

WHITE PAPER

Get Ready for Big Data:

How Scale-Out NAS Delivers the Scalability, Performance,
Resilience and Manageability that Big Data Environments Demand

FUSIONSTORM™

www.fusionstorm.com

EXECUTIVE SUMMARY

“Big data” has become a term of art in the IT industry, describing not only the massive amount of information generated today but its enormous potential value. Big data promises to revolutionize business decision-making, and the ability to draw intelligence from big data has become a top priority in many organizations.

The primary sources of big data growth are customers, sales, marketing and operations. To a great extent, this data is not transactional but rather unstructured information captured in files, emails, social media and the like. In order to reach big data nirvana, organizations must break down the silos separating these diverse information sources and combine them into a common pool. Once the data is aggregated, informatics and data science can be used to discover patterns in the data that humans cannot perceive.

Taking that first step is far from easy, however. Big data places tremendous pressure on the storage infrastructure as organizations grapple with petabytes or exabytes of information. Big data is ill-suited for storage area networks (SANs), which were designed for high-performance OLTP applications rather than unstructured data and file sharing. However, traditional file-based network attached storage (NAS) lacks the scalability, reliability and performance to handle big data demands.

These problems are particularly acute in data-intensive environments such as financial services, where the speed and depth of analyses impacts line-of-business operations and competitive position. Financial services firms require high-performance storage that can support these mission-critical applications across growing volumes of unstructured data.

Scale-out NAS can help organizations cope with the onslaught of big data. Scale-out NAS is a proven, enterprise-class technology that scales horizontally using commodity storage in a clustered configuration. With scale-out NAS, organizations can build an aggregate storage pool that expands rapidly and cheaply yet delivers predictable performance, high availability and optimum utilization.

This paper discusses the value of scale-out NAS for big data environments, with the financial services sector as an example use case. It focuses on EMC Isilon, a software-managed scale-out NAS solution that overcomes the limitations of legacy file-based storage systems.

THE BIG DATA CHALLENGE

Today's storage environment is under tremendous pressure from the inside due to rampant data growth. According to IDC's 2012 Digital Universe Study, 2.8 zettabytes (ZB) of data was created and replicated in 2012, most of it unstructured. IDC projects that the digital universe will reach a mammoth 40ZB by 2020, an amount that exceeds previous forecasts by 14 percent. Spending on the hardware, software, services and staff that form the "infrastructure" of the digital universe will increase by 40 percent between 2012 and 2020, with targeted areas such as storage management growing considerably faster.

One way to deal with this data storage growth is to scale vertically with higher-capacity storage devices. In the short run, this consolidates the storage environment and simplifies management by moving data onto fewer devices. Ultimately, however, data volumes are going to exceed the maximum capacity of those devices, and additional systems will have to be deployed. Big data environments often wind up with tens or even hundreds of storage silos, most of which are underutilized. The resulting storage sprawl increases capital outlays and energy costs, and causes severe management headaches.

Traditional scale-up storage systems also lack the horsepower to complete big data operations efficiently. Many organizations add more spindles to the environment and reduce the amount of data stored on each disk in order to avoid performance bottlenecks and provide the throughput needed for large numbers of I/O requests. This practice also leads to a bloated yet underutilized storage infrastructure.

Although the cost per gigabyte for data storage has plummeted, the money spent in creating, storing and managing data has increased 50 percent to \$4 trillion since 2005, and is expected to reach \$5.2 trillion by 2015. Organizations seeking to rein in data storage costs need fluid scalability that maximizes utilization. They need to unify storage subsystems in order to successfully manage ever-increasing data storage demands. At the same time, they need to optimize the information management environment by seamlessly moving data to the appropriate platform for both performance and cost savings. None of that is practical in the traditional scale-up NAS environment.

UNDERSTANDING SCALE-OUT NAS

Scale-out NAS is ideally suited to handle growing volumes of unstructured data. A scale-out NAS architecture aggregates commodity storage devices to create a common storage pool with a virtualization/abstraction layer that makes the devices behave like a single system. The system is easily scaled by adding more devices to the cluster rather than increasing the capacity of a particular device.

