# EMC ISILON SCALE-OUT NAS: AN ARCHITECTURE FOR RESILIENCY, HIGH AVAILABILITY, AND DATA PROTECTION

## ABSTRACT

This paper demonstrates that the EMC Isilon file system can remain online while a cluster sustains multiple failures of underlying components. The EMC® Isilon® OneFS® architecture, data protection scheme, and high-availability features deliver resiliency, reliability, and data availability. This paper covers EMC Isilon OneFS Version 7.

July 2014

REDEFINE

EMC WHITE PAPER

EMC²

To learn more about how EMC products, services, and solutions can help solve your business and IT challenges, contact your local representative or authorized reseller, visit www.emc.com, or explore and compare products in the EMC Store

Part Number H13330

## TABLE OF CONTENTS

# INTRODUCTION

Enterprise IT administrators, catering to an increasingly global workforce, are under pressure to provide around-the-clock services that operate without disruption. The pressure for continuous services now extends beyond mission-critical applications. File services ranging from email and home directories to web content repositories are expected to be online, all the time.

Minor disruptions in operations can have a major impact on business, leading enterprise service-level agreements to stipulate levels of uptime and availability. IT must ensure that unexpected disruptions are avoided and rapidly resolved if they occur.

## Diverse Requirements for 24x7 Availability

Availability becomes more important in the face of diverse, globalized requirements. In a globalized economy, business operations never stop—corporations provide services 24x7x365 or use a "follow-the-sun" model of operations and centralized IT services. In some cases, IT operations that directly affect lives, like health care and government, require non-stop operations. The following are examples of 24x7 operations that are proliferating amid diverse requirements:

- Financial services such as retail online banking and the data for automated teller machines
- Medical data services such as consolidating medical imaging into a vendor neutral archive
- User home directories and group file shares for multinational organizations seeking to collaborate and exchange documents
- Digital design firms ranging from media and entertainment to electronic design automation (EDA) for which downtime results in lost productivity and lost revenue

## Continuous Applications

Software applications impose additional pressure on IT systems to deliver highly available infrastructure. The growth in the sophistication and complexity of applications means that the applications either take a longer time to complete or in some cases never cease.

Design firms, for example, run jobs that can take days or weeks to finish. In many cases, these jobs cannot be restarted from where they left off. Any disruption in data access that causes an application to lose its connection requires the entire job to be restarted from the beginning. Depending on when the job was interrupted, how many jobs were running at the time, and the downstream dependencies of the job, the downtime can reduce revenue, undermine productivity, and increase costs. For firms operating in markets with sensitive time-to-market requirements, such as electronic design automation, downtime can affect profit margins.

## Workload Consolidation

Other pressures, too, are coming to bear on IT infrastructure. As enterprise IT infrastructure becomes more powerful and more robust, there is a tendency to run more workloads on a single system. System architects are pointing dozens of workloads and applications at a single storage system because consuming a single pool of storage capacity is more economically efficient than using independent storage silos.

Pooling capacity, however, increases the importance of the storage system's resilience. As the storage system expands and as more workloads are consolidated into it, it is critical that single failures do not turn into cascading failures that imperil the workloads.

Any disruption in data access impairs the delivery of critical IT and data services. As such, the storage systems that provide applications, clients, and virtual machines with data and storage capacity are the foundation of data availability.

# STORAGE AVAILABILITY

SLAs, 24x7 operations, applications, virtual machines, and client computers all to some extent depend on storage systems. Given the pressures of a globalized marketplace and diverse requirements that IT operations impose, a storage system must remain highly available at all times to serve data. Data availability is paramount.

To remain online, to protect data, and to ensure its availability, a storage system must meet a range of high-level requirements for both its hardware and software, including the following examples:

- Highly available architecture with no single point of failure
- Capacity to handle hardware and disk failures
- Data protection
- Backup and restore
- Disaster recover

- Redundant network interfaces
- Continuous monitoring

For the purposes of this paper, availability is defined as the file system being online so that users and applications can read and write data. This paper demonstrates that an EMC Isilon file system can remain online while a cluster sustains multiple failures of underlying components.

