DataDirect NETWORKS

www.datadirectnet.com

White Paper

Best Practices for Architecting a Lustre-Based Storage Environment

Summary of Best Practices

1. A Best Practices Lustre storage system is built without any single point of failure – consisting of both storage and server networking hardware that is completely capable of automatic failover.
2. External, highly-available metadata storage should be used in all cases. Best Practices Lustre deployments use an externally RAIDed (active-active) storage target which is mirrored and highly optimized for transactional data service – usually configured as either a RAID 10 or RAID 50 partition.
3. Best Practice Lustre storage environments should employ remote power switches, as opposed to IPMI, to manage Lustre STONITH and failover services.
4. Additional failover resilience can be built into the storage architecture by connecting the OSS and MDS failover pairs together with a separate serial cable, to create a second monitoring network which is in addition to the primary 10/100/1000 Ethernet monitoring network.
5. To deliver high performance, Best Practices Lustre systems are built from high performance storage and server elements. This approach ensures reduced I/O overhead, allowing for maximum computation time – but also ensures lower administration costs by reducing system sprawl and reducing single points of failure.
6. Because drives can and will fail – Best Practices storage system performance should be measured over the system lifespan. It is important to understand and plan for how systems withstand and handle drive error and repair events.
7. Predictable striped file performance can only be derived from storage devices that are capable of withstanding and reducing the impact of common component failures in a striped file environment.
8. Long-term data integrity must be protected by intelligent RAIDing methods. Beyond traditional RAID features, Best Practice persistent Lustre data stores require: RAID 6 (or more), low drive rebuild times, read parity verification and predictive failure monitoring.
9. Best Practices deployments stay on, or very near to, the supported Lustre kernel path. These environments also receive the most responsive support as troubleshooting and issue re-creation are more easily performed.
10. Best Practices deployments are built from experience – storage systems present different challenges to different users and it is advisable to understand and deploy storage technologies that have been successfully adopted by the Lustre community to reduce deployment complications.
11. Best Practice MGS services are built on stand-alone MGS nodes that have access to externally RAIDed storage.

Table of Contents

Introduction

The rise of clustered computers has created a proliferation of scientific, analytic and research data. Initially, the exclusive domain of government laboratories and universities, corporations are now abandoning SMP-style computational approaches and adopting cluster computers for applications such as seismic data processing, financial analysis, product development and simulation, pharmaceutical development and web content serving. This computing revolution has created a growing storage infrastructure challenge as traditional storage methods struggle to keep pace with the speed and parallel service requirements of scalable compute environments.

To address the challenges that cannot be solved by traditional storage approaches, clustered and parallel file systems are being adopted to virtualize the storage infrastructure and scale capacity and performance beyond what is possible with single NAS, SAN or DAS systems. Most of these technologies leverage open storage platforms to scale I/O services to the compute farm – technologies such as IBM's GPFS, Sun's open source Lustre File System and the iBrix Fusion file system have been proven to support concurrent file and file system access across thousands of file system clients – utilizing dozens to hundreds of file servers all presenting a common namespace to a load-balanced cluster.

Introducing: The Lustre File System

Lustre is a leading technology in this new class of parallel I/O technologies and is an emerging open-source standard for scalable HPC and cluster computers. The Lustre File System currently powers over two Petaflops of aggregated computing capability measured across scalable cluster computers all over the world. This next generation file system technology is currently used on nearly 1/3 of the world's *Top100* fastest computers. Applying intelligence throughout its unique architecture, the file system separates metadata services away from the data path and through intelligent lock management – scales throughput and file operations performance as the system grows.

Lustre turns commodity servers into smart storage management devices that serve and store data objects. The objects are dynamically distributed horizontally across the servers, shattering the performance limitations common on traditional storage systems and achieving single volume throughput levels greater than 100GB/s.

Lustre Best Practices: Born from Experience

Designed to suit the compute and deployment requirements of a broad array of computer users, the Lustre file system is highly configurable and supports a broad array of permutations. Built on the open-source Linux operating system, Lustre software is made available to the open source community under the GPLv2 open source license and can be adapted and evolved within the community for feature addition and bug-fixing purposes. The open source community has embraced this technology and Lustre is now deployed within a wide variety of government, university and corporate storage environments.

The broad configurability and adaptability of the Lustre file system presents a combinatory challenge when deciding how to deploy a Lustre environment designed for performance, reliability and operational efficiency.

DataDirect Networks is a leading provider of scalable storage systems for performance and capacity-driven applications and the company's Silicon Storage Architecture (S2A) appliance and storage

systems are commonly used to enable next-generation cluster scaling - enabling cluster I/O to several hundred gigabytes per second and petabytes of storage capacity. With over 5 years experience in working with Sun / Cluster File Systems on the world's largest Lustre deployments, DataDirect Networks' has developed an extensive body of experience with Lustre storage environments and a deep understanding of deployment and operational best practices. S2A technology is a leading and trusted storage platform chosen by Lustre users worldwide. Computational organizations, listed in Table 1, all have selected DataDirect Networks S2A Storage and the Lustre File System for their high performance production computation requirements.

Country	Organization	Highest Top500 Rank (Nov 2007)
UK	Atomic Weapons Establishment	35
France	Commissariat à l'énergie atomique [CEA]	19
USA	Indiana University	42
USA	Lawrence Berkeley National Laboratory [LBNL]	9
USA	Lawrence Livermore National Laboratory [LLNL]	1
USA	Louisiana State University [LSU]	32
USA	Maui High Performance Computing Center [MHPCC]	25
USA	NASA Ames Laboratory	20
USA	National Center for Supercomputing Applications [NCSA]	14
USA	Oak Ridge National Laboratory [ORNL]	7
USA	Sandia National Laboratory [SNL]	6
USA	Texas Advanced Computing Center [TACC]	22
Germany	The University of Dresden	125
USA	US DoD Engineering Research & Development Ctr [ERDC]	31

And many more government, corporate and university Lustre + S2A deployments worldwide.

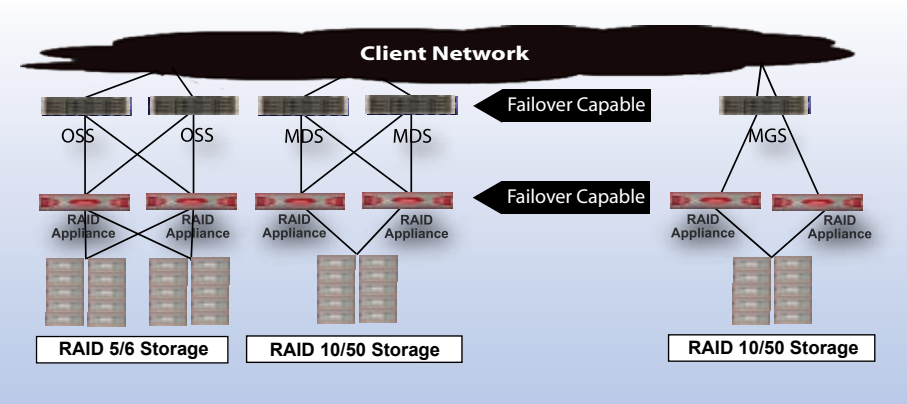
Table 1: Organizations running the Lustre File System with DataDirect Networks S2A Storage Systems. Routine, world-class storage deployment experiences with the Lustre File System have helped DataDirect Networks develop an extensive understanding of Lustre Best Practices.

What follows is a series of known "Best Practices" to consider when deploying a Lustre storage environment. These methods and concepts were developed through numerous large-scale, mission-critical Lustre deployments and optimization efforts. This guide is designed with the sole purpose of decreasing the time-to-deployment and enabling highly reliable, high-performance Lustre environments.

Best Practice #1: High Availability Requires Storage Infrastructure Failover

Environments that require high-availability also require the highest levels of system resiliency and redundancy. Because the Lustre File System is capable of virtualizing a file system namespace over a number of storage and server elements, and because servers are not fault tolerant – the storage infrastructure must be built from a collection of RAIDed storage elements and provide multiple paths to Lustre metadata and object storage servers to enable high-availability failover services and uninterrupted service continuity.

Diagram: Lustre Environment with Failover



Note: Cabling, monitoring and power control infrastructure not displayed here – detailed later in this document. OSS Cluster nodes are configurable from 2-400+ OSSs. MDS and MGS services do not necessarily need to operate from separate RAID storage systems, they only need separate partitions.

Shared storage is commonly used with SAN and cluster file systems to provide the performance and failure avoidance required to meet production computation requirements. Also known as "SAN storage", "Monolithic storage" or "external storage", these systems are built from dual, active-active RAID controllers and manage a collection of hard disks in separate disk enclosures. These disk enclosures also are commonly built with redundant interfaces to further support the data availability capabilities of the system.

Shared storage is designed to maximize data integrity SAN and DAS environments – and can be deployed in a SAN or Direct-Connected configuration within a Lustre storage environment. In a SAN configuration – a single Fibre Channel, Infiniband or iSCSI SAN fabric can manage all Object Storage Servers and Targets across one or many shared storage devices. However, SANs are not necessarily required with the Lustre File System. The parallel storage management capabilities of the file system enable separate scalable storage building blocks composed of Lustre OSSs that are direct-connected to a shared storage system, designed to handle multiple levels of system failure while delivering high levels of throughput These building blocks can be iterated within a single Lustre environment to scale both performance and capacity of the storage architecture.

