High Availability in Lustre®

John Fragalla
Principal Solutions Architect
High Performance Computing
## Goal for Lustre® HA

- Build resiliency and architect enterprise-class features for Lustre®
- Detect failures and provide architecture to deal with any level of failure
- Provide continued access to data with no to minimal I/O interrupts for applications
- Multiple redundant components is the basis for Lustre® Storage Availability
Software Resiliency and Reliability

• Software is based high availability to configure, manage, and monitor services

• Tolerate any type of failure to provide continuous data access

• Other benefits of High Availability
  • Live software upgrades
  • Live system maintenance
  • Live hardware upgrades/maintenance

• Data protection layer
  • No need to utilize dedicated RAID controllers
Data Protection Layer

• At minimum design solution around RAID 6 or write data based on 8+2 (8 data blocks and 2 parity blocks)
• Need to plan for rare cases of multiple drive failures within a single storage enclosure
  • Use of Hot Spares
  • Balance Front-end OST Performance with back-end disk rebuilds
• Adjusting rebuilds on the fly can protect against rare cases
Individual HA Domains

- Two Lustre Servers in a HA pair having access to all the OSTs
  - Reliable redundant Heartbeats
    - Two different networks or the use of passive back-plane within a storage enclosure and external network
  - Dual-Path Drives and SAS Paths
  - Separate Fabric Switches for Server in the Pair
HA Event Detection

- HA services detect and failover resources when loss of data from clients occur
  - Fabric Failure
  - SAS Interruption to Disks
  - Heartbeat Compromised
  - Software Interrupt
- HW Failure of any kind
- Rules and timeout values are critical in HA services to ensure the discovery of a failure is handled within short period of time
HA Failover

• When a partner detects an issue with a problem node, want to avoid dual-mount, split brain, or an event causing two nodes to fail in a HA Domain

• Defining the rules is critical and the use of STONITH

• In an event a healthy nodes questions it’s partner, Shoot The Other Node In The Head to avoid the problem node causing issues to healthy partner

• When partner node powers off unhealthy node, resources will failover automatically

• On failure, once Lustre Targets are mounted, resources can still be unavailable to client due to recovery time
  • Depends on number of Clients, and amount of data to replay when Failover occurred

xyratex
Failback

• When a HA event occurs, Failback should be manual
  • Automatic failback can be problematic if the node is still unhealthy and can cause more problems
• Failback is done manually for various reasons
  • Admins can do maintenance to ensure the problem causing the failover/STONITH is fixed
• Avoid further outage due to Lustre Recovery Time on failback
• Depending on architecture, suggest having module solution so on failure it will effect performance on a small percentage of overall solution
Fabric Connectivity

• If connecting all storage to a single Leaf Switch or Core Switch, this is the SPOF
• Ensure the solution can survive a complete fabric switch failure in addition to
  • HCA Failure
  • HCA Driver issue
• Cable Issue
• If using Leaf-Core module, multiple fabric links from Leaf to each core is highly recommended to provide redundant cable connectivity
UPS and Power

- UPS is critical to protect against unexpected power glitches or interrupts to power.
- Without UPS, any power interruption can compromise data integrity or disk drive MTBF.
- UPS can provide enough power to survive short power interrupts or allow for graceful shutdown of Lustre.
- Multiple Power grids to feed the redundant rack PDUs provide redundancy if one power feed is compromised.
Integration and Testing in Factory

• Hardware integration
  • Integrating the storage solution designed from the ground up into a single chassis proves to be highly available compared to non-integrated solutions

• Building, configuring and testing the entire storage solution in factory improves increased reliability and reduces on-site integration

• Repeatable performance on benchmarks for proven throughput
Disk Drive Testing and Integration

- **First Phase**
  - Rigorous disk drive testing before integrating the solution into a chassis

- **Second Phase**
  - Disk drive testing integrated into a chassis
  - Running various low level testing for connectivity
  - Vibration testing

- **Third Phase**
  - Disk drive testing with the software installed
  - Rigorous test cases
Integration and Solution Testing in Factory

• Hardware integration
  • Integrating the storage solution designed from the ground up into a single chassis proves to be highly available compared to non-integrated solutions

• Building, configuring and testing the entire storage solution in factory improves reliability and reduces on-site integration

• Repeatable performance benchmarks for proven throughput
Integrated System Testing (IST) is a patented 3-Stage testing process embedded within manufacturing and designed to remove hidden quality problems.

**Features**
- Optimized 36 Hour Manufacturing & Test
- Adaptable Test Automation
- Standard Across the Globe

**Benefits**
- Reduces solution warranty and service costs
- Reduces Infant Mortality
- Up to 1.5X drive reliability improvement over 3 Yrs.
  - AFR Reduction to < 0.5%, regardless of disk supplier
  - 67% less disk drive failures in first 3 months
Full Factory Integration – Pre-cabled & pre-installed

xyratex
Benefits For ClusterStor Based Architectures

• Technology based on ClusterStor, e.g. Sonexion, is highly available integrated Lustre storage solution providing end to end system management without sacrificing performance, leveraging industry standard components

• Built and optimized for Lustre®

• High availability and redundancy is the basis of ClusterStor

• ClusterStor benefits from Xyratex strong roots in Disk Test equipment and traditional OEM solutions

xyratex
Thank You

John_Fragalla@xyratex.com