### Traditional NAS vs. Scale-Out NAS

Scale-Out storage differs from traditional storage systems such as storage-attached networks, direct-attached storage, and scale-up network-attached storage. In a scale-out architecture, the loss of a drive or a node is handled differently from dual-controller architectures, in which single chassis failures can take out an entire cluster if the HA pair's capacity is insufficient.

Traditional architectures for NAS appliances rely on a failover relationship between two controllers that form a high-availability (HA) pair in which the two controllers provide redundancy. When failure occurs, the clients, connections, and storage operations that were handled by a single storage controller are moved to a neighboring storage controller.

However, an HA pair can create utilization problems. Loading a traditional NAS system beyond 50 percent can result in lengthy downtime when a hardware component fails. In the event of an unplanned outage that stems from a failure, the second controller in the HA pair must contain enough capacity to handle the entire load from the failed controller. The second controller must also include the performance characteristics and operational functionality to handle the entire load. Therefore loading either of the system's controllers beyond 50 percent places the HA pair at risk should a failure occur. Malfunction of the single controller is not considered to be a failure that results in the inability to access data unless the resulting performance decline exceeds the specification limits for the entire system.

### Availability with Isilon Scale-Out NAS

Scale-out NAS systems are different. The architecture of an EMC Isilon scale-out NAS system contains no single master for the data and no concept of an HA pair. Instead, Isilon scale-out NAS is a fully distributed system that consists of nodes of modular hardware arranged in a cluster. The distributed Isilon OneFS operating system combines the memory, I/O, CPUs, and disks of the nodes into a cohesive storage unit to present a global namespace as a single file system.

The nodes work together as peers in a shared-nothing hardware architecture with no single point of failure. Every node adds capacity, performance, and resiliency to the cluster. As nodes are added, the file system expands dynamically and redistributes data, eliminating the work of partitioning disks and creating volumes. The result is a highly resilient and scalable storage architecture.

The distributed OneFS operating systems handles a failure by distributing the load of a failed node to the remaining nodes. The system keeps just enough redundant information to reconstruct the data on a node or disk that fails, and the amount of overhead to protect against failure decreases as nodes are added to the cluster.

As a result, storage administrators can extract much higher utilization rates from the storage system without risking excessive downtime. Compared with traditional scale-up NAS systems, a scale-out architecture provides a more resilient foundation for data protection and data availability.

In its 2013 report titled "Critical Capabilities for Scale-Out File System Storage," Gartner rated EMC Isilon highest among storage vendors for resiliency—the platform's capabilities for provisioning a high level of system availability and uptime.[1]

## ISILON ARCHITECTURE FOR AVAILABILITY

The design of Isilon's clustered architecture supports the following availability objectives:

- No single point of failure
- Unparalleled levels of data protection
- Tolerance for multi-failure scenarios
- Fully distributed single file system
- Pro-active failure detection and preemptive drive rebuilds
- Fast, scaleble drive rebuilds
- Flexible, efficient data protection
- Fully journaled file system

## Hardware

An Isilon cluster is built on a highly redundant architecture governed by the hardware premise of shared nothing. The fundamental building blocks are platform nodes. As a rack-mountable appliance, a node includes the following components in a 2U or 4U rack-mountable chassis with an LCD front panel: memory, CPUs, RAM, NVRAM, network interfaces, InfiniBand adapters, disk controllers, and storage media. The redundant InfiniBand adapters provide the distributed system bus that connects all the nodes. Each node houses a battery-backed file system journal. NVRAM is grouped to protect write operations from power failures.