There are, however, other technologies by which Lustre users can build Lustre environments. These technologies often do not enable either controller or OSS failover services and inject single-points-of-failure into a Lustre environment. The trade-off associated with an architecture that cannot withstand a server or storage controller failure can be serious – and has an overall impact on total cluster productivity levels. Architectures without failover capabilities reduce the long-term duty cycle of a compute cluster and sporadic system downtime should be factored into overall performance-delivery calculations and uptime-requirements planning.

Architecture Design Consideration: Storage Servers vs. Shared Servers

Eg. Storage Server – composing both the host and the storage target.

Definition: Storage Servers

Storage Servers are network-connectable block and file devices, generally built from standard, off-the-shelf motherboards and designed to house 12-60 drives - managed by an embedded SW or HW RAID device. Additional drives can be added and supported by that system via a JBOD or SBOD connection to additional drive shelves.

The systems are connected to the storage environment as "hosts" to the data network - which is typically built from Gigabit Ethernet or 10 Gigabit Ethernet switching or with increasingly popular Infiniband network technology.

Examples of possible Storage Servers used with the Lustre File System include:

- 16, 24, 36 & 48 Drive "whitebox" Storage Servers
- HP Proliant 380 Storage Server
- Dell Powervault Disk Storage Enclosures w/ Poweredge Servers
- Sun x4500 "Thumper" Storage Server

Because Storage Servers are built from mainly commodity technology and are designed with less internal and external system redundancy – they are in many cases less expensive to build a Lustre environment with as compared to an environment built from shared storage systems. While the initial economic appeal of these systems is clear – there are data availability ramifications associated with deploying these systems in striped and non-striped file Lustre environments both which are severe and not considered Best Practice components. Furthermore, the availability of data in a cluster environment will impact the effective value of the total storage cluster.

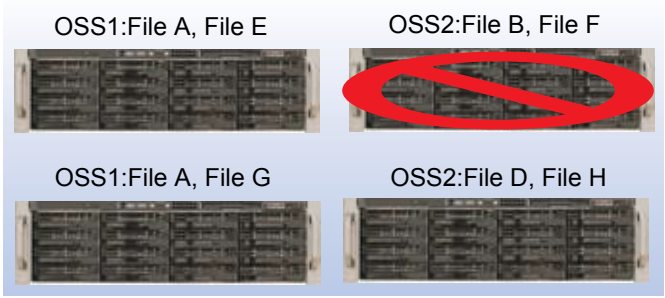
Spotlight: Lustre File Striping

The Lustre File System provides the ability to stripe files (by distributing intelligent data objects) across a number of Lustre Object Storage Servers (Host Systems) and Lustre Object Storage Targets (LUNs) to enable fast, concurrent file write and read capability. The additional CPU, network and disk resources that can be brought to bear when leveraging Lustre File Striping capabilities ensure that maximum storage cluster resources participate in a file I/O event.

Scenario #1: Non-Striped File Environments

In the case of a Lustre storage cluster built from Storage Servers where the policy is such that each file is only striped within one Storage Server (because that Storage Server is limited by its lack of failover capability) when any Storage Server fails, all of the files which are resident on the failed server node will be unavailable for the duration of the failed system's outage period. Consider the following diagrams:

Example #1: Healthy system with 4 x servers attached to disks (single line to disk symbol).



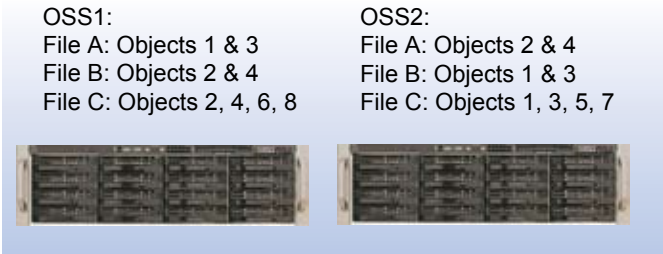
Example #2: Same configuration; failed node – OSS2's files not available due to an OSS failure.

As demonstrated, the inoperative state of any server will render the data on the failed server unavailable for the duration of the system outage. Because Storage Servers offer a single path to the data they manage, these servers are single-points-of-failure.There is no ability to failover or multi-path services to access the offline data.

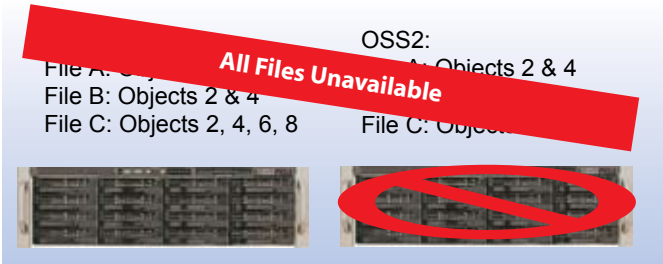
Scenario #2: Striped File Environments

Lustre File Striping is a desirable method of accelerating file performance by bringing many to all of the storage cluster resources to bear during a file read or write operation. Examples of file types that benefit from striped file performance include: application checkpoint files, seismic data sets, satellite ingest files, visualization files and more…

In the case of a Lustre storage cluster built from Storage Servers where the policy is such that each file is striped across multiple Storage Servers (because Storage Servers are single points of failure and have zero failover capability) when any Storage Server fails, all of the files which touch the failed server node will be unavailable for the duration of the failed system's outage period. This results in a considerably more severe operational impact as compared to Scenario #1. It should be noted that cross-OSS file striping is highly inadvisable in Lustre environments built from Storage Servers. Consider the following diagrams:



Example #3: Healthy system with 2 x OSS storage servers with 16 disks each. A file is striped across both servers and the objects are round-robin placed for load balancing.



Example #4: Same configuration; failed node. Because the architecture contains many single points of failure, all striped files touching the failed node are unavailable during server outage.

As demonstrated, the inoperative state of any server will render any file (that is striped over the failed server) to be unavailable for the duration of the system outage. If the file system policy is set so that all files are striped across all resources for maximum system performance – the entire storage volume is highly susceptible to data unavailability because it is built from a collection of single-points-of-failure. As the storage environment grows with additional Storage Servers, the file system's propensity for failure increases at a greater than linear rate.

Application Note: Network Mirror of Storage Servers (DRBD)

The UNIX community has developed various open source methods for RAIDing networked Storage Servers – creating cross-server RAID 1 (mirrored) devices. The open-source DRBD Linux utility is a popular choice for this style of network device mirroring. In the case of a DRBD-based cluster, Lustre OSTs are cross-mirrored across a pair of DRBD servers so that they manage an active/passive set of LUNs. This approach presents several challenges with respect to scaling and availability:

- Cost: The cost of duplicating storage and server resources substantially reduces storage HW Return On Investment, increases the overhead of system administration, and brings the system cost to a point where it is as expensive or more expensive than more capacity-efficient, highly-redundant RAID systems such as mid-range RAID 5 or RAID 6 shared storage devices
- Reliability: the consistency and coherency guarantees of a failover-capable DRBD cluster are less rigid than more traditional storage architectures. If a server fails without completing the mirroring services to the failover OST – it is possible that the Lustre metadata server becomes inconsistent with the failover OST state – which could create data corruption instances.

Metadata Storage Best Practices

It should be noted that in all cases, it is advisable to place Metadata Storage on shared storage that is accessible from a pair of failover-capable servers. This shared storage be built from a mirrored RAID set which provides optimum fault tolerance and performance to maximize the reliability and availability of the clustered storage environment.

Because the availability of the entire storage infrastructure is tied to the MDT, and because the MDT is a relatively small investment compared to the HW costs of the OST storage, an additional investment in highly-available metadata storage is justified by the safeguarded cluster uptime.

Architecture Design Consideration: Failover Networking

Lustre failover services are critically dependent on a healthy and predictable failover networking foundation. If resources are not properly powered down and powered up to failback – the storage cluster can be highly susceptible to data corruption – as Object Storage Server (OSS) or Metadata Server (MDS) resources are improperly managed and are left available to write concurrently to a single Object Storage Target (OST) or Metadata Target (MDT). As such, a few simple configuration guidelines can help ensure that OSSs and MDSs never concurrently write to any single volume. These guidelines include:

Remote Power Switching vs. IPMI Power Management

To manage failover, Lustre systems require some technology to effectively power on and off a Lustre server. This level of power management ensures that Lustre OSS X never automatically "powers on" while its failover pair (OSS Y) is writing to OSS X's OSTs. If OSS X and OSS Y ever concurrently write to a common OST – that OST will very likely become corrupted and data integrity will be compromised.

The solution to this problem is to Shoot The Other Node In The Head (aka: STONITH) to ensure that the failed node does not prematurely wake up and unexpectedly begin writing to a failed-over OST. Common Linux HA utilities such as SuSE's Heartbeat and RedHat's Cluster Manager (aka: clumanager) are the policy engines by which STONITH services are managed.