Scale-out NAS delivers a number of key benefits:

- **Reduced storage acquisition costs.** Organizations can leverage high-capacity, low-cost disk storage instead of more expensive high-performance storage capacity.

- **Faster, non-disruptive scalability.** IT managers no longer need to overprovision storage resources to ensure that essential services are not disrupted due to rapidly changing demands. Organizations can increase storage utilization rates and add capacity in days, not months.
- **Better performance.** In a scale-out NAS environment, performance scales in the same linear fashion as capacity. Clustering provides the throughput and bandwidth needed for high-performance environments, and enables applications to take advantage of that performance seamlessly.
- **Simplified management.** Scale-out NAS features a single file system and single point of management, enabling one administrator to manage tens of petabytes as easily as 100TB.
- **High availability.** Although built on commodity storage devices, the clustered architecture is designed to survive multiple hardware failures.

Scale-out NAS also supports automated, policy-based storage tiering. This enables organizations to use the right storage resources for specific workloads and data management requirements. Organizations can also optimize the storage environment by moving data among price/performance tiers across its lifecycle.

EMC ISILON SCALE-OUT NAS

The EMC Isilon scale-out NAS solution is based upon the OneFS storage operating system, which combines the three layers of traditional storage architectures — file system, volume manager and data protection — into one unified software layer. Industry-standard storage devices function as nodes connected via an Infiniband high-speed interconnect, with OneFS creating a single intelligent file system that spans all nodes within an Isilon cluster. When devices are added to the cluster, OneFS automatically joins the new nodes and redistributes data evenly across the cluster.

EMC Isilon supports industry-standard file-based protocols, including CIFS, NFS, HTTP and FTP, and the block-based iSCSI protocol, in a single, shared pool of storage. EMC Isilon also natively integrates with the Hadoop Distributed File System (HDFS), removing the complexity associated with integrating disparate open source components and hardware in a Hadoop environment.

With EMC Isilon data management tools, organizations can easily configure, manage, clone, thin provision, snapshot, tier, replicate and secure storage systems. EMC Isilon SmartPools creates a single point of management for all storage tiers and automatically aligns application needs with performance, capacity and cost. Organizations can consolidate a wide range of applications onto a single storage resource, eliminating the need for manual data migration between multiple systems. SmartPools accelerates data access for mission-critical business processes while reducing storage costs and management.

EMC Isilon InsightIQ provides analytics that help administrators identify and eliminate performance bottlenecks. It delivers actionable insight into data usage trends, enabling immediate and effective response to changes in workflow requirements and user demand, improving resource utilization and business agility.

With EMC Isilon, organizations also gain increased resilience through faster drive-rebuild times and superior data protection. Should a drive failure occur, an EMC Isilon scale-out cluster can rebuild active and archive data in less than one day, compared to multiple days or even weeks with traditional storage systems. The FlexProtect file striping technology built into OneFS provides N+4 data protection to help reduce the risk of data loss and improve availability, even as the cluster grows.

IDC POSITIONS EMC AS A 'LEADER' IN SCALE-OUT NAS

EMC has been positioned as a "Leader" by IDC in its "IDC MarketScape: Worldwide Scale-Out File-Based Storage 2012 Vendor Analysis" report. EMC earned this ranking based on the functionality and success of its EMC Isilon portfolio.

IDC reports that "EMC's portfolio encompasses a variety of storage products for the entry, midrange and enterprise levels that serve numerous use cases. EMC's backbone for the scale-out file-based storage segment is Isilon. EMC acquired Isilon in late 2010 and positions Isilon's unique clustered architecture for customers that face rapid growth of data, high bandwidth needs and high availability demands." The IDC report further notes: "EMC has consistently dominated the storage space for several years. It has been able to do so by identifying upcoming trends and bringing to market timely and appropriate storage solutions. EMC has also benefited from strategic alliances and acquisitions to enhance its portfolio, giving it an edge over the competition."

USE CASE: FINANCIAL SERVICES

EMC Isilon scale-out NAS is a good fit for medical imaging, gene sequencing, seismic exploration, media and entertainment, satellite images, product development/chip design and high-performance computing. It is also well-positioned for large-scale home directories, archives and analytics. Financial services applications provide a useful illustration of EMC Isilon at work.

