PARTNER TECHNICAL BRIEF

Storage Designed for High Availability

Lenovo® Storage S3200 & SANsymphony™-V

Designing a data storage infrastructure for high availability requires the selection of the most reliable hardware components, coupled with intelligent storage services software to ensure continuous access to critical business data in the face of equipment and facility outages, as well as ongoing expansion and upgrades.

This paper outlines the design techniques employed to achieve enterprise-class high availability by combining the robust hardware and software in the Lenovo Storage S3200 storage array with synchronous mirroring and remote replication from the DataCore SANsymphony-V Software-defined Storage platform.

DataCore Ready Certified

The combination of SANsymphony-V and Lenovo Storage S3200 storage array, certified under the rigorous DataCore Ready Program, offers a rock solid storage foundation to meet the most stringent uptime requirements and disaster recovery objectives. Yet the solution is attractively priced to keep within budget constraints.

Keeping Users and Applications Running Undisturbed

Highly reliable storage hardware is a vital foundation for business continuity, but many IT organizations fail to account for the multitude of other factors that greatly contribute to planned and unplanned downtime. DataCore prevents these more frequent sources of storage-related disruptions from ever affecting applications, ensuring enterprise-class high availability in a cost-effective manner by leveraging redundancy across devices and sites.

The DataCore Ready Program Value Proposition

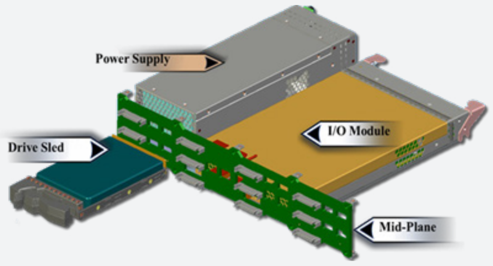
DataCore Ready identifies solutions trusted to strengthen SANsymphony-V- based infrastructures. While DataCore solutions interoperate with popular open and industry-standard products, the DataCore Ready designation ensures that these solutions have successfully executed a functional test plan and additional verification testing to meet a superior level of joint solution compatibility.

SOLUTION HIGHLIGHTS

- Synchronous mirroring for enterprise-class, zero downtime, zero touch high-availability (HA)
- Asynchronous replication to remote sites for disaster recovery (DR)
- Eliminates single points of failure

LENOVO STORAGE S3200 HIGHLIGHTS

- DataCore Ready certified
- Designed for high availability: demonstrated 99.999% availability
- NEBS & MIL-STD-810G compliant
- Works with VMware vSphere, Microsoft Hyper-V and other server operating systems and hypervisors



Lenovo's mechanical design enables the power supply and fan, I/O module and controller, and disk drives all to be serviced quickly as hot-swappable Field Replaceable Units (FRUs). Being able to replace redundant FRUs while the system is fully operational further enhances availability.

Customers who leverage DataCore Ready offerings benefit from quality assurance, reduced risk and lower integration costs. The DataCore Ready logo helps customers quickly identify products and solutions that are optimized for SANsymphony-V.

Designing for High Availability

High availability for the Lenovo storage arrays is achieved through a combination of three design elements:

- High reliability (measured by the Mean Time Between Failures or MTBF) of the complete system
- Redundant subsystems to eliminate as many single points of failure as possible
- Rapid repair of any failure (measured by Mean Time to Repair or MTTR)

The following equation for availability demonstrates the vital role of serviceability in the system's design. Maximum availability can be achieved only by minimizing the time it takes to affect a repair, which is reduced significantly by using FRUs.

$$\text{Availability} = \frac{\text{MTBF}}{\text{MTBF} + \text{MTTR}}$$

To achieve maximum availability, Lenovo designs for reliability and serviceability, as well as for manufacturability.

Design for Reliability & Serviceability (DFRS)

Designing hardware for high reliability and serviceability involves both the system and its subsystems. To achieve high availability at the system level, Lenovo integrates reliability into the design process in several ways. The first and most obvious is the use of disk drive redundancy with RAID configurations and dual power supplies, each including its own fan to prevent over-heating (and thereby, accelerated component failures).

Even higher availability is achieved by using redundant controllers. By eliminating single points of failure in these critical subsystems, the system itself continues to operate normally during a failure of any single FRU. While such a failure does factor into the subsystem's MTBF (its rated reliability), it does not diminish the availability of the system. Lenovo Storage S3200 architecture features full redundancy for every subsystem requiring a significant number of active components. The mechanical chassis itself cannot be redundant, of course, and there is a single mid-plane that performs the simple function of connecting the redundant controllers to the redundant disk drives. The mid-plane has minimal active components, however, and Lenovo selects these for the highest possible reliability. The result is an extraordinarily high MTBF for the chassis and its mid-plane, and therefore, virtually no impact on system availability.