To execute STONITH services – it is also required that some remote server power management technology is configured into the Lustre OSS and MDS architecture. Two common technologies are:

- The Intelligent Platform Management Interface [aka: IPMI]: IPMI is a server controller which operates independently of the server operating system (OS) and enables remote console administration. IPMIAISO performs automatic systems management and administration (which is operated from a separate network and power interface from the rest of the server). The IPMI processor can be used to completely power control the system (based upon policies set to be triggered by server and network events), but also performs additional tasks such as server environmental monitoring.
- Remote Power Switches: These remotely controllable power distribution units are rackmounted power switches. They respond to power and network events and can switch power outlets on and off 'per outlet' on a policy-basis.

Best Practices: Because DataDirect Networks has witnessed events where both IPMI cards fail simultaneously – although this is a rare occurrence – the Best Practice recommendation is to use a data-center grade remote power switch to perform STONITH services in a Lustre storage cluster.

Redundant Networking leveraging System-System Serial Cabling

In addition to the standard 10/100/1000 Ethernet failover and systems monitoring network which is advisable for Lustre deployments, DataDirect Networks also sees great value in adding a separate point-to-point failover monitoring capability to the Lustre cluster to effect greater resiliency.

By connecting two failover servers via an additional serial cable into a failover pair, a series of independent point-to-point server networks can be deployed to monitor and manage failover services. This avoids potential corruption incidents which result from primary monitoring network failure or flakiness.

Best Practice #2: Maximizing Computational Capability by Minimizing I/O Overhead

The design goals of the Lustre File System have all centered around the enablement of massively scalable cluster computing and delivering the highest levels of I/O performance in order to minimize the overhead of data read and write service. This allows clusters to spend less time waiting for storage and more time computing.

Because many storage services performed with the Lustre File System are defensive in nature, such as system checkpointing, it is incumbent upon system architects to ensure that defensive I/O activities do not overly burden the productivity of the computation workflow. Additionally, long wait times for either read or write intensive I/O operations dilute the overall effectiveness of the cluster computer and can and should be avoided. This can be achieved through the implementation of several technologies.

High-Performance Storage Platforms

Because storage devices perform at different speeds and have different block and object service capabilities, it is important to develop an architecture that meets or exceeds peak I/O requirements. "Peak I/O requirements" are not defined as the maximum that a system can perform I/O at (which is usually a product of network bandwidth), but is rather the maximum amount of combined read and write capability that the system will need at any one point in time. To architect a system for high performance, it is advisable to configure a system with high performance storage subsystems that can efficiently serve data to & from high-performance Lustre servers (OSSs and MDSs). Advances in commodity computing and networking have brought COTS servers to a point where it is possible to deliver 500MB/s to 1GB/s+ from a single OSS. Using the right underlying storage to serve this level of throughput, a Lustre environment can deliver as much as 10GB/s with as few as 10 commodity Object Storage Servers.

Storage systems also tend to perform differently according to differing storage parameters on deployment. While a default storage mode that a system is shipped with may provide an intended level of performance – parameters which are decided upon and set at the time of deployment (to increase data protection levels) may decrease the delivered level of performance from the selected system.

Examples:

1. RAID 5 generally performs faster than RAID 6 but often provides a lower level of data protection.
2. Cache mirroring can reduce the effective performance of a system by as much as 50%, but is required if a user wants to protect write data held in cache.
3. Parity checking during read operations does not come for free (performance-wise), but deficiencies in SATA ECC capabilities make read parity checking desirable, if not necessary.

Quality of Service

In addition to a storage subsystem's performance statistics, it is important to consider the storage hardware's ability to withstand drive, controller and enclosure failures. In the case of a Lustre environment deployed with storage servers (described above in Best Practice #1 section a), obviously the motherboard/controller failover capability is negated by the single purpose architecture. Outside of that, it is important to consider:

The first level quality of service is data availability:

- Quality of Service: Data Availability via RAID Controller Failover

  Because single RAID controllers are not inherently fault resilient, care must be taken when selecting a RAID controller system with dual RAID control elements that operate independently, with independent power supplies, that have cross failover capability.

  In cases where array performance is not capable of meeting the cluster's storage requirements (and the controllers' write caches are enabled), RAID controller cache mirroring is a critical feature designed to ensure data integrity.

- Quality of Service: Maintaining Performance During System & Drive Error Cases

  Because storage systems exhibit different levels of performance as they undergo system and drive error cases and correction events – it is important to measure the performance of your storage infrastructure over time and not simply measure peak or point-in-time performance.

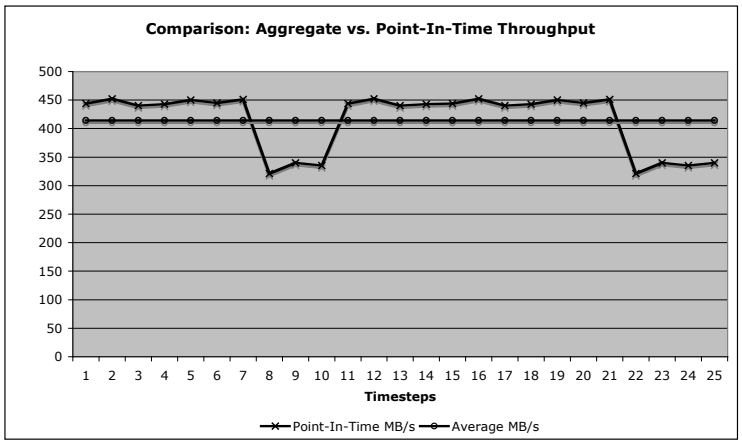
  This measurement becomes especially important when the value of the storage procurement is measured on a $/performance basis. In order to ascertain the true performance of a storage system, users must understand the failure characteristics of the drives that the system will use, the MTBF of all of the system components, and how the system manages hardware errors.

  Any number of drive error events can contribute to a storage system performance decrease. Contributing factors include, drive slowdowns, drive time-outs, drive restarts, request retries, and full and partial drive rebuilds. Of all the various drive error/correction events that can impact quality of service and performance - hard drive rebuilds are especially detrimental to system performance. This is because:

  o The rebuild event takes cycles from the parity engine

  o Internal system bandwidth is taken from normal I/O process and reallocated to rebuilding data

  o The healthy disks in a parity group experience additional read activity because the new drive data needs to be provided from the rest of the RAID set. This read activity also interrupts normal reads and writes to the parity set undergoing the rebuild and reduces the predictability of the I/O in the degraded parity group during the course of the rebuild.



Graph 1: Analysis of system throughput over time. Aggregate performance is 9% lower than peak performance due to the system's inability to protect application performance against drive failure issues.

As seen in Graph 1, the system point-in-time performance demonstrates a certain peak MB/s performance level (while in a healthy state) – however the system exhibits a lower aggregate performance level over a course of time as it experiences and deals with hard disk drive issues. The graph depicts a real-life scenario where 1TB SATA drives have taken the storage system into a degraded state for over 25 hours per rebuild. The result is a nearly 9% overall system performance decrease over time. These results will vary across storage technologies from both a performance and from a failure-management perspective.

Reducing Administration Overhead by Reducing System Sprawl

The benefits of cluster computing are fundamentally derived from being able to scale computation cost-effectively with a series of cluster nodes, as opposed to scaling computation within a node through adding additional CPU and memory resources. This commodity scaling approach allows cluster computer centers to scale resources cost effectively by leveraging low-cost server technology. Clustered, distributed file systems apply the same cluster principles to storage scaling by enabling the aggregation of independent storage servers and arrays into a single, common volume.

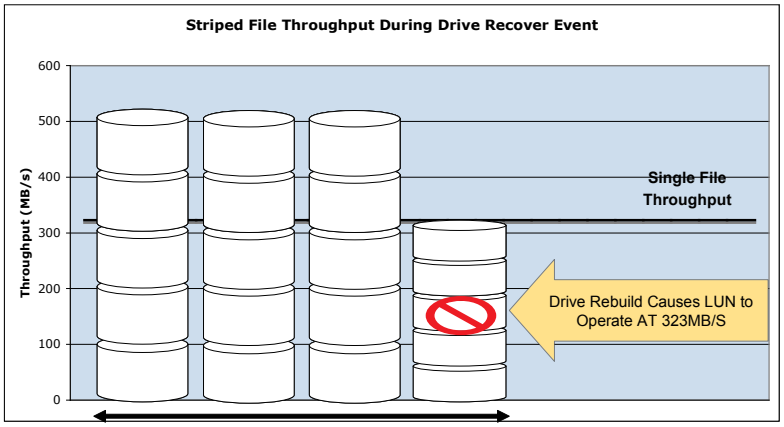
While clustered storage has enabled file systems to scale to previously impossible levels of capacity and performance, this is not without its own set of implicit costs. As storage clusters grow larger and larger, there is a direct correlation between the number of systems to be managed and the effort required to manage these systems. This administration effort has a direct effect on system administration costs and budgets. By reducing the amount of network, control and server resources in any given storage environment, data center managers can minimize administration time and costs associated with deploying clustered storage. This approach also reduces the amount of controller and server failure elements and can increase system uptime.

Best Practice #3: Ensuring Predictable Striped File Performance

Scalable file system namespaces can be composed from hundreds to tens of thousands of hard disk drives managed by a series of Lustre Object Storage Servers and underlying storage subsystems. Because the chance of drive failure and system error correction increases on a more than linear basis (as the amount of drives increases in any given storage cluster), it is important to consider the failure impact on the overall performance of a striped file system.

