or years in advance of actual need.
- ☒ Continuous and real time view of the past with IBRIX Snap: a versioning file system that allows users to access the file system as it appeared at the instance of the snapshot. The IBRIX Storage Systems support over 1M snapshots within the namespace, and snaps can be taken of complete file systems, directories or individual files. They are read only, immutable, and allow recovery of objects in event of deletion or for backup purposes.
- ☒ Reliable and usable data retention: Administrators can set up separate file systems including ones designated as write once, read many (WORM) Data that supports data validation (computes and saves checksum for each file) and reporting (stats for data retention and space utilization) that are required for specific data retention and privacy regulations.
- ☒ Evergreen (persistent) storage: Using the "in-the-box" auto-tiering feature, the system can quickly and non-disruptively move files from performance pools to capacity pools based on usage or data type policies. IT teams can also use this function to non-disruptively migrate files to new (higher performing or more cost effective) nodes as they are introduced.

A key element in HP's delivery of scale out file-based storage solutions is the recognition that individual vertical industries and specific user communities have

unique performance, information management, or archiving requirements. One specific market where HP the X9000 plays an important role is Healthcare Storage.

HP's developed strategic partnerships with a number of leading Healthcare PACS ISV's. It worked with them to create pre-tested and pre-certified converged medical storage solutions designed for continuous, secure, and cost effective medical record archiving and access. The solution leverages the X9000's built-in data movement technology to enable the right placement of records in the right place at the right time. It also allows organizations to dial in just the right amount of image cache storage versus archive storage as needs change.

HP also recognizes that long term archiving of both structured and unstructured data is a critical requirement for many organizations. HP and partners such as Commvault, Symantec, iTernity, and Quantum are working to make sure that companies can leverage the underlying data storage and management capabilities of the X9000 IBRIX Storage Solution in conjunction with advanced, intelligent archival software solutions.

---

## **Challenges and Opportunities**

Continued rapid data growth and the increasing role of rich content in organizations new application and services plans are posing ever changing storage and information management challenges for HP's content driven customers and prospects. HP needs to address a number of requirements as it expands its role in organizations' content intensive environments:

- ☒ Continue to improve underlying storage hardware capacity, performance, and power management efficiencies through technologies such as data compression, even more tunable/intelligent automated data tiering, and added support for higher performance SSD as well as even denser/more power efficient HDD solutions
- ☒ Establish closer technical and business ties with leading content creation, distribution, and analytics (Big Data) software solutions suppliers in all important vertical industries that will make it easier for customers to fully exploit the information stored within X9000s file systems. As one important execution point in this effort, HP plans to release integrated metadata search capabilities as part of the IBRIX software in the next 12 months
- ☒ Extend the reach of the HP solutions to better address the growing content needs of medium and small business as well as large enterprises. This effort should include expanding the ecosystem of cloud-based storage service provider partners.

## **ESSENTIAL GUIDANCE**

The digital assets of organizations around the world will continue to expand and they will spend more time and money collecting, storing, and monetizing rich pools of content for the foreseeable future. IDC expects companies like HP to develop and

deliver increasing number of tools and practices to help information's guardians – the CIOs of the world – deal with the here-and-now of their own particular content depots.

In the end, however, a CIO's challenges dealing with his or her own exploding universe of digital information will be not just technical – how to find information, manage it, and protect it – but also organizational. CIOs must take a leadership role in driving the adoption of the new, information-taming technologies and practices, like HP's scale out NAS storage solutions. Making the leap will require organizational change, not just a few new systems or more software. The success of many enterprises in the coming years will be determined by how successfully the CIO drives the required enterprise-wide adjustment to the new realities of the content driven enterprise.

---

### **Copyright Notice**

External Publication of IDC Information and Data — Any IDC information that is to be used in advertising, press releases, or promotional materials requires prior written approval from the appropriate IDC Vice President or Country Manager. A draft of the proposed document should accompany any such request. IDC reserves the right to deny approval of external usage for any reason.

Copyright 2011 IDC. Reproduction without written permission is completely forbidden.