Abstract

This white paper reviews the challenges organizations face as they deal with the growing need for “always-on” levels of service availability. It illustrates how the EMC Isilon OneFS architecture provides the tools needed to address these challenges.

December 2012
# Table of contents

**EMC Isilon OneFS and nondisruptive operation** .............................................. 4  
Introduction ........................................................................................................ 4  

**Nondisruptive operation** ............................................................................... 4  
Client connection failover and balancing ............................................................ 5  

**Nondisruptive hardware upgrades** ................................................................. 6  
Node addition ..................................................................................................... 6  
Hardware refresh ............................................................................................... 6  
Pull-out node maintenance ................................................................................ 6  

**Nondisruptive software upgrades** ................................................................. 6  
Minor version rolling upgrade .......................................................................... 6  
Major version simultaneous upgrade ............................................................... 6  
Replication failover/failback ............................................................................. 8  
Fast recovery from unplanned outages ............................................................. 8  

**Client connection behavior** ......................................................................... 8  
NFS hard mounts .............................................................................................. 9  
NFS soft mounts ............................................................................................... 9  
SMB ................................................................................................................ 9  
Other protocols ............................................................................................... 9  
Applications ................................................................................................... 9  

**Conclusion** .................................................................................................. 10  

**About EMC Isilon** ........................................................................................ 10  
Contact EMC Isilon ......................................................................................... 10
EMC Isilon OneFS and nondisruptive operation

Introduction

In today’s world, enterprises are not only dealing with an enormous increase in the quantity of data required to successfully run their businesses, but also a growing pressure toward “always-on” levels of service availability. Maintenance windows are shrinking and stringent availability requirements afford only a handful of minutes or seconds of downtime per year for planned events—hardware refreshes, software and firmware upgrades, and infrastructure maintenance—plus any unplanned disruptive occurrences.

Organizations across the globe are faced with the daunting challenge of scaling to manage both an ever-increasing number and variety of clients, plus these demanding enterprise mandates for failure resiliency—all this, typically with little-to-no increase in storage administration resources.

In a recent study, a fundamental “lack of best practice storage architecture redundancy”¹ was highlighted as a primary reason for unavailability of enterprise IT services. It’s worth noting here that there is a considerable difference between uptime (system responsiveness) and availability (service responsiveness). Just because a storage system is up does not guarantee that it’s serving data correctly. Additionally, an external network event can as easily render data unavailable as a failed storage processor.

Fortunately, EMC® Isilon® enables IT managers to simply and efficiently satisfy the demands of this always-on 24x7x365 world by helping to ensure the highest levels of performance and availability. Through a highly parallel, “shared-nothing” network-attached storage (NAS) architecture, EMC Isilon OneFS® is able to deliver industry-leading levels of data protection and availability. When paired with intelligent client connection load balancing and failover support, Isilon dramatically simplifies and optimizes scale-out availability and performance for the enterprise.

Nondisruptive operation

Nondisruptive operation (NDO) is the process of managing a storage infrastructure without interrupting I/O to connected clients and their associated applications. In addition to guarding against unplanned outages, nondisruptive operation also includes proactive maintenance events such as software upgrades and hardware refreshes. Two main approaches are employed to address this, depending on the characteristics of the underlying storage architecture.

Legacy “scale-up” storage systems typically have two storage controller heads in an active-active configuration which front a large number of Fibre Channel Arbitrated Loop (FC-AL) attached hard drives. In this arrangement, one controller head is upgraded while the other handles all the client connections and storage management tasks. The downside here is that 50 percent of the storage system’s bandwidth and resources are rendered unavailable to service client I/O during this time. Additionally,

¹”Blueprints for High Availability,” Evan Marcus and Hal Stern.
system redundancy no longer exists, heavily impacting resilience to failures during the upgrade.

Other scale-up storage solutions employ an active-passive storage controller architecture, again with a large number of fibre channel attached drives. In this scenario, although there’s no reduction in system resources during an upgrade, redundancy is jeopardized. Additionally, this represents a poor return on investment (ROI), as 50 percent of the storage controller hardware sits idle during normal operating conditions.

