

WHITE PAPER

Managing Data Growth and Monetizing Information Value: The Role of Scale-Out Storage Solutions in the Expanding Universe of Digital Information and Big Data

Sponsored by: EMC

Richard L. Villars

July 2012

IDC OPINION

We are living in a world in which billions of devices (PCs, laptops, smartphones, tablets) are being deployed — over 4 billion by 2015. When sensors generating log data are added to the mix, we are really talking about trillions of devices in the next decade. We are also dealing with an explosion in content. IDC expects the amount of capacity deployed to store all that content to grow more than sixfold over the next five years. All of these developments have some common characteristics: They generate more data. They tell us where we are. They tell us what we want. They tell us what we are doing.

The continued expansion of business-critical information and rich content within extended enterprises is changing the storage dynamic in a wide range of industries and organizations. Extracting value from this expanding universe of digital information is becoming a core business mandate. IT organizations must place a high priority on boosting the efficiency and reliability of their platforms for collecting, storing, protecting, and delivering this information.

The top storage challenges that your organization will face with existing dispersed storage assets (e.g., file servers and scale-up NAS devices) include:

- Inefficient use of deployed storage assets
- Continuous need for administrator intervention (rebalancing loads)
- Expensive and time-consuming upgrade/replacement cycles

For organizations across all industries, addressing these limitations requires the deployment of new classes of storage solutions (scale-out storage solutions) that are optimized for rapid data ingest, efficient storage management, and reliable access. Aggressive use of a scale-out storage solution such as EMC's Isilon portfolio of hardware and complementary software will allow your IT organization to implement a number of best practices that can boost storage asset use, reduce operational overhead, and meet high data availability expectations.

THE DATACENTER MEETS CONTENT IN THE EXPANDING DIGITAL UNIVERSE

File-based data sets (e.g., rich content, machine-generated sensor data, and virtual machine images) continue to grow at a rapid pace. In spite of unstable economic conditions that are having a significant impact on datacenter investments, investing in the storage solutions for large (multipetabyte) and fast-growing information assets remains a necessity in both established organizations and emerging cloud IT environments. The forces driving this explosive growth in file-based storage include:

- ☒ Rapid expansion of mobile and sensor devices in both consumer and industrial uses that are providing a constant feed of unstructured data
- ☒ Greater use of NAS solutions as the storage management layer supporting virtualized and convergent IT infrastructure environments
- ☒ Increased use of disk-based and cloud-based storage solutions (based on NAS platforms) for both backup/recovery and centralized data retention/archiving
- ☒ Increasingly diverse industries that are moving to a "digital first" or "digital only" business model, which requires buildout of highly scalable NAS solutions for data ingest, processing, archiving, and delivery
- ☒ Emergence of Big Data analytic environments that require ongoing but unpredictable access to large and rapidly proliferating data sets

The increasing dependence on unstructured data types (e.g., files, documents, records, images, call detail records, sensor logs, and video) for competitive differentiation as well as effective operations is causing an explosion in demand for NAS storage capacity versus more traditional block-based storage capacity. In 2012, spending on file-based storage solutions such as scale-up and scale-out NAS systems is expected to be \$13 billion. By 2015, new capacity shipments for file-based storage (86,290 petabytes) will be more than three times the size of block-based storage.

Limitations of Existing Storage Solutions for Large Content Environments

The rapid growth of traditional file-based storage environments based on dispersed file servers and scale-up NAS systems poses a number of challenges for datacenter executives.

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 tasks for many IT executives and administrators in corporate datacenters.

This shift is already having a significant impact on how companies assess storage solutions in terms of system performance, operational efficiency, and system intelligence. IDC spoke with IT administrators who indicated that existing storage solutions are often part of the problem, not part of the solution, in the following areas:

- ☒ Inefficient and uncoordinated dispersal of storage assets within datacenters and across the extended enterprise (Companies with large numbers of dispersed file servers and low-end NAS systems often find it difficult to boost effective utilization rates beyond 50% while also struggling to meet new regulatory/eDiscovery requirements for data disposition.)
- ☒ Inability to intelligently optimize/tier information assets based on either performance or capacity requirements (Organizations with growing pools of CAD documents, sensor logs, or video assets report situations where storage administrators spend more than half of their time constantly rebalancing content on multiple independent NAS systems to minimize data bottlenecks.)
- ☒ Difficulties in classifying, reorganizing, and rapidly analyzing original content and metadata associated with newly created and archived files (Organizations are unable to locate and effectively use historic transaction and file data for risk governance and business-expanding data analytics.)
- ☒ Limited (and expensive options) to set up business continuity, disaster recovery, and data migration processes and policies for very large (multipetabyte) content pools (IT staffs worry that backup times for content are exceeding acceptable time windows while they are also experiencing an explosion in demand for secondary storage capacity to back up old, infrequently accessed files over and over again.)