In a striped file environment, the performance of a file read/write operation is determined by the slowest performing component in the stripe set. As such, storage architects must take great care to avoid designing a system that cannot support the drive failures (which can be commonplace in a system built from thousands of hard disk drives).

In an example where a system checkpoint happens over 5 LUNs/OSTs spread across five RAID controllers - which are each comprised of four data disks and one parity disks (five disks total per LUN, twenty disks across five LUNs) - it is likely that a system will eventually experience one or more drive error events concurrently in the system within one or many prolonged I/O operations.



Graph 2: File striped across five LUNs – each in separate RAID groups is performance-limited to the slowest device in the stripe set (as the device does not protect application performance against drive failure issues).

As seen in Graph 2, while four of the five devices are performing healthily, the checkpoint application cannot finish until the 5th device completes its (slower) write operation. Because the file is striped across all 5 devices, all of the compute nodes involved in the checkpoint operation cannot resume normal compute operations until the slowest storage element has finished the file I/O operation.

Considering the near-certain likelihood of spurious drive event(s) in a clustered storage environment (particularly when deploying commodity SATA drive technology), it is an advisable Best Practice to deploy storage devices that are capable of withstanding and reducing the impact of common component failures in a striped file environment.

Best Practice #4: Protecting Large, Persistent Data Stores

While the Lustre File System has previously been used for /scratch or /temp space – the continuous availability and stability improvements made have enabled storage managers to deploy Lustre technology to house persistent data stores. Large, persistent data storage projects at facilities such as CEA, Indiana University and others have highlighted the benefits of using Lustre as more than just a temporary storage service.

The Need for Intelligent RAID

SATA disk technology is often selected to store this persistent data because of the cost and space efficiency of SATA technology (as compared to enterprise FC and SAS drives). As storage managers continue to expand their Lustre environments to store increasingly persistent and important data, Best Practice storage architectures are designed to protect SATA data over the on-disk lifespan. Because additional considerations must be made with storage systems for long-term data, as opposed to temporary data which can be easily re-created, storage managers must keep a keen eye on long-term data integrity and availability. This issue generally forces a discussion toward more intelligent RAID technologies, which include:

- Double Disk Failure Protection: RAID 6

  The emergence of SATA drive technology, as a clustered storage standard, has heightened the need for large SATA pools to be built from RAID 6 protected storage. In a n+2 parity group, a RAID 6 drive set is configured with two on-line parity drives to prevent a data corruption episode resulting from the loss of two drives in any one interval. The presence of the second parity drive provides system administrators with a comfortable window to repair or replace a failed drive without fearing a second drive loss.

  In large SATA pools, a RAID 6–based OST configuration is without question a Best Practice storage system design choice.

- Drive Rebuild Acceleration to Minimize Vulnerable State

  Incremental increases in hard drive capacity have not been accompanied with a corresponding increase in drive rotational or seek performance. Because drives have not gotten materially faster as areal drive density has increased, the amount of time it takes to rebuild data on a replacement or previously-offline drive has now exceeded the 24 hour mark. In the case of 1TB 7200RPM SATA hard disk drives, drive rebuild times can take over 24 hours. During this 24+hour period, the system is effectively in a RAID 5 mode, whereby there is only a single parity drive in an n+1 configuration – and is therefore susceptible to a double-disk failure for more than a day.

New drive maintenance technologies have emerged which allow for systems to accelerate the drive rebuild cycle in various error cases. These include:

  - Distributed drive rebuild capability
  - Partial drive rebuild capability for reset and rebooted drives

These Best Practice technologies can reduce the burden of hard drive rebuilds and can shrink rebuild times for the largest drives from a day to as little as minutes

- Read & Write Data Integrity Verification & Correction

  Known as Unrecoverable Bit Errors (UBE) or Silent Data Corruption Events - SATA UBEs are triggered when a when a disk's magnetic bits lose the ability to hold data - or data is corrupted on media. Another source of corruption is the result of a buffer or head misread. Because SATA drives are not equipped with enterprise ECC capability – these are the leading causes of read data corruption (because storage systems are not generally equipped to handle such events).

  With a combined, installed capacity of over 100PB worldwide, much of this using SATA technology, DataDirect Networks has identified and witnessed UBE's as a very real issue in the SATA storage industry. Great care must be taken to select a Best Practice architecture which can identify and correct read corruption events.

- Predictive Drive-Failure Diagnostic Tools

  Increasingly, hard disks are able to provide more predictive information to storage administrators. This enables administrators to take a more proactive role in hard disk replacement. By knowing in advance that a certain drive may fail, the failure instance can be avoided or reduced by ensuring that:

  - Time is allotted to service the system (prior to the drive failure)
  - All necessary equipment and spares are available to service the system

  A hard drive administration standard is emerging for the monitoring, logging and analysis of drive information about their operational state – this tool is known as the S.M.A.R.T. (aka: SMART) utility. The Self-Monitoring, Analysis and Reporting Technology utility is used on a wide variety of platforms to log and present precautionary steps to avoiding a hard drive crisis.

  In addition to SMART, there are many robust drive/system diagnostic tools available from leading storage controller manufacturers that may also be deployed for Best Practice drive monitoring.

Checksums

Beginning with Lustre version 1.6.3, some versions have the default system parameter that the file system run a checksum against all read operations to verify integrity of the data that is being read. While having this feature enabled is absolutely a Lustre Best Practice – there are some ramifications associated with keeping this feature enabled at either runtime or compile-time.

Lustre checksum: The Lustre checksum uses a CRC32 algorithm to detect single bit errors and swapped and/or missing bytes. This algorithm requires a fairly significant amount of CPU power to execute and can impact performance adversely if the OSS and MDS server hardware is not powerful enough to handle both the I/O, file system administration and checksum services concurrently.

Lustre Checksum vs. Storage Without Checksums:

Because most storage arrays do not have the ability to check and correct the integrity of on-disk, in-flight data during a read operation – it is a highly advisable that checksums are enabled to guarantee the consistency of read data.

The Best Practice recommendation in this case is to enable checksums and to be sure to have state-of-the-art OSS CPUs to handle the increased computational load.

Lustre Checksum vs. Storage With Checksums:

While most storage systems available today do not have the ability to verify the integrity of read data, there are a few that do have this ability, such as the DataDirect Networks S2A family of storage arrays. These systems have the ability to check parity data going to the hosts, and can also correct this data in real-time. This real-time data correction capability is a by-product of the system being designed to read both source and parity data during each read operation and maximizes data integrity to the HBA.

In the case of an array that has the ability to validate and ensure data integrity of the host, it is advisable (but not required) that storage administrators configure Lustre file systems with the checksum capability enabled. This configuration will further ensure the integrity of the read data and protect the overall storage cluster from possible bit-flip in the HBA buffer cache and ensure end to end data integrity.

In the interest of maintaining performance in the storage system, it is an advisable Best Practice to configure OSSs and MDSs with state of the art CPUs, enable Lustre checksums, and design into a storage cluster a RAID subsystem that can correct the majority of UBEs (usually resulting from SATA ECC inadequacies) without forcing a reread from the OSS.

Best Practice #5: Proven Architectures = Proven Success

The final Best Practice presented in this document is not so much of a recommendation for any one specific configuration – as it is an advisement for keeping a mindful awareness of the Lustre community's undertakings. Because Lustre is highly configurable, the open source nature of the technology can invite adaptation for specific environments and for a specific functionality. It is important to keep in mind the advantages of staying within a known, good range relative to what is deployed in the Lustre community and what is supported by Sun/CFS.

Kernels

While it is possible to modify Linux kernels to suit a variety of kernel flavors that are not supported or provided by Sun/CFS, it is important to note:

- Because of the variances in Linux kernels, moving from a supported kernel to an unsupported kernel is not as simple as just recompiling Lustre code onto a new kernel. There are differences in the dependencies and the mappings within the Lustre code base - where versions are very often tailored to specific kernel releases. Therefore, a migration to a non-supported kernel may often result in a lack of functionality, usability and/or stability. This may require additional code changes.
- As Sun/CFS has a finite amount of resources available for the validation and troubleshooting of Lustre releases with specific enterprise kernels, as it may be possible to build a working Lustre version on a non-supported kernel. The unsupported nature of the kernel upon which that release is built upon will make troubleshooting and issue replication extremely difficult and may result in either undesirable issue resolution latency or even lack of support.

Therefore, it is considered a Best Practice recommendation to use the Linux kernel versions that are supplied by and supported by Sun/CFS and their authorized partners.