## Network

Client computers can access any node in the cluster through dual 1 GigE or 10 GigE network connections. Client connections are, by default, distributed across the cluster with round-robin load balancing. On the network side, Isilon's logical network interface (LNI) framework provides a robust, dynamic abstraction for easily combining and managing differing interfaces for network resilience. Multiple network interfaces can be trunked together with LACP and LAGG to aggregate bandwidth.

An EMC® Isilon® SmartConnect™ automatic client-connection, load balancing, and failover software license adds additional network resilience with IP address pools that support multiple DNS zones in a subnet as well as IP failover, also known as NFS failover.

## File System

The cluster's highly extensible file system provides mirrored volumes for the root and /var volumes with the Isilon Mirrored Device Driver (IMDD), stored on flash drives. For further resilience, OneFS also automatically saves last-known good boot partitions.

## Data Protection

OneFS is designed to withstand multiple simultaneous component failures while preserving unfettered access to the file system. OneFS implements data protection as part of the file system; data protection does not depend on hardware or RAID controllers.

For efficiency and reliability, OneFS stripes data to guard it with parity blocks at the file level instead of parity disks. A larger cluster can lose up to four 36-drive nodes without loss of data and without disrupting client connections.

At the same time, OneFS protects data with Reed-Solomon Forward Error Correction, or FEC—a highly efficient method of reliably protecting data. FEC encodes a file's data in a distributed set of symbols, adding space-efficient redundancy. With only a part of the symbol set, OneFS can recover the original file data. In a cluster with five or more nodes, forward error correction delivers 80 percent or more storage efficiency. On larger clusters, FEC provides as much as four levels of redundancy.

OneFS supports N+M FEC levels of N+1, N+2, N+3, and N+4. In the N+M data model, N represents the number of data stripes, and M represents the number of FEC stripes. For example, with N+2 the cluster can lose two drives on different nodes or lose two nodes. OneFS also supports N+M:B. In the N+M:B notation, M is the number of disk failures, and B is the number of node failures. With N+2:1 protection, for example, the cluster can lose two drives or one node without losing data. As an alternative to FEC codes, OneFS can also protect data with up to to eight mirrors.

## OneFS Availability Software and Features

In addition to data protection with striping and forward error correction, OneFS includes the following software modules and features to help protect the integrity, availability, and confidentiality of data. Some of the software modules require a separate license.

**SMARTPOOLS**. SmartPools establishes multiple file pools governed by file-pool policies. For data availability, SmartPools policies move data to disk pools for tiering and for segregating workflows. As such, SmartPools aligns the availability requirements of the data with the right class of storage. An Isilon cluster can provide multiple pools that support a range of availability SLAs within a single file system—a resource pool model that aligns with the current IT trend of private and hybrid cloud initiatives.

**CLONES**. Isilon provides full read and write copies of files with clones that share blocks with other files to save space. OneFS also provides virtual machine linked cloning through VMware API integration.

**AUTOMATED CLUSTER REPLICATION AND FAILOVER**. SyncIQ replicates data on another Isilon cluster and automates failover and failback operations between clusters. If a cluster becomes unusable, SyncIQ fails over to another Isilon cluster.

**SNAPSHOTS**. SnapshotIQ protects data with a snapshot—a logical copy of data stored on a cluster. A snapshot can be restored to its top-level directory. SnapshotIQ provides various features and options to meet recovery point objectives.

**FILE SYSTEM JOURNAL.** A journal, which records file-system changes in a battery-backed NVRAM card, recovers the file system after failures, such as a power loss. When a node restarts, the journal replays file transactions to restore the file system.

**VIRTUAL HOT SPARE**. When a drive fails, OneFS uses space reserved in a subpool instead of a hot spare drive. The reserved space is known as a virtual hot spare. In contrast to a spare drive, a virtual hot spare automatically resolves drive failures and continues writing data. If a drive fails, OneFS migrates data to the virtual hot spare to reprotect it. As an option, storage administrators can reserve as many as four disk drives as a virtual hot spare.