To combat these limitations, organizations need to invest more in scale-out NAS solutions that are designed to meet ever more diverse storage and information management needs.

PICKING THE RIGHT STORAGE SOLUTION FOR THE CONTENT-DRIVEN ORGANIZATION

While many people think rapid file-based data growth is limited to a few industries such as media or healthcare, the scale of content growth across all companies is significant. Major forces driving this growth include:

- ☒ Increased archiving of data for compliance or use in analysis of historical trends
- ☒ Greater use of virtualized servers and desktops
- ☒ Consolidation of home directories
- ☒ Disk-based backup and tiering of applications, databases, and log data

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 midsize organizations, this transition will involve greater use of

cloud-based storage services. Most large enterprises (not just organizations in content-driven industries or providers of cloud-based storage services) will need to deploy new scale-out NAS solutions optimized for content ingest, storage, and access.

What Is Scale-Out NAS?

Scale-out NAS refers to file-based solutions that provide a single name space or object storage pool across physical storage and/or process modules. Some scale-out NAS systems allow for incremental scaling of performance and capacity independent of each other, which can deliver considerable flexibility in asset management.

When evaluating scale-out NAS solutions, you need to consider the following fundamental, common elements:

- ☒ **Multidimensional scalability.** Solutions that offer elastic scalability will arm customers with the ability to better manage rapidly expanding and varying loads. Many pools of unstructured content start modestly but can quickly overwhelm the initial NAS environment. Scalability isn't just about expanding hardware capacity or providing more consolidated management of dispersed NAS systems, however. In performance- and content-intensive environments, scalability must also encompass throughput, file size, and file volume. The best solutions will allow you to scale performance and capacity simultaneously and independently to ensure the optimal configuration for your specific needs.
- ☒ **Storage efficiency.** Solutions in this area will enable easy adoption of a continually expanding range of storage optimization technologies. These capabilities include data reduction technologies (e.g., thin provisioning) and data optimization technologies (e.g., automated data tiering). The best solutions will allow you to introduce or recalibrate these capabilities without major disruptions or system migrations.
- ☒ **Intelligent information management.** Solutions in this area will allow organizations to continuously observe data usage, 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.
- ☒ **Petabyte-level data protection.** As content pools grow, data protection processes and methods need to employ new approaches to reduce file rebuild times, eliminate the need for backup/migration downtime, and provide for rapid disaster recovery at replicated facilities.

The remainder of this white paper examines EMC's Isilon scale-out NAS solutions and assesses how well they address organizations' growing need for petabyte-scale storage systems that address rapid and increasingly complex content creation, access, and archiving needs.

EMC'S ISILON STORAGE SCALE-OUT NAS SOLUTIONS FOR LARGE CONTENT POOLS

EMC's Isilon Storage Division (Isilon) is a global leader in the development and deployment of scale-out NAS solutions for a wide range of content-driven organizations and large content depots as well as a growing range of enterprises across all industries that are struggling with datacenter transformation and the current data explosion.

As of mid-2012, EMC Isilon had over 3,000 customers worldwide deploying over 40,000 of its modular system elements. The Isilon division's stated mission is to remove the barriers that exist between businesses and their data by helping customers get to the information they need and manipulate and act on that information. It seeks to meet this goal by delivering a scale-out NAS platform with near limitless capacity, unprecedented levels of performance, and advanced data services that make the storage and delivery of content more efficient and simple.

Isilon's Scale-Out Platform: Delivering Scale in Multiple Dimensions

The foundation of Isilon's scale-out NAS solution (the Isilon cluster) is the OneFS operating system, currently in its sixth generation. It provides a unified software foundation for basic storage services (file system, volume manager, and RAID) that can be distributed across multiple storage hardware nodes and tuned for different requirements.