Modular FRU Design with Rapid Fault Notification

To enhance system serviceability for the shortest possible MTTR, Lenovo utilizes several complementary design techniques. The first is the use of a modular chassis with FRUs. The ability to swap out a confirmed failed subsystem quickly and easily minimizes the time it takes to repair an installed system and restore it to full operation. By utilizing such a modular design, which provides convenient access to all subsystems, Lenovo Storage S3200 can be maintained seamlessly with minimal or no disruption in service during most repairs.

The second serviceability technique is immediate notification of any failure. The longer it takes to detect a failure, the longer it will take to repair. Time is of the essence for another reason, however: The failure of a redundant subsystem creates a temporary single point of failure that increases the risk of a system-level outage. For this reason, the firmware in all Lenovo systems is designed to detect, isolate and confirm any failure, initiate a fail over to a redundant subsystem, and provide immediate notification. The actual "messaging" of the notification can also be configured to match operational procedures to ensure that on-duty staff is properly and quickly notified.

Designing for Maximum MTBF

At the FRU or subsystem level, Lenovo utilizes four separate design techniques to maximize the MTBF of each, while at the same time also maximizing the inclusion of leading-edge SAN features.

The first is reducing the part count. Because any individual part can fail, the fewer there are, the higher the inherent reliability of the subsystem. Lenovo's engineers endeavor, therefore, to minimize the parts required on all printed circuit boards and other subsystem FRUs.

High Quality Component for Lower TCO

The second technique is to use only high quality parts. Higher quality parts cost more, of course, but their superior performance and longer service lives normally contribute to a lower total cost of ownership in the long-run. Despite the higher per-part cost, minimizing the part count, while concurrently enhancing functionality, helps to improve the overall price/performance of a highly-reliable design. For these reasons, Lenovo utilizes only the highest quality parts available from reputable suppliers.

Increased Operating Margins

The third technique involves the de-rating of selected parts. Operating any part or component at or near its rated capacities inevitably shortens its useful service life. For critical parts, Lenovo selects only those that will be able to operate at approximately 50% of their maximum allowable specifications for voltage, power and/or current. This can substantially increase the service life, and therefore, the MTBF of the subsystem.

Ensuring Software Maturity

The fourth technique is unique to Lenovo: designing for software reliability. In modern designs, software reliability is just as important as hardware reliability, and in some ways even more important. The reason is: Software bugs (including those in firmware) that cause downtime normally take significantly longer to resolve than the more obvious hardware failures. Bugs are often dependent upon system state (the set of circumstances leading up to the failure), making them difficult to reproduce and isolate quickly, and any patch or update must be tested before it can be released. Both add considerably to the MTTR for software failures, thereby adversely impacting on system serviceability and availability.

Ensuring Software Maturity

To maximize software reliability, Lenovo monitors the improvement in the Mean Time to Discovery (MTTD) of bugs to assess the maturity of all software and firmware during development. It is important to note that MTTD is not an industry metric, but is instead a software maturity metric created by Lenovo as part of the company's commitment to quality and reliability. All designs must have a sufficiently high and stable MTTD before being finalized.

Stringent Design Verification

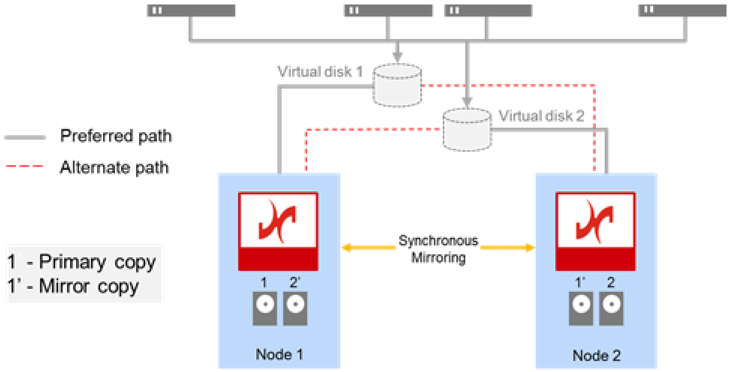
The design is released to manufacturing only after passing three comprehensive tests. The Engineering Verification Test (EVT) and the Design Verification Test (DVT) ensure that the system and/or subsystem(s) fully satisfy all design specifications, including those for high reliability of both the hardware and software. These tests also confirm that marginal variations in parts from component suppliers will not compromise system reliability over the product's useful life of a minimum of 10 years. The Reliability Demonstration Test (RDT) is a separate and rigorous evaluation of the final production hardware that verifies its calculated reliability, availability and serviceability. Whereas some vendors use only a few samples in a fairly short demonstration test, Lenovo's RDT uses 18 to 20 fully-configured systems in a 13-week test that must produce zero hardware failures to pass.

DataCore Synchronous Mirroring

Having looked in detail at the techniques used to ensure hardware reliability, we turn now to consider how an enterprise-class high-availability architecture is achieved in combination with DataCore SANsymphony-V Software-defined Storage platform.