**ANTIVIRUS**. OneFS can send files to servers running the Internet Content Adaptation Protocol (ICAP) to scan for viruses and other threats.

**NDMP BACKUP AND RESTORE**. OneFS can back up data to tape and other devices through the Network Data Management Protocol.

**SMARTLOCK**. The SmartLock tool prevents users from modifying and deleting files. With a SmartLock license, security managers can commit files to a write-once, read-many state: The file can never be modified and cannot be deleted until after a set retention period. SmartLock protects critical data from malicious, accidental, or premature alteration or deletion to help ensure compliance with SEC 17a-4 regulations.

**INTEGRITYSCAN**. An IntegrityScan job examines the file system for inconsistencies by systematically reading every block and verifying its associated checksum. Unlike traditional 'fsck' style file system integrity checking tools, IntegrityScan runs while the cluster operates, eliminating the need for downtime. If IntegrityScan detects a checksum mismatch, OneFS generates an alert and automatically attempts to repair the block.

**DYNAMIC SECTOR REPAIR**. OneFS includes a Dynamic Sector Repair (DSR) feature that forces bad disk sectors to be rewritten elsewhere. When OneFS fails to read a block, OneFS invokes DSR to reconstruct the missing data and write it to either a different location on the drive or to another drive on the node, ensuring that subsequent reads of the block do not fail. DSR is fully automated and completely transparent to system administrators. Disk sector errors and CRC mismatches use almost the same mechanism as the drive rebuild process.

**MEDIASCAN**. A MediaScan job checks disk sectors and uses DSR to force disk drives to fix sector ECC errors. A low-impact background process, MediaScan is fully distributed to exploit the benefits of Isilon's parallel architecture.

**PROACTIVE DEVICE FAILURE**. OneFS proactively removes, or smartfails, a drive that reaches a threshold of ECC errors and reconstructs the data from that drive elsewhere. Both SmartFail and the subsequent repair process require no administrator intervention and no cluster downtime.

**ISILON DATA INTEGRITY.** Isilon Data Integrity (IDI) protects file system structures against corruption with 32-bit CRC checksums. All Isilon blocks use checksum verification. Metadata checksums are housed in the metadata blocks themselves, whereas file data checksums are stored as metadata, thereby providing referential integrity. All checksums are recomputed by the initiator, the node servicing a particular read, on every request. If a recomputed checksum does not match the stored checksum, OneFS generates a system alert, logs the event, and returns the corresponding parity block to the client and attempts to repair the data block.

**PROTOCOL CHECKSUMS**. In addition to blocks and metadata, OneFS also provides checksum verification for Remote Block Management (RBM) protocol data. RBM is a unicast, RPC-based protocol developed by Isilon for use over the internal InfiniBand network. Checksums on the RBM protocol are in addition to the InfiniBand hardware checksums provided at the network layer. The RBM checksums detect and isolate machines that contain faulty hardware components.

**FAULT ISOLATION**. Because OneFS protects its data at the file level, any data inconsistencies are isolated on the unavailable device—the rest of the file system remains intact and available. If, for example, a ten-node cluster that is protected at N+2 sustains three simultaneous drive failures, one in each of three nodes, the data striped across the other two hundred and thirty-seven drives would remain unaffected. In contrast, with a traditional RAID6 system, losing more than two drives in a RAID-set renders the system unusable and requires a full restore from backups.

**ACCELERATED DRIVE REBUILDS**. The time it takes a storage system to rebuild data from a failed disk drive is crucial to the data reliability of the system. With the advent of four terabyte drives and the creation of increasingly larger single volumes and file systems, typical recovery times for classic RAID implementations can extend to a week or more. During this period, storage systems are vulnerable to additional drive failures and the data loss and downtime that can result.