By contrast, with Isilon nondisruptive upgrades, only a single node—a small piece of the overall cluster architecture—is upgraded at any time. Therefore, the bulk of the cluster’s bandwidth and redundancy is available throughout the entire upgrade process.

**Client connection failover and balancing**

Within Isilon OneFS, the EMC Isilon SmartConnect™ front-end connection load-balancer contributes heavily to data availability by supporting dynamic failover and failback of attached clients. This helps ensure that, during a rolling upgrade, network file system (NFS) client sessions are automatically handed off to another node in the cluster to finish its operation with minimal user or application interruption. Microsoft® Windows® clients also benefit by easily being able to remount a server message block (SMB) share using any other available node in the cluster.

During failover, clients are evenly redistributed across all remaining nodes in the cluster, ensuring minimal performance impact. If a node is brought down for any reason, including a failure, the dynamic IP addresses on that node are seamlessly migrated to another node in the cluster. When the offline node is brought back online, SmartConnect can automatically rebalance the NFS clients across the entire cluster to ensure maximum storage and performance utilization. For periodic system maintenance and software updates, this functionality allows for per-node rolling upgrades, affording full availability throughout the duration of the maintenance window.
Nondisruptive hardware upgrades

Node addition
With Isilon, adding additional capacity, performance, and consolidating applications is a simple, nondisruptive process. Once connected to the back-end InfiniBand and front-end Ethernet networks, new nodes are powered on and easily joined to the cluster from either the front liquid crystal display (LCD) control panel or the WebUI. New capacity and performance is fully available after a 60-second node join process, and the cluster automatically balances data and client connections across the new hardware.

Hardware refresh
Old hardware can also be seamlessly removed, or SmartFailed, from the cluster and replaced by more current node varieties. During this process, the data and metadata blocks on the SmartFailed node are restriped across the remainder of the cluster, and client connections are migrated and balanced across the other available nodes.

Pull-out node maintenance
Individual nodes can also be temporarily and nondisruptive removed from an Isilon cluster for maintenance and minor field upgrades by using “suspend” and “resume” commands.

Nondisruptive software upgrades

Minor version rolling upgrade
For minor releases, the Isilon cluster is upgraded node-by-node in a rolling upgrade process. During each node’s brief reboot period, existing client connections are migrated to other nodes and new connections are routed to other available cluster resources. As such, all the node reboots are serialized within the outage window. OneFS minor releases contain no major file system or kernel structural changes; an example would be upgrading from OneFS 6.5.5.7 to OneFS 6.5.5.8.

Major version simultaneous upgrade
For major releases, when there have been significant changes to the file system, kernel, or firmware, the upgrade procedure typically involves a quick, full-cluster reboot, or simultaneous upgrade. This fast reboot typically occurs in about two minutes, regardless of cluster size, composition, or capacity usage.
All cluster upgrade preparation tasks are completed ahead of time in order to minimize the reboot window. The brief, two-minute upgrade interval is measured in terms of protocol availability or “service responsiveness,” as defined earlier. Additionally, any data layout changes are implemented post-reboot, when the cluster is fully available again. An example of a major release upgrade would be moving from OneFS 6.5 to OneFS 7.0.

In preparation for an upgrade, a comprehensive pre-upgrade check is run against the cluster and its array of data management and protection modules. This helps ensure that the cluster is in prime condition for upgrade. The administrator is alerted to, and required to resolve, any outstanding cluster configuration, free space, and other environmental issues before an upgrade job is allowed to run.

**Figure 2: OneFS WebUI upgrade options**
**Replication failover/failback**

Another approach to managing both major and minor release upgrades, as well as other cluster maintenance, is utilizing a target cluster mirror and EMC Isilon OneFS SyncIQ™ parallel replication. In this scenario, clients are redirected, or “failed over,” to another cluster while the original source cluster is upgraded as shown in Figure 3. Once complete, clients are “failed back” to the original source cluster, and the target is demoted back to a read-only replica.

![Figure 3: Failover/Failback with OneFS SyncIQ](image)

**Fast recovery from unplanned outages**

Fast reboot is not only applicable for major upgrades. Other physical events—data center power outages, over-temperature shutdowns, hardware, or software failures—also benefit heavily from the ability of individual nodes, or the entire cluster, to fast boot in a consistent state.