The DataCore software synchronously mirrors I/Os between separate Lenovo storage arrays to eliminate single points of failure. The benefits include:

- Zero downtime, zero-touch, enterprise-class high availability
- Mirrors data synchronously at high speeds between physically separate locations
- No manual intervention or scripting needed for failover, re-synchronization and failback
- Prevents equipment and site outages, both planned or unplanned, from disrupting business operations
- Uninterrupted data access from application and/or server clusters even during the loss of an entire site



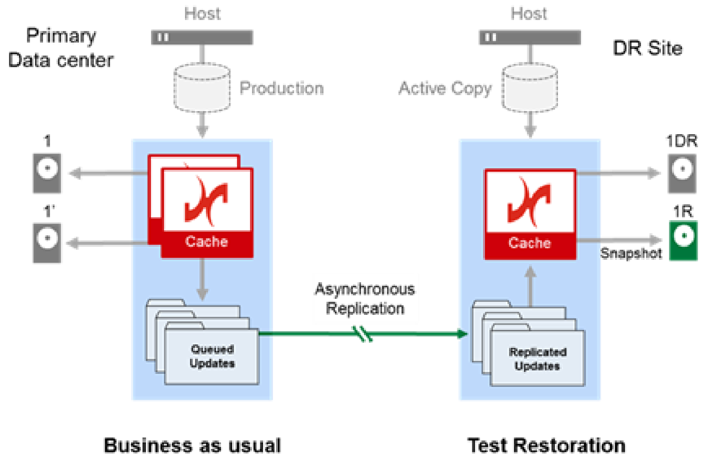
- Active-active mirrored copies at geographically separate sites appear as shared disks on redundant paths to local and metro clusters
- The simplicity of fully autonomous nodes operating as a redundant N+1 grid, without the complexity, limitations and extra cost of clustered storage devices

SANsymphony-V creates redundant storage pools by synchronously mirroring between DataCore nodes. For any mirrored virtual disk, one DataCore node owns the primary copy and another holds the secondary copy. Those are maintained in lock step. In the diagram above, Node 1 owns the primary mirror copy labeled 1 and Node 2 holds the secondary copy labeled 1'. The preferred path from the host computer to the virtual disk is generally assigned to the node that holds the primary copy of the mirrored set.

Under normal operation, all read and write requests issued to that virtual disk will be serviced by the primary copy. The secondary copy need only keep up with new updates arriving from the mirroring function. Generally, nodes are configured to control primary copies for some virtual disks and secondary for others, thereby evenly balancing their read workloads. Alternatively, N+1 configurations consisting of 3 or more nodes, may rely on a common node to keep the mirrored backup copy. Should any errors be encountered on the preferred path, the host's multipath drivers automatically fail over to the alternate path without disrupting applications. The same is true if a node needs to be taken out-of-service for maintenance or upgrades. If the node encounters any problems trying to reach the disks where the primary copy is stored, it will redirect the request to the node holding the mirrored copy.

From a physical standpoint, best practices call for the DataCore nodes to be maintained in separate chassis at different locations with their respective portion of the disk pool so that each can benefit from separate power, cooling and uninterruptible power supplies (UPS). The physical separation reduces the possibility that a single mishap or facility problem will affect both members of the mirrored set.

Round-trip network latencies govern the maximum distance between mirrored nodes. Current technologies support inter-node distances up to 100 kilometers. The actual limits depend on application delay sensitivity and the network latency experienced between locations.



Asynchronous Remote Replication

DataCore's remote replication function addresses requirements for secondary copies to be housed beyond the reach of synchronous mirroring, as in distant disaster recovery sites, branch offices and satellite facilities. It relies on a basic IP connection between locations, and works in both directions. That is, each site can act as the disaster recovery spot for the other.

The software operates asynchronously, meaning that it does not hold up the application waiting on confirmation from the remote end that the update has been stored in both places.

Instead, it offers to do its best to keep up to date with changes at the local site, but makes no guarantees. It's far better than trying to constantly make backup tapes and ship them to a safe house or paying extra for point-products to handle only this task. The advanced protocol handles prolonged transmission delays or link outages allowing you to set the priority of which virtual disks should be allocated the most bandwidth.

You can quickly get a remote site initialized by cloning the primary site's remote disks onto transportable media and shipping them to the disaster recovery center. Then the software applies the changes that transpired while in transit. To help you build strong, verifiable disaster recovery practices that you can confide in, SANsymphony-V enables you to test restoration at the remote site while production updates continue to arrive. Any changes made during the simulated recovery are then discarded and the standby copies refreshed with the newest updates.

About Lenovo Storage Systems - For further information about Lenovo please visit www.lenovo.com/storage.

0515

For additional information, please visit **www.datacore.com** or email **info@datacore.com**

© 2015 DataCore Software Corporation. All Rights Reserved. DataCore, the DataCore logo and SANsymphony are trademarks or registered trademarks of DataCore Software Corporation. All other products, services and company names mentioned herein may be trademarks of their respective owners.