Since OneFS is built upon a highly distributed architecture, OneFS can exploit the CPUs, memory, and spindles from many nodes to efficiently reconstruct data from failed drives in a parallel process. Because Isilon is not bound by the speed of any one drive, OneFS can recover from drive failures extremely quickly— an efficiency that increases with cluster size. As such, a failed drive in an Isilon cluster is rebuilt an order of magnitude faster than RAID-based storage devices without the need for hot-spare drives.

Isilon availability software and options contribute to ensuring that an enterprise can meet its recovery time objective (RTO), which is the allotted amount of time within a service level agreement to recover and restore data. For complete information about the data availability features of OneFS, see the white paper titled High Availability and Data Protection with EMC Isilon Scale-Out NAS.

### Monitoring

OneFS monitors every node in a cluster. System administrators can view status with the OneFS web administration interface or the command-line interface. With SNMP versions 1, 2c, and 3, administrators can remotely monitor hardware components, CPU usage, switches, and network interfaces. OneFS also includes a RESTful application programming interface to automate monitoring and to retrieve statistics. In addition, Isilon supports monitoring and auditing through Varonis solutions.

**INSIGHTIQ**. The InsightIQ virtual appliance monitors and analyzes the performance of an Isilon cluster to help optimize storage resources and forecast capacity requirements.

**SUPPORTIQ.** OneFS logs contain data that Isilon Technical Support personnel can securely upload with an administrator's permission and then analyze to troubleshoot cluster problems.

**EMC SECURE REMOTE SUPPORT (ESRS)**. The ESRS gateway is a secure, IP-based customer service support system that includes 24x7 remote monitoring and secure authentication with AES 256-bit encryption and RSA digital certificates. It can monitor every node, send alerts, and provide support personnel with remote access for troubleshooting.

## PRESERVING UPTIME AND PROTECTING DATA

While it would be desirable to expect that enterprise hardware never fails, the reality is that components fail. When they do, it is critical that the failures do not affect the users and applications that rely on them. The uptime of the system must be preserved in the face of hardware failures.

This paper defines system availability as the OneFS file system being online with read and write quorum—meaning that users, clients, and applications can read from and write to the cluster.

Because OneFS distributes data across drives and across nodes, the likelihood that a drive failure results in data loss is, in general, extremely low.

As a guideline, the probability associated with drive failures means that they are, to a certain extent, predictable. As a result, a certain number of spare drives—the number depends on the size of the cluster and the total number of disks—can be kept on hand to replace a failed drive. In addition, OneFS includes SmartFail, which automatically quarantines a drive with a problem and reprotects the data. SmartFail can also decommission old drives so that they can be replaced before they become problematic.

### Analyzing System Availability

System availability is the percentage of uptime during a year. The "nines" of uptime describes that the system is down no more than a certain period each year to maintain an uptime level that is expressed as a percentage with as many as three nines after the decimal point, as the following table illustrates. To reach five nines of uptime—that is, 99.999 percent—the system can be down no more than 5.26 minutes per year.

| UNIT | AVAILABILITY PERCENT | DOWNTIME DURING A YEAR |
|---|---|---|
| One nine | 90 | 36.5 days |
| Two nines | 99 | 3.65 days |
| Three nines | 99.9 | 8.76 hours |
| Four nines | 99.99 | 52.56 minutes |
| Five nines | 99.999 | 5.26 minutes |

There is a difference between uptime and availability: A system can be up and running but unavailable for many reasons, such as a network outage or network routing problems. Uptime means that the system is online for reading and writing data.

In calculating uptime, EMC Isilon models the cluster as a repairable k-out-of-n system because it is a distributed system with independent nodes that relate to one another as peers and store redundancy information. Such a system functions when k components are functioning. With an Isilon cluster, n refers to the number of nodes and k refers to n − p where p is the protection level. Although evaluation of a non-repairable k-out-of-n system is straightforward, the inclusion of a repair process complicates formal attempts to evaluate availability.[2]

As with other distributed systems, a cluster must have a quorum to work properly. A quorum prevents data conflicts—for example, conflicting versions of the same file—in case two groups of nodes become unsynchronized. To prevent such conflicts, an unsynchronized node is separated, or split, from the other nodes in the cluster until it can resynchronize with the nodes and rejoin the cluster.