Financial services organizations rely upon advanced transaction modeling to maximize returns and/or reduce risk. These organizations require a high-performance data center infrastructure that can perform quantitative financial analyses based upon sophisticated algorithms. In the past, much of the data manipulated in these models was stored in the enterprise SAN. Today, transactional and trading information is increasingly file-based, creating a need for high-performance NAS within the financial services sector.

EMC Isilon scale-out NAS provides the capacity and performance needed in the financial services environment. EMC Isilon has been proven to enable 50 percent better turnaround time at 30 percent lower cost than legacy NAS systems, and can expand data access independent of storage for deeper analysis. EMC Isilon also allows snapshots at the individual file level, which is important in iterative flows where new data is written to the same file based upon the analysis. EMC Isilon SmartLock protects critical data from unauthorized deletion or alteration, helping financial services organizations to meet stringent SEC 17a-4 requirements.

An EMC Isilon system can support Isilon-certified components from a variety of manufacturers to scale from terabytes to petabytes. The EMC Isilon solution also includes three platform nodes, enabling financial services firms to choose the right storage for specific tasks:

- **S-Series** — high I/O performance and a large, scalable, globally coherent cache for complex quantitative financial analysis tied to high-performance compute clusters
- **X-Series** — high concurrent access and throughput for traditional quantitative analysis, portfolio analysis, compliance checking, risk analysis and other line-of-business workflows
- **NL-Series** — cost-effective nearline archival for legal and regulatory compliance requirements

EMC Isilon can scale on demand in terms of capacity, performance and throughput, without downtime or configuration changes. This enables financial services firms to manipulate resource-intensive models without interruptions for adding capacity or retargeting applications to additional storage systems.

FUSIONSTORM: PROVEN STORAGE EXPERTISE

FusionStorm is an EMC Signature Velocity Partner, with proven expertise in the deployment of EMC Isilon scale-out NAS. FusionStorm engineers can help organizations migrate to EMC Isilon from traditional storage solutions without workflow disruption, while reducing complexity and management overhead.

CONCLUSION

Big data promises great opportunities but also presents tremendous storage challenges. Organizations need more efficient and cost-effective ways of building and managing large, rapidly growing information repositories in order to tap the business intelligence potential of big data. The increasing volume of file-based information has outpaced transactional data, driving demand for highly scalable NAS solutions.

Scale-out NAS is the ideal platform for big data environments. It provides scalability, performance and resilience in an efficient, easy-to-manage environment that can be built on low-cost commodity storage. Automated, policy-based storage tiering enables organizations to optimize the storage environment based upon workload demands and information lifecycle requirements.

Scale-out NAS is particularly beneficial in certain industry sectors, such as financial services, that rely upon data-intensive applications. Financial services firms need storage solutions that deliver the performance to support complex financial analyses, the throughput for line-of-business workflows and cost-effective capacity to meet legal and regulatory compliance requirements for data retention. In these environments, scale-out NAS is displacing expensive enterprise scale-up storage equipment with more agile, cost-effective solutions.

With its seventh-generation operating system and suite of data management tools, EMC Isilon is a leader in the shift toward scale-out NAS. EMC Isilon scale-out NAS is changing the face of file-based storage in big data environments through next-generation performance, efficiency and availability as well as cost-effective scalability.

About FusionStorm

FusionStorm delivers best-of-breed technology solutions that give our customers a competitive edge. Our deep technical expertise and business acumen enable us to design and implement solutions aligned with each customer's unique business requirements. We maintain the highest certifications from industry leaders and offer an array of services designed to maximize the efficiency of our customers' operations.

Headquartered in San Francisco, FusionStorm has offices and data centers across the globe, and a culture that emphasizes long-term relationships. By partnering with FusionStorm, customers gain a team of seasoned professionals with a track record of success and an unwavering commitment to customer satisfaction.



800.228.TECH | info@fusionstorm.com | www.fusionstorm.com

Headquarters

FusionStorm Inc.
2 Bryant Street, Suite 150
San Francisco, CA 94105

PHONE: 800 228-TECH
FAX: 415 623-2630