Storage

While it is possible to deploy a Lustre file system on top of any SCSI device (which is supported by a Lustre-supported Linux SCSI driver), there are a number of storage permutations possible within a Lustre storage environment. These can take storage deployments down relatively rocky paths. As such, it is a Best Practice to understand and deploy storage technologies that have been successfully adopted by the Lustre community to reduce deployment complications and leverage the community's work with components such as:

- Fibre Channel and Infiniband Adapters
- Multipathing Drivers
- Active/Active Storage Appliances
- LUN and Port Mapping
- and more …

MGS Servers

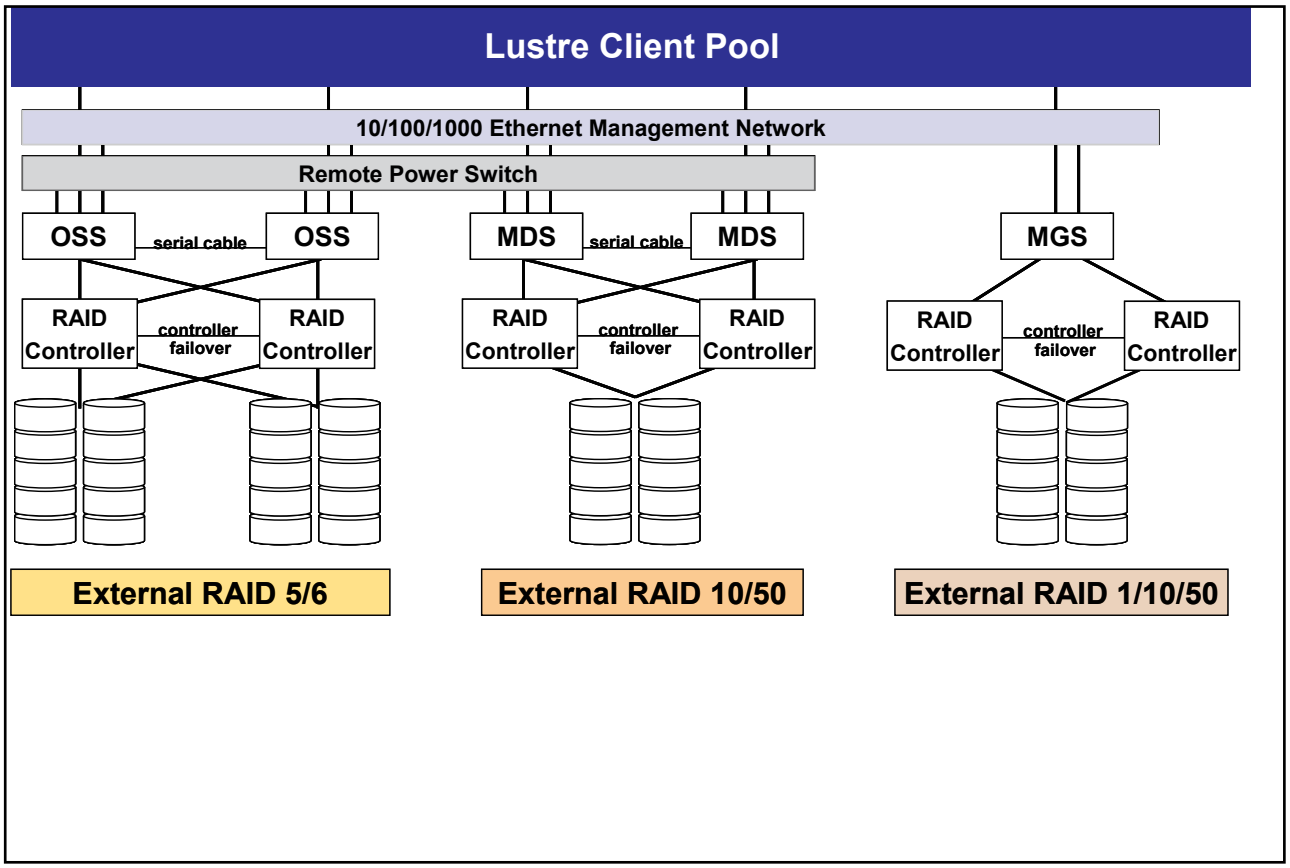
The MGS server is a relatively new convention within the Lustre configuration. Introduced in Lustre 1.6 as part of the mountconf feature, the Lustre MGS automates cluster configuration awareness and management, allowing for the easy addition of resources to an existing Lustre environment.

As this concept is relatively new to Lustre configuration, it is important to remember how to configure the MGS and how best to protect MGS information.

- A Best Practice MGS service is deployed on a stand-alone server to independently monitor the activities of MDS, OSS and Client services. DataDirect Networks has witnessed cases where an MGS has been hosted on a MDS and has observed additional failover complexities when both an MDS and MGS are on the failed server node.
- A Best Practice MGS partition should be hosted on an external storage device which is both RAIDed and available from an active/active set of RAID controllers. This will enable rapid MGS server hot-swap ensuring that MGS partition RAID services are not dependent on the MGS server node.

Best Practice Reference Architectures

Best Practices **Reference** Architecture Diagram



DataDirect Networks S2A Storage Technology: A Commonly Used Storage Foundation for Best Practices Lustre Environments

The Lustre File System has a deep history with DataDirect Networks' S2A Storage technology. Dating back before Lustre version 1.0, the S2A storage system was used as early as 2003 with the Lustre File System at the National Center for Supercomputing Applications. Even in the early days, the benefits of coupling these two technologies was easily seen as CFS and DataDirect Networks broke a worldwide storage performance record to achieve 11.1GB/s on a storage cluster composed of 104 OSS and 140TB. Since that time, the relationship between Lustre and S2A technologies has grown much deeper.

The S2A has become the leading storage platform of choice for high-performance Lustre environments. This is due to the industry leading performance and because the Lustre File System can scale elegantly and predictably with S2A storage technology. S2A storage powers systems designed with several petabytes of capacity, 10s of 1000s of processors and systems designed to deliver over 100GB/s in single volume throughput.

S2A storage is chosen to power Best Practice Lustre File System deployments for many reasons, including:

- **S2A Storage is Designed to Support Lustre Data Availability and Integrity**

  High Availability:
  Designed with Active/Active storage applaince capabilities, the S2A (combined with a number of failover server pairs) is configured to eliminate any single point of failure in a Lustre File System storage cluster. The S2A is especially resilient, and can withstand any number of storage system failures/outages – including drive failures, enclosure failures, drive channel failures and storage appliance failures. All of these reliability features combine to maximize data availability and are especially important as the vulnerability of a data set increases in striped file and striped file system environment.

  Maximum Data Integrity:
  Because long-term data needs long-term data integrity – the S2A is designed to protect the health of a file system designed to contain persistent data sets. The S2A's SATAssure technology are redefining best practices for deploying SATA-based Lustre storage to extend and maximize resiliency without compromising parallel I/O performance.

  By understanding that drives will fail, the S2A and SATAssure compensate for SATA's inherent shortcomings by providing needed, additional protective and preventive measures to maintain data integrity and sustain full access to data - all while sustaining full system performance. This allows the enterprise to deploy cost effective, scalable SATA capacity with confidence.

The S2A SATAssure technology is DataDirect Networks' innovative SATA storage management suite designed for:

o Real-Time, High-Performance RAID 6: DataDirect Networks has designed a highly efficient RAID 6 implementation whereby the storage system derives full performance from a RAID set consisting of 8 data drives and 2 parity drives. Unlike other systems that suffer a performance penalty for protecting against double-disk failures. The S2A's Direct RAID6 ensures that the storage system is protected from the occasional double-disk failure – all while delivering peak system performance.

o Real-Time Data Integrity Verification and Correction: With an understanding of the issues associated with deploying SATA hard drives (that has been developed from deploying hundreds of storage systems into Lustre environments), DataDirect Networks has developed an innovative approach to protecting data which stops corruption at the system level. S2A SATAssure, leverages the S2A's DirectRAID HW-accelerated parallel RAID engine, which performs parity calculation on every read operation as well as on every write operation. This additional integrity check ensures that only known-good data is delivered to the OSS and that corrupted data is corrected in real-time as to eliminate the need for a re-write or re-try at the host level.

In addition to eliminating silent data corruption in real-time, the self-healing capabilities of the S2A transparently monitors the entire storage volume and scrubs/repairs data on disk - which has been corrupted since the write event. This intelligent storage verification technology ensures that the S2A finds and ensures the repair of the bad data before the Lustre read operation occurs.

S2A read parity checking, combined with preventative data corruption detection and repair capabilities ensure the health of persistent Lustre data for the years that it must reside on-disk.

- **S2A Storage is Performance-Designed for High-Speed Lustre I/O**

Through the proliferation of several flagship Lustre systems, which use DataDirect Networks' S2A storage hardware as the underlying high-speed storage platform, both DataDirect Networks and Sun/CFS engineers have engaged to further optimize their respective platforms in the interest of enabling peta-scale I/O and deriving maximum efficiency and performance from a storage cluster. The result of this work can be found in a number of combined capabilities:

Complimentary Transfer Sizes

The design objective, when maximizing file system performance, is to keep both the network pipes full and the disks writing or reading as much as possible during any read or write event. In cases where the I/O performance is hampered by inconsistent network throughput or where storage arrays are excessively seeking and not serving data – the performance of the storage system is not being used to its full potential.

In an effort to reduce the amount of RPCs in a read or write operation – thereby increasing the efficiency of each I/O operation through the data network and on/off of a storage subsystem – Lustre has been designed to support data transfers and block sizes of 1MB and 2MB. These large transfer sizes enable a highly-sequential data transfer from the client to the Object Storage Target and help the disks spend more time writing and reading data and less time seeking – thereby reducing the I/O time of any file operation.