For a quorum, more than half the nodes must be available over the cluster's internal InfiniBand network. A seven-node cluster, for example, requires a four-node quorum. A 10-node cluster requires a six-node quorum. If a node is unreachable over the internal network, OneFS separates the node from the cluster.

The degraded states of nodes—such as offline, smartfail, read-only, and so on—are ignored when analyzing the availability of the cluster because, as long as there are enough available nodes to maintain quorum, a degraded node does not affect the uptime of the cluster. When OneFS can reconnect with a degraded node, OneFS merges it back into the cluster. Unlike a RAID system, an Isilon node can rejoin the cluster without being rebuilt and reconfigured.

For the cluster to be considered available, a certain number of nodes (k) must be operating out of a total number of nodes (n). Thus, returning to the earlier discussion of the N+M nomenclature, k=M and n=N. The k, that is, represents the number of nodes out of the total that must be available for the cluster to have a quorum. If the total number of nodes (n) in a cluster, for example, equals 5, then k must equal 3.

As a distributed system with independent nodes, each node's availability is identical to that of the other nodes—while any node's failure is independent from those of other nodes. By examining information from retired nodes to identify node failures, the EMC Isilon reliability model estimates its average failure rate (AFR) for nodes at 0.1 percent.[3] (The average failure rate does not take into account software faults, user error, environmental issues, or planned maintenance. A major upgrade is considered to be planned maintenance, and a rolling upgrade assumes that client connections can be disconnected or failed over to another node while the node is upgraded.)

By using the k+n model, the theoretical system availability of an example five-node Isilon cluster, when subjected to the node average failure rate of 0.1 percent, can be as high as 0.999999—which exceeds five nines of availability. The number, which is a guideline, not a guarantee, assumes that system administrators have maximized data protection and implemented Isilon’s high-availability features to their fullest extent.

Isilon's resiliency is recognized in the industry. Gartner, in its 2013 report on critical capabilities for storage, said, "Among the products evaluated, Isilon stands out in capacity, performance, manageability and resiliency."[1]

In the Gartner report, resiliency refers to "the options and capabilities offered in the platform for provisioning a high level of system availability and uptime. Options offered can include high tolerance for simultaneous disk and/or node failures, fault isolation techniques, built-in protection against data corruption and other techniques (such as snapshots and replication) to meet customers' recovery point objectives (RPOs) and recovery time objectives (RTOs)."

## Mean Time to Data Loss

With Isilon scale-out NAS systems, there is a tradeoff between data availability and storage efficiency. Higher protection levels typically consume more space than lower levels because storing FEC codes requires some space, however small. The overhead for data protection depends on the protection level, the file size, the number of nodes in the cluster, and other factors. Since OneFS stripes both data and FEC codes across nodes, the overhead declines as administrators add nodes to the cluster. Even with small clusters, however, the overhead is much less than with traditional storage systems.

With OneFS 7.0 or later, the default data protection for files is set to where the reliability of the cluster renders the five-nines paradigm obsolete. A different paradigm considers how long it would take for enough multiple failures to occur at the same time to cause data loss—in other words, the mean time to data loss (MTTDL).[4] Reliably protecting data, however, can ensure that even multiple failures at unusual rates nearly eliminate the likelihood of data loss. Isilon recommends a minimum protection level that is based on the size of the cluster and the types of nodes it contains to establish an estimated minimum mean time to data loss (MTTDL) of several thousand years.

Because so many variables interact to determine the optimal protection level for an environment, a best practice is to consult an EMC Isilon representative about selecting a protection level. Isilon can analyze the cluster to identify its mean time to data loss and then suggest an optimal policy.