At present, Isilon delivers five different specialized hardware nodes:

- ☒ S-Series — IOPS-Intensive Applications (including SSD support)
- ☒ X-Series — High Concurrent and Sequential Throughput Workflows
- ☒ NL-Series — High Density/High Capacity for Archival Content
- ☒ Performance Accelerator — Independent Scaling of Throughput
- ☒ Backup Accelerator Node — High-Speed and Scalable Tape Backup and Restore Solution

The combination of OneFS and these tuned hardware nodes 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 more than 1 petabyte of storage capacity with a single full-time equivalent (FTE) administrator.
- ☒ **Highly tunable and scalable capacity and throughput levels.** Organizations can deploy over 15 petabytes of capacity, achieving in excess of 100GBps of throughput or supporting up to 1.64 million IOPS in a single file system.
- ☒ **Rapid system expansion and upgrades.** Administrators can nondisruptively add capacity and/or performance resources within 60 seconds and thus no longer need to buy/overprovision capacity months or years in advance of actual need.

- ☒ **Evergreen (persistent) storage.** Using the OneFS autobalance feature, IT teams can quickly and nondisruptively migrate data from older nodes (due for replacement) to new, higher-performing or more cost-effective nodes as they are introduced.
- ☒ **High reliability.** The combination of independent scalability at the node level along with the advanced data protection/replication capabilities of OneFS allows the Isilon system to become more resilient as it grows.

Isilon's Application Software Portfolio: Enabling Storage Efficiency and Information Management

Providing a cost-effective and future-proof scale-out storage platform isn't sufficient to meet the increasingly complex needs of content-intensive organizations. IT teams in these organizations also require a complementary set of application and management software solutions to address operational and administrative challenges.

Isilon, in conjunction with several partners, also provides a suite of advanced application and management solutions to address customers' storage efficiency and information management requirements. These solutions include the following:

- ☒ **SmartQuotas** provides the ability to set up, control, thin provision, and limit storage usage subpools in an Isilon scale-out storage cluster to meet storage challenges for different content types and use cases.
- ☒ **SmartPools** provides automated, policy-based file migration; workflow isolation; higher utilization; and independent scalability for different content pools in a single Isilon cluster.
- ☒ **SyncIQ** includes sophisticated, policy-based file replication management tools to quickly and efficiently move mission-critical data through a globally distributed network for load balancing or business continuity/disaster recovery.
- ☒ **SmartLock** features write once, read many (WORM) policy-based immutability and retention to provide protection against accidental, premature, or malicious alteration or deletion of data.
- ☒ **SmartConnect** simplifies the management and maximizes the performance of large numbers of clients that access content from Isilon clusters.
- ☒ **SnapShotIQ** enables faster and more efficient snapshot scheduling to provide more flexibility in meeting recovery point and recovery time goals.
- ☒ **InsightIQ** includes advanced analytics/reporting tools to optimize applications, correlate workflow and network events, and more accurately forecast future storage needs.
- ☒ **Aspera for Isilon** is a predictable, high-performance wide area file and content delivery solution designed to move large files long distances (jointly developed/provided by Aspera and Isilon).

Challenges/Opportunities for EMC's Isilon Division

Given continued rapid data growth and the increasing role of rich content in organizations' new application and services plans, the storage and information management challenges posed by Isilon's content-driven customers and prospects will only increase in the coming years. EMC 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, further enhancements to SmartPools to enable even greater levels of tunable/intelligent automated data tiering, and support for even denser/more power-efficient SSD and HDD solutions
- ☒ Establish closer technical and business ties with leading content creation, distribution, and analytics (Big Data) solutions suppliers that will make it easier for customers to fully exploit the information stored within Isilon clusters (An early response from the Isilon division is its enhancement enabling native Hadoop Distributed File System [HDFS] access to Isilon-based data sets.)
- ☒ Extend the reach of the Isilon solutions to better address the growing content needs of small and medium-sized businesses as well as large enterprises (This effort should include expanding the ecosystem of cloud-based storage service provider partners.)

CONCLUSION

The digital universe will continue to expand, and organizations will spend more time and money collecting, storing, and monetizing fast-growing data sets and rich pools of content for the foreseeable future. IDC expects companies such as EMC Isilon to develop and deliver increasing numbers of tools and practices to help the guardians of organizations' digital assets — the CIOs of the world — deal with the here and now of their own particular unstructured content depots and Big Data analytic investments.

In the end, however, a CIO's challenges will be not just technical — how to find, manage, and protect information — but also organizational. CIOs must take a leadership role in driving the adoption of new, information-taming technologies and practices, such as EMC Isilon's scale-out 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 successful CIOs and their IT teams are in driving the required enterprisewide adjustment to the new realities of the data- and 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 2012 IDC. Reproduction without written permission is completely forbidden.