**Client connection behavior**

All the clients attached to a cluster will have their I/O requests suspended during SmartConnect failover and rebalance operations. Although all clients experience a suspension of I/O during session transfer or rebalance, many will recover their sessions gracefully and resume operations. However, some—particularly Windows clients—may need to re-establish their connections. This is dependent on the protocol, the length of the session failover, and the characteristics of the particular application. Additionally, NFS failover is only currently available for clients using NFS versions 2 or 3.
**NFS hard mounts**

NFS clients that have a hard-mounted OneFS export will continue to attempt reconnection indefinitely. Therefore, client sessions are unaffected by a node reboot, unless the application issuing the request happens to time out waiting for an NFS response. This can typically be addressed by extending the application timeout window.

**NFS soft mounts**

Soft-mounted NFS clients continue their reconnection attempts until their configurable session timeout limit is reached. However, although soft mounts may reduce the possibility of client instability during failover, they can expose applications to the potential for session inconsistency. As such, soft NFS mounts are recommended only in cases in which client responsiveness is the main priority.

If Transmission Control Protocol (TCP) soft mounts are not possible, the risk of soft NFS mounts over User Datagram Protocol (UDP) can be reduced by specifying long retransmission timeout values and a relatively large number of retries in the mount options. For example, in order to set the NFS timeout to sixty seconds and five retransmission attempts, the following client mount options may be used:

```plaintext
timeo=60, retrans=5.
```

**SMB**

Cluster connections from Windows SMB v1 and v2 clients typically result in a dropped session and a subsequent re-attach. This protocol shortcoming is a challenge common to all storage systems.

Ideally, cluster upgrades are scheduled in advance and preceded by requests for SMB users to drop their connections voluntarily. However, any remaining SMB sessions can be ended via the “terminate TCP connections” options in the upgrade section of the cluster’s WebUI.

**Other protocols**

Stateless protocols such as HTTP will typically recover gracefully from a SmartConnect failover operation. With stateful protocols (for example, File Transfer Protocol [FTP]), state is typically lost and the operation must be retried by the client.

**Applications**

Application session behavior during a SmartConnect failover operation is very much specific to the application itself. Generally, if timeout-based, application parameters can be tuned to increase timeout intervals to exceed node reboot times as a means of avoiding application disruption. Please consult the particular application’s best-practices guide for specific details.
Conclusion

Organizations of all sizes around the globe are dealing with a deluge of digital content and unstructured data that is driving massive increases in storage needs. As these enterprise datasets continue to expand to unprecedented sizes, always-on data availability becomes ever more crucial. As such, a new approach is needed to meet the availability, protection, and performance requirements of this era of Big Data.

EMC Isilon enables organizations to linearly scale capacity and performance up to 20 petabytes, 106 GB per second, and 1.6 million SPECsfs 2008 IOPS. Moreover, they can do this within a single file system—one which is simultaneously performant, resilient, and simple to manage. Built atop enterprise-class hardware and powered by the highly acclaimed OneFS distributed file system, Isilon scale-out NAS solutions deliver the following key tenets:

- Unparalleled levels of data availability and protection
- No single point of failure
- Fully distributed single file system
- Industry-leading tolerance for multi-failure scenarios
- Pro-active failure detection and pre-emptive, fast drive rebuilds
- Flexible, file-level data protection
- Fully journalled file system
- Transient availability and nondisruptive operation
- Extreme scalability paired with management simplicity

About EMC Isilon

EMC Isilon is the global leader in scale-out NAS. We provide powerful yet simple solutions for enterprises that want to manage their data, not their storage. Isilon products are simple to install, manage and scale, at any size and, unlike traditional enterprise storage, Isilon stays simple no matter how much storage is added, how much performance is required, or how business needs change in the future. We’re challenging enterprises to think differently about their storage, because when they do, they’ll recognize there’s a better, simpler way. Learn what we mean at www.emc.com/isilon.


Contact EMC Isilon

www.emc.com/isilon

505 1st Avenue South, Seattle, WA 98104
Toll-free: 877-2-ISILON | Phone: +1-206-315-7602
Fax: +1-206-315-7501 | Email: sales@isilon.com