### Performance Under Failure

An EMC Isilon cluster can continue to perform when components fail. Because an Isilon cluster is a pure scale-out architecture coupled with a distributed operating system that gracefully handles component failures, the OneFS file system can continue to support input-output operations while drives or nodes are being repaired or replaced. With the Isilon scale-out architecture, several failure scenarios that undercut traditional scale-up storage systems are rendered irrelevant.

For instance, with a traditional scale-up architecture, overloading a controller in an HA pair can undermine performance when a failure occurs. The second controller in an HA pair must include the performance characteristics to handle the entire load in case the first controller fails. A problem often arises, however, because system administrators gravitate toward tapping unutilized resources—they tend to point more connections at the controllers than one of them alone could handle. When a controller goes down, the remaining controller may not be able to support the additional load. As a result, the storage system can slow down by 50 percent or more or fail entirely. The loss of the single controller may not be considered to be a failure that completely results in the inability to access data unless the resulting performance decline exceeds the specification limits for the entire system.

The Isilon scale-out architecture is immune to such a failure scenario. Isilon does not contain HA pairs of controllers and thus does not experience a performance bottleneck on node failure. Because every node in effect acts as a controller, the failure of a node results in the distribution of the node's workload to other nodes. If a node goes offline, the cluster's overall performance degrades only by a small percentage. If one node in a 10-node cluster goes offline, for example, the cluster's performance diminishes only by about 10 percent.

Traditional RAID architectures compromise on data protection when drives fail. In a typical RAID DP group of 14 disks, 12 of the disks contain data, and 2 disks of capacity are set aside for parity. If too many disks in the group fail, however, the entire aggregate can go offline and lead to data loss. With an Isilon scale-out NAS cluster, in contrast, the distributed operating system protects data by distributing it across nodes in a cluster. Even in the unlikely event that three drives in a node fail, the data remains available from other nodes and protected from loss.

Traditional RAID systems also compromise on drive rebuild times. The time it takes to rebuild drives increases with the size of the drives. With an Isilon scale-out cluster, on the other hand, drive rebuild times are inherently low because the OneFS operating system distributes the work of rebuilding the data by using all the disks and CPUs in a set of nodes.

Finally, OneFS storage pools can automatically align data protection and performance with the value of data to help insulate important data from failures. Storage pools, which can be implemented at any time without having to reconfigure client computers, can isolate different datasets to increase performance, availability, and data protection for important data and decrease storage costs for less important data.

## CONCLUSION

An EMC Isilon cluster delivers resiliency through a scale-out architecture that so gracefully handles multiple hardware failures that it renders the five-nines paradigm obsolete. The resiliency of the Isilon architecture recasts availability in terms of mean time to data loss. By following the best practices for protecting data, storage administrators can implement availability features and data protection policies to ensure a high level of data availability.

## REFERENCES

[1] "Critical Capabilities for Scale-Out File System Storage," Gartner, Inc., published Jan. 24, 2013, http://www.gartner.com/technology/reprints.do?id=1-1DYP0VR&ct=130206&st=sb.

[2] For a primer on reliability theory and a discussion of availability calculations, see W. Kuo and M.J. Zuo. Optimal Reliability Modeling: Principles and Applications. John Wiley & Sons, 2002. Available from: http://books.google.com/books?id=vdZ4Bm-LnHMC

[3]For an overview of the statistical hypothesis testing behind calculating AFR, see J.L. Romeau. Reliability Estimations for the Exponential Life. RAC START, 10(7).

[4] For an overview of clustered storage reliability and its relationship to MTTDL, see KK Rao, James Lee Hafner, and Richard A. Golding. Reliability for Networked Storage Nodes. IEEE Transactions on Dependable and Secure Computing, 8:404–418, 2011. http://doi.ieeecomputersociety.org/10.1109/TDSC.2010.21