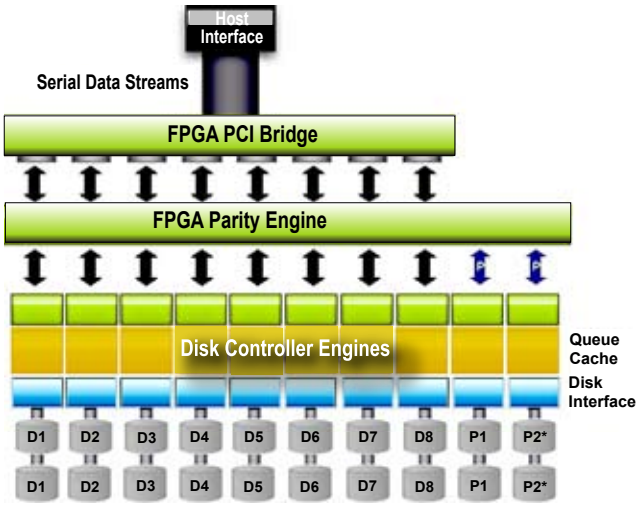
DataDirect Networks has designed the S2A to store and retrieve Lustre transfers with optimum efficiency. There are two primary approaches that DataDirect Networks has taken to achieve this:

• RAID 3 – Style Data Transfer: DataDirect Networks DirectRAID algorithms divide each Lustre block into eight even segments. As seen in Image 1, these block segments are then streamed in parallel on and off of the 8 data drives in a Lustre OST. As such, block I/Os can happen as much as 8 times faster than with normal RAID 5 storage systems. S2A systems stream data onto hard drives as much as 3x faster per drive than competing platforms.

• No Read/Modify/Write Penalty: Unlike traditional storage systems, which upon initially writing data need to read a data block and then modify the LUN to include parity data in the LUN - DataDirect Networks' DirectRAID S2A parity generation technology enables single-operation data writes through the real-time generation of parity during the initial write event. The result of this work is: whereby traditional RAID 5 systems that exhibit multiple I/Os for each block write operation – the S2A increases storage efficiency by generating parity information in real-time and allowing the system to perform more I/Os within the same interval.

High-Performance Building Blocks

By limiting the amount of HW resources required to deliver a desired level of performance, Lustre users can achieve a greater level of system performance (through the consolidation and reduction of components) and a scalable architecture by which performance can be easily scaled. As seen in Table 2, S2A storage systems have historically provided some of the fastest storage performance available in their class. By deploying fewer, faster devices – storage managers reduce storage system administration and ease the scaling units required to achieve 10s to 100s of GB/s.



Additionally, S2A Storage Systems have the unique capability of writing as fast as they read. This is especially important for write-intensive operations such as cluster checkpointing, satellite ingest, content storage and archiving. The high performance write capability of the S2A is as much as 8x faster than that of competing shared storage systems.

DataDirect NETWORKS	 S2A9900	 S2A9550
Supported Disk Technology	SAS & SATA Mix behind single appliance	Fibre-Channel or SATA
RAID Parity Protection	RAID 6 8+2	RAID 3 (8+1+1), RAID 6 8+2
Sustained Throughput	5.3GB/s - 6GB/s Read & Write	2.4 GB/s - 2.8GB/s Read & Write
Scalability	1200 Drives (SAS and/or SATA)	960 Drives
Max IOPS	40,000	14,000
Cache Size	5.0GB ECC/RAID Protected	5.0GB ECC/RAID Protected
Disk Side Ports/Port Type: **Total Back-End Bandwidth**	20 / SAS 4 Lane 24GB/s	20 / FC-2 5GB/s
Host Side Ports	8 x IB 4x DDR or 8 x FC-8	8 x IB 4x SDR or 8 x FC-4

Table 2: Comparison of DataDirect Networks 7th and 8th Generation Silicon Storage Architecture (S2A) Appliances. Among the fastest systems in existence, the S2A A enables Lustre environments to scale efficiently, cost-effectively and without the administration associated with system sprawl that is common with other, lesser performing storage architectures.

Enabling File Striping

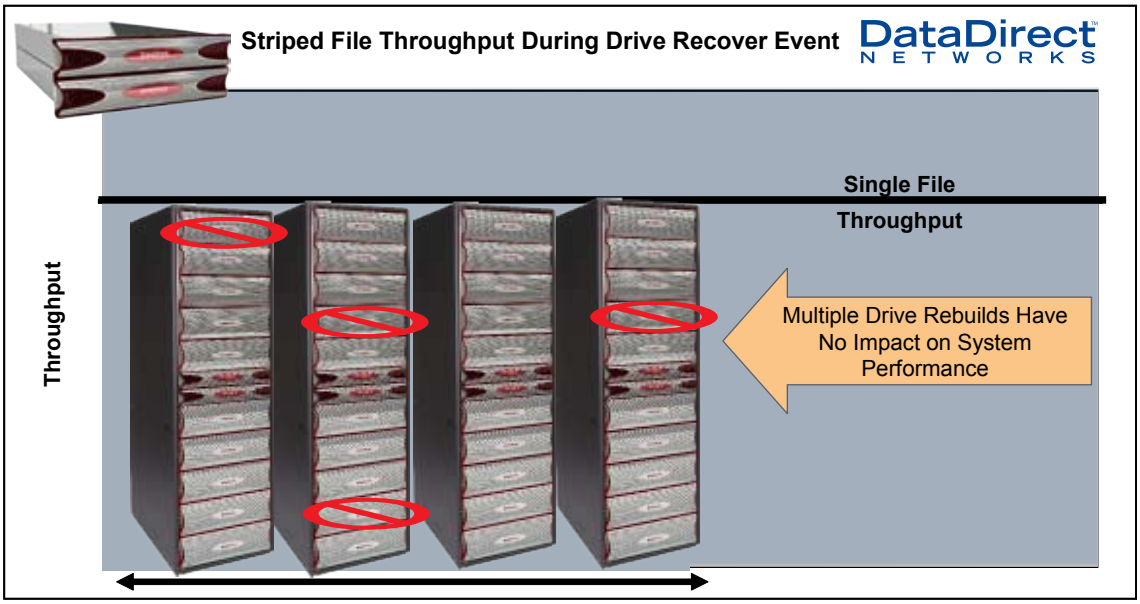
Because drives fail – DataDirect Networks engineers have taken great care in designing a system is capable of withstanding drive failure and supports the predictable I/O levels required when striping data across storage systems. As such, the S2A has been built with very high levels of system resiliency to automatically manage and protect applications from typical component failures, including:

- Zero-Impact Drive Rebuilds: Each S2A Storage System is capable of performing up to four concurrent drive rebuilds without impacting host performance at all. Additionally, intelligent drive rebuild management can optionally ensure that no more than four rebuilds ever happen at one time to guarantee performance predictability.

- Partial Drive Rebuilds: In cases where a reset or a power cycling of a failed SATA drive is required, SATAssure performs partial rebuilds to minimize the rebuild time by only updating information which has been journaled by the S2A while the drive was offline. This capability ensures that drives are rebuilt faster and that the system performance does not fall prey to downtime associated with lost LUNs (because more than 2 drives failed in a LUN where full rebuilds leave the system too vulnerable).

- Zero-Impact Enclosure Failure: S2A systems can lose up to 1/5 of their storage enclosures without exhibiting any performance degradation. Because S2A DirectRAID Parity Engines read both data and parity data in real-time – 1 out of every 5 enclosures can go missing without compromising access or application performance.

In the event of a drive enclosure outage – the S2A's journaled rebuild capability will rapidly bring the offline drives back to full health once the enclosure comes back online. A high-density enclosure consisting of 48 1TB drives can be brought to full health in as little as 20 minutes as opposed to the 12 days that would be required for all of the 48 drives to completely rebuild from scratch.



Stripe Set for a File: Striped Across 4 S2A Storage Systems All but one performing concurrent drive rebuilds

As depicted in Graph 3, the S2A's capabilities ensure that no matter how wide a file needs to be striped to increase file I/O performance – the application will receive a predictable level of performance and not fall prey to degraded levels which are common with other storage systems unable to withstand routine component failures.

Graph 3: A single visualization file is striped across five S2A storage systems – unlike the previous example – the S2As ability to shield the application from drive management issues enables the file I/O to transpire predictably and without degradation.

Conclusion

While many approaches can be taken to building up a Lustre environment, it is important to be aware of the trade-offs associated with various design decisions. This document articulated a number of the design implications and provides the reader with information to make informed decisions to scale high-performance cluster file I/O.

DataDirect Networks technology has been deployed with the Lustre File System for over 5 years now – across 4 generations of S2A technology. While there are a number of storage systems that can be configured with the Lustre File System - only DataDirect Networks S2A technology is the storage platform of choice for so many of the world's largest and fastest Lustre File System environments.

Definitions

DIrectRAID: DirectRAID™ is the S2A's scalable, high-performance, hardware-accelerated RAID engine which is a core component of DirectOS operating system. DataDirect Networks' DirectRAID technology is designed with intelligent algorithms and leverages a high-speed internally parallel system architecture to completely streamline and parallelize the data path resulting in real-time data transfers, write speeds which are as fast as read speeds and on-the-fly RAID parity calculations/correction.

