Large Scale, Real-Time HPC SAN Design

Robert Plaster
IT Architect
RPI Consulting Services
rplaster@rpi-cs.com

Description

Real-time, high-bandwidth and mission critical systems present special design challenges. As they are mission critical they must perform their mission even if multiple SAN components fail. Because they are large-scale, duplicating stripe groups, which provide needed bandwidth to achieve high availability is prohibitively expensive. This talk discusses field-tested tools and techniques for designing SANs for use in such extreme applications.

Define Topology

- Big Bucket
  - Expensive approach as Real-Time requirements apply to whole system.
  - One Anomaly can take out entire system.
- Workflow Pattern (Saw tooth)
  - More Economical.
  - Decouples Real-Time as soon as feasible.
  - More Robust and fault-tolerant as a system.

Define File Workflow

- Ingest
- Sorting (File Type and Size) aka Pre-Processing
- Processing/Encoding
- Delivery/Staging

Define Topology for Workflow Stations

- Ingest
- Sorting/Pre-Processing
- Processing/Encoding
- Delivery/Staging

- 2N
- N+1
- N+1
- NN

Note: Apps and Drivers support are necessary to achieve hardware PRAS (Performance, Reliability, Availability, Supportability)

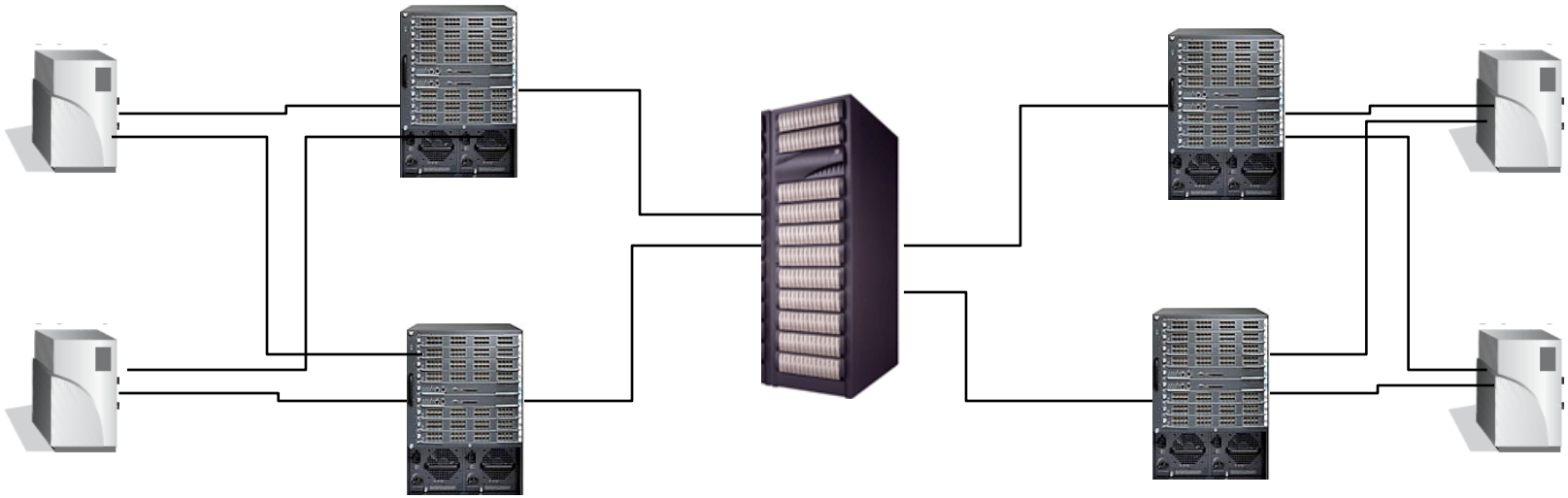
RPI
Consulting Services, Inc

Reduce Complexity and component count.

- Create an atomic structure for our disk and networking components.
- Benchmark that structure to understand it's limitations
- Use that structure as our scaling mechanism.

Avoid ISL (Inter-Switch Links)

- ISL's can kill performance, adds latency and complexity.
- Leveraged Storage Connections to bridge switches and expand connectivity



RPI
Consulting Services, Inc

Standard SAN Fail-Over Concerns

- How to handle path failures?
  - Most vendors limit path availability to two.
  - Shared file systems = shared failure
  - Symmetric RAIDs
    - Reduced performance as load balance is compromised unless we account for failure and increased channel demand.
    - Increased Cost to have excess performance available.
  - Asymmetric RAIDs
    - 50% decreased performance as load is on non-preferred channel.
    - Middleware in the mix to avoid channel ping-pong
    - Potentially crippling ping-ponging if no middleware directs all devices to other channel.

RPI
Consulting Services, Inc

Large Scale HPC SAN Fail-Over Concerns

- What will each device failure do to performance and availability?
- How do we isolate path failures to keep data running at RATE for non-affected hosts?
- How do we reduce the number of software components in the stack?
- How do I know that the system is optimal and healthy?
- How can we ring out the architecture and ensure that the system is built properly without taking days and weeks of testing and using super skilled SAN trained technicians?
- How do I integrate SAN sub-systems from multiple sub-contractors and avoid issues caused by dissimilar design approaches and configuration techniques?
- Host x just spit out a SCSI error. Is it local or global?

Fail-over redefined

- Worked with HBA vendor to have N+i paths.
- Created tools and processes to define SAN paths and give system a known performance for each failure scenario.
- HBA driver accepts pre-defined path definitions from configuration file loaded at boot.
  - Configuration file is created by tool.
- No middleware needed to direct RAID array as pre-defined failure definitions ensures single point failure resolution is optimal solution.
- Able to isolate HBA failures to a host to avoid channel saturation and hot spots.

Automated Configuration and Testing

- Created tools to automate configuration steps
  - Eliminated human induced configuration errors
- Automated Failure Tests to drive configuration and verify it is working as designed.
  - Proof system would react as designed when failures occur
- Error checking
  - When integration occurs, tools check that each subcontractor is compliant with design rationale and rule sets
  - Allows for a Configuration Management positive feedback.
- Created query tools that verified cables were plugged in where the drawings said they would be.
- Tools completed verification on small hosts 4-8 HBA's (8-15 paths) in 3 minutes. Fully populated large platform with over 300 paths completed in 18 minutes.

SAN Name Space

- Large Shared SAN needs human readable names to speed troubleshooting and maintenance
- Used SCSI addresses as unique RAID device names.
  - Every host binds a unique Target wwpn to the same Target ID.
- Used Volume Labels as sub-identifiers
  - OS limited character lengths but still valuable info for isolating devices on a large SAN.
- Configuration Tools enforced naming conventions

Test the Stack

- Created a regression test for hardware and driver components.
- Combined all the components to create a stack and tested them at rate.
- Any new version triggers a retest of the entire integrated stack.

Maintenance Issues

- File System Fragmentation
  - HSM vs Disk Only
- Backup an HSM?
  - Just the Metadata
- Catastrophic Device Failure
  - Have a plan
- Firmware Induced changes
  - Firmware changes LUN Parametric (FS may barf)
- Device End of Life replacement
  - HDA,Tape Drives
    - Larger HDAs change performance norms for LUN
  - RAID Array (new wwpn?)