Lustre: Lustre® is a high-performance, multi-network, fault-tolerant, POSIX-compliant network file system for Linux clusters. The key features of Lustre:

- Capacity to run over a wide range of network fabrics
- Fine-grained locking for efficient concurrent file access
- Failover ability to reconstruct the state if a server node fails
- Distributed file object handling for scalable data access

Lustre is a complete, software-only, open-source solution for any hardware that can run Linux. It has native drivers for many of the fastest networking fabrics. Lustre can use any storage medium that looks like a block device.

MDS: The Metadata Server (MDS) provides the network request handling for one or more local MDTs. MDS servers can be deployed in failover pairs, however no more than one active MDS can be deployed in a single Lustre cluster.

MDT: The MDT provides back-end storage for metadata for a single file system.
The metadata managed by the MDT consists of the file hierarchy ("namespace"), along with file attributes such as permissions and references to the data objects stored on the OSTs.

MGS: The Management Server (MGS) defines configuration information for all Lustre file systems at a site. Each Lustre target contacts the MGS to provide information, and Lustre clients contact the MGS to retrieve information. The MGS can provide live updates to the configuration of targets and clients. The MGS requires its own disk for storage. However, there is a provision that allows the MGS to share a disk ("co-locate") with a single MDT. The MGS is not considered "part" of an individual file system; it provides configuration mechanisms to other Lustre components.

OSS: A server node which manages one or more OSTs through performing I/O with the Lustre clients and coordinating file locking with the MDS.

OST: An OST provides back-end storage for file object data (effectively, chunks of user files). Typically, multiple OSTs provide access to different file chunks. The MDT tracks the location of the chunks. On a node serving OSTs, an Object Storage Server (OSS) component provides the network request handling for one or more local OSTs.

RAID 0: RAID 0 is a data striping method whereby data is striped across disks within an array or across arrays within a clustered file system, the data is broken down into blocks and each block is written to a separate device.

RAID 1: RAID 1 blocks are duplicated and mirrored between two hard disk drives. This data protection method provides fault tolerance against disk errors/failures and increases read performance.

RAID 10: RAID 10 data is implemented as a striped array whose stripe components are RAID 1 arrays. RAID 10 arrays can sustain multiple simultaneous drive failures across various stripe components.

RAID 3: RAID 3 blocks are striped across disks with dedicated parity. RAID 3 stripe sets break blocks into bytes and byte-stripe data across all of the data and parity disks. RAID 3 stripe sets provide comparable fault tolerance to RAID 5.

RAID 5: RAID 5 stripes both data and parity blocks across three or more drives. An entire data block is written on a single data disk and parity information is generated and written to a different drive in the parity group.

RAID 50: RAID 50 data is implemented as a striped array whose stripe components are RAID 5 arrays. RAID 5 arrays can sustain simultaneous drive failures across different stripe components but is not configured to withstand as many failures as RAID 10 configurations.

RAID 6: According to SNIA, the definition of RAID 6 is: "Any form of RAID that can continue to execute read and write requests to all of a RAID array's virtual disks in the presence of any two concurrent disk failures. Several methods, including dual check data computations (parity and Reed Solomon), orthogonal dual parity check data and diagonal parity have been used to implement RAID Level 6.

DataDirect Networks RAID 6 implementation uses a Reed-Solomon combined with a custom FPGA to protect against double-disk failures while delivering full write performance.

SATAssure: DataDirect Networks' SATAssure™ technology is an intelligent and robust SATA drive management technology and is a core element of the S2A DirectOS™ operating system, built natively into the S2A Appliance. SATAssure delivers enterprise-class data protection by making it possible to confidently deploy very large pools of SATA storage while maintaining data availability, reliability and full system performance.

S2A: The Silicon Storage Architecture (S2A) Appliance is an intelligent data management device designed by DataDirect Networks to deliver uncompromised levels of storage performance, reliability and quality of service.

References

Lustre Manual
http://manual.lustre.org/

DataDirect Networks Website:
http://www.datadirectnet.com

DataDirect Networks S2A 9550 Data Sheet
http://www.datadirectnet.com/pdfs/S2A9550_Brochure_110907.pdf

DataDirect Networks S2A 9900 Data Sheet
http://www.datadirectnet.com/pdfs/S2A9900Brochure110907.pdf

DataDirect Networks White Paper: Best Practices: Enterprise SATA Deployment with High Performance and Reliability
http://www.datadirectnet.com/index.php?option=com_content&task=view&id=298&Itemid=307

DataDirect Networks S2A9550: Capacity Optimized Speed and Reliability
Brian Garrett (with Claude Bouffard) - Enterprise Strategy Group Report: January 2008

Seagate Website- The evolution of S.M.A.R.T
http://www.seagate.com/support/kb/disc/smart.html

Wikipedia RAID Levels Explanation:
http://en.wikipedia.org/wiki/Standard_RAID_levels

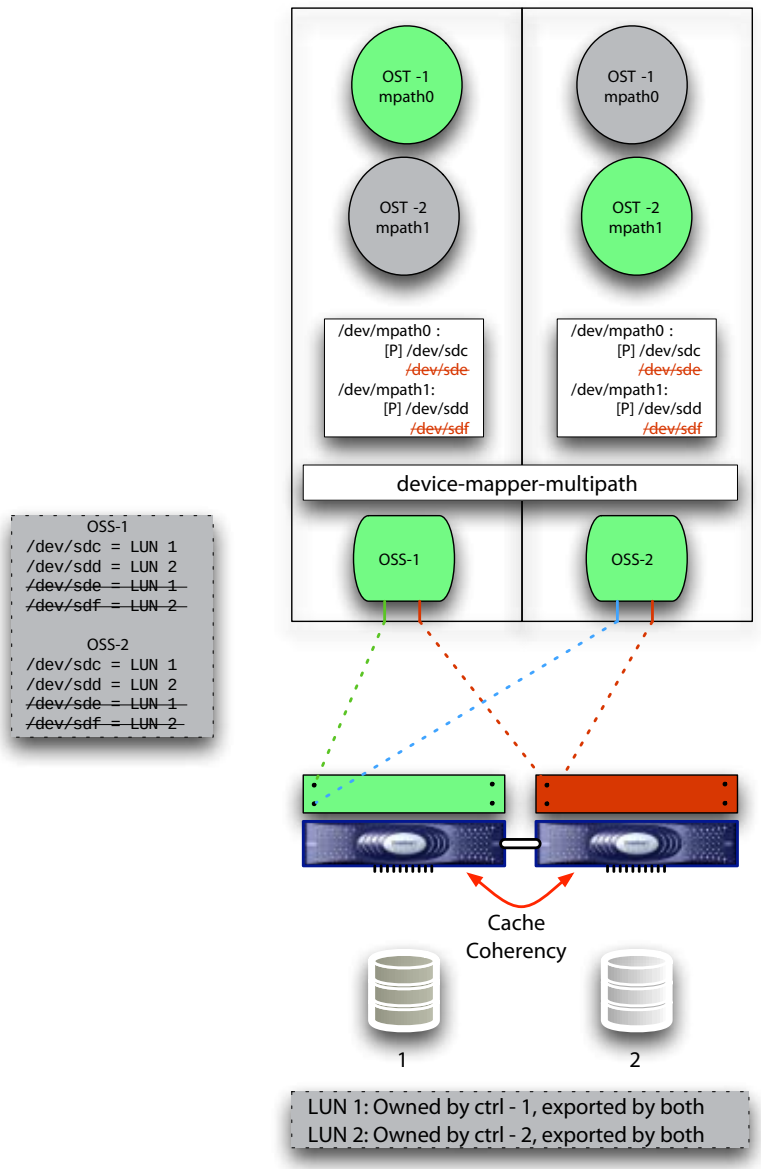
Wikipedia SMART Page:
http://en.wikipedia.org/wiki/Self-Monitoring%2C_Analysis%2C_and_Reporting_Technology

Lustre Center of Excellence Report: French Atomic Energy Commission (CEA)
clusterfs-intra.com/cfscom/ images/lustre/LUG2007/cealug2007.pdf

Indiana University Data Capacitor
http://datacapacitor.researchtechnologies.uits.iu.edu/

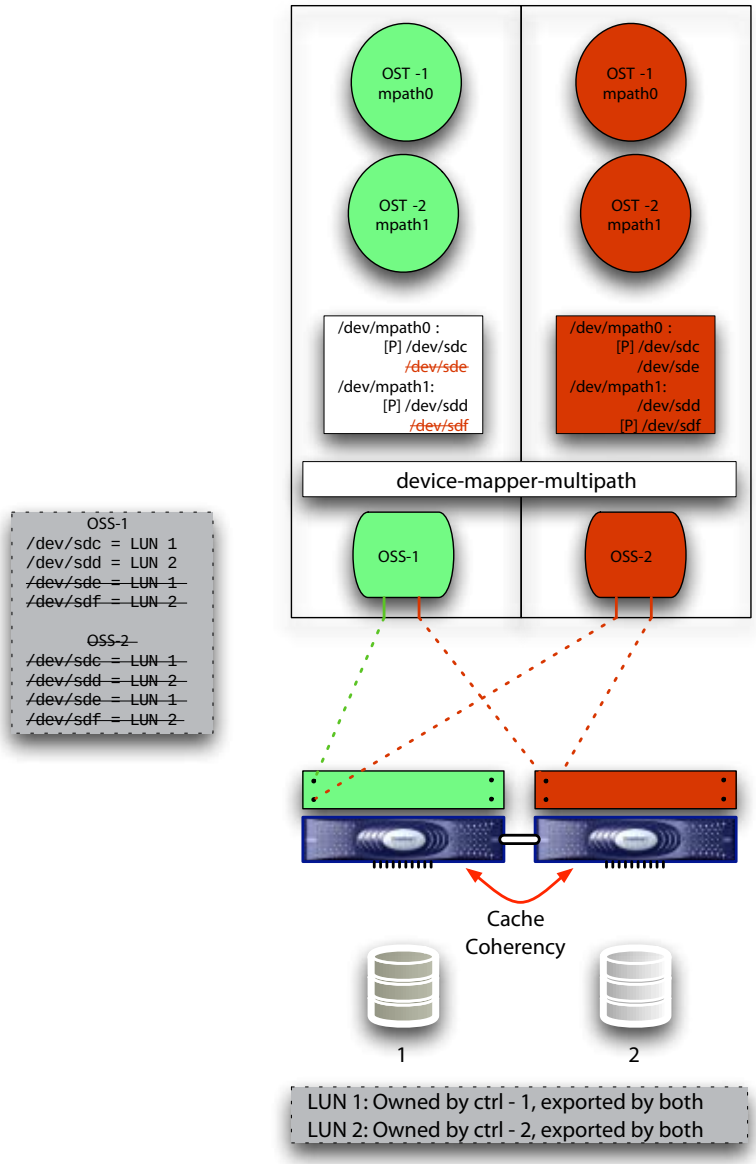
HPCWire Announcement: NCSA Performance Breakthrough:
www.hpcwire.com/hpcwire/hpcwireWWW/03/1119/106493

Appendix: Lustre Failover Configuration Diagram



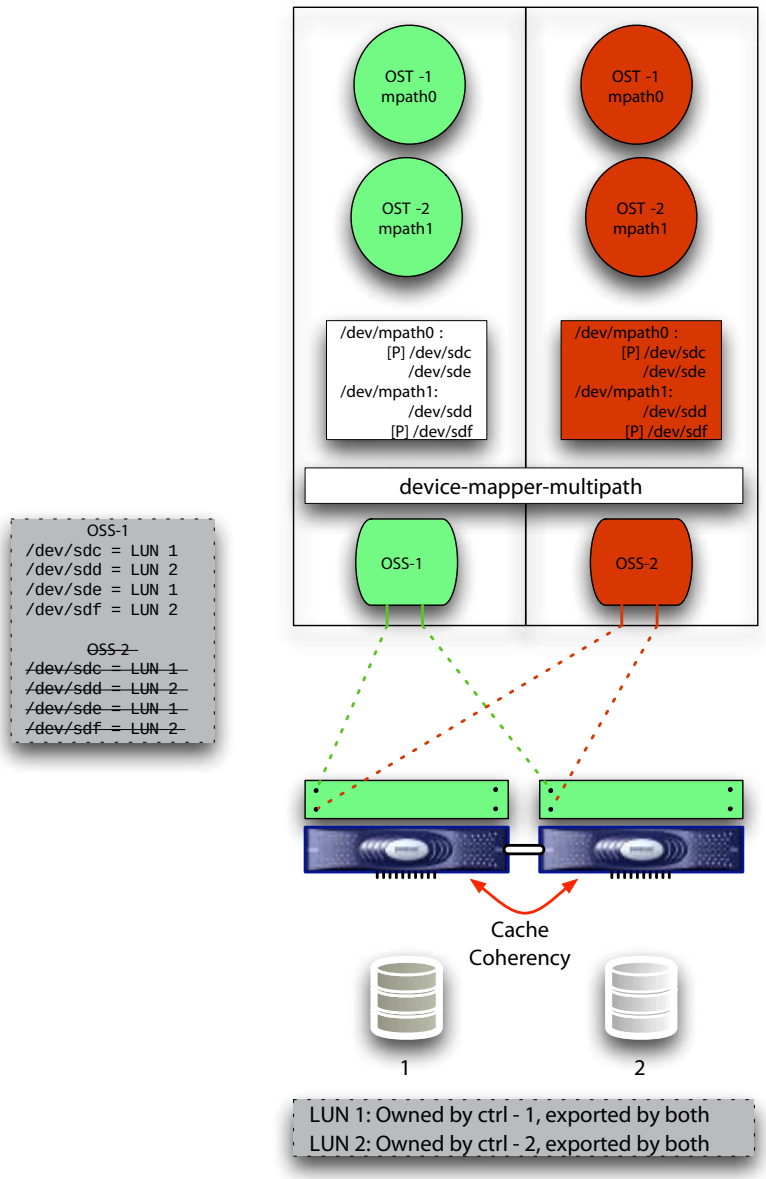
Appendix: Lustre Failover Configuration Diagram (con't)

Failed OSS Node and S2A Appliance



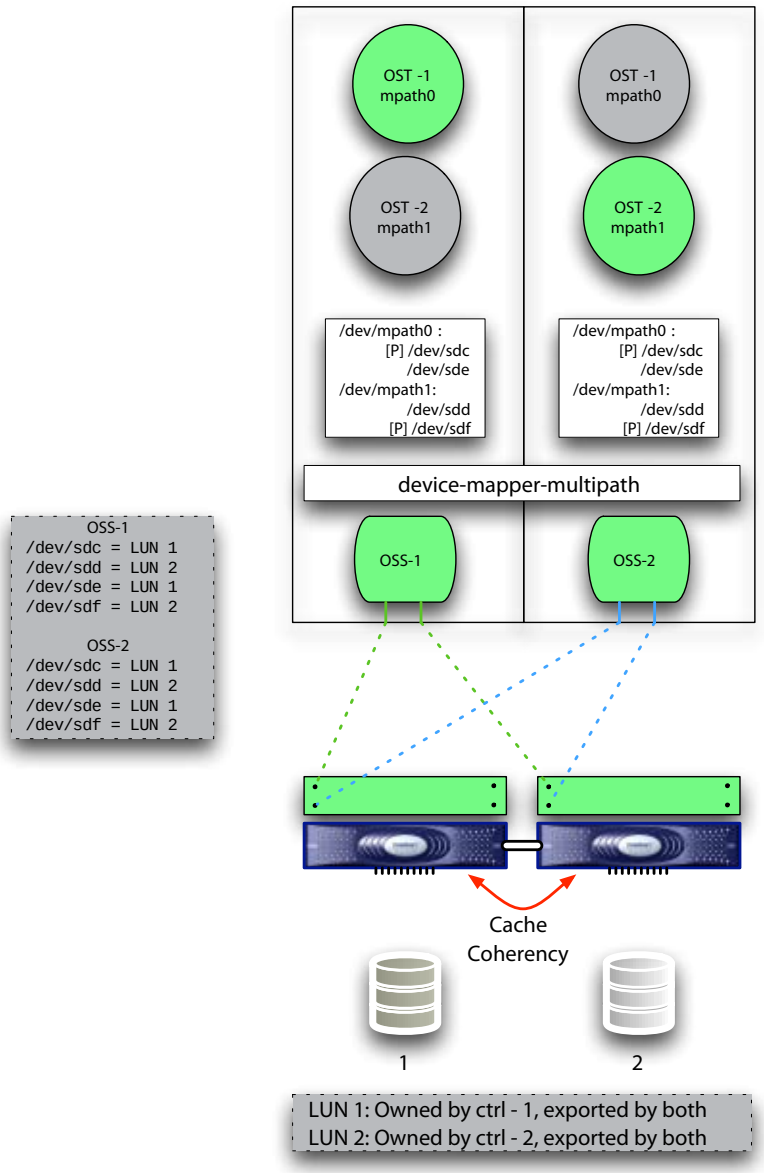
Appendix: Lustre Failover Configuration Diagram (con't)

Failed OSS Node and Functional Controllers



Appendix: Lustre Failover Configuration Diagram (con't)

All Functional Components



S2A TECHNOLOGY

DataDirect Networks is the leading provider of open, scalable storage systems for performance and capacity driven applications. DataDirect's S2A (Silicon Storage Architecture) appliance enables modern applications such as video streaming, content delivery, modeling and simulation, backup and archiving, cluster and supercomputing, and real-time collaborative workflows, that are driving the explosive demand for storage performance and capacity. DataDirect's S2A technology and solutions solve today's most challenging storage requirements, including providing shared, high-speed access to a common pool of data, minimizing data center footprints and storage costs for massive archives, reducing simulation computational times, and capturing and serving massive amounts of digital content.

Major corporations, supercomputing centers and rich media organizations, including AOL, Ascent Media, Boeing, CINECA, CGGVeritas, CNN, Disney, Federal Reserve Board, Ford, Hess, Kodak Gallery, Lawrence Livermore National Laboratories, NASA Ames, RIOT, Sandia National Laboratories, Sony, Technical University Dresden, Technicolor, Time Warner, Thomson, Trinity College Dublin and Universal, utilize DataDirect high performance, high capacity solutions.

DataDirect NETWORKS

Performance. Capacity. Innovation.

9351 Deering Avenue . Chatsworth . California 91311
phone +1.800.TERABYTE (837.2298) . fax +1.818.700.7601
sales@datadirectnet.com
www.datadirectnet.com