Global commerce and the demand for 24/7 access have made information availability vital to enterprise success. A lack of data access during maintenance, hardware or software failure, or any other period of downtime can lead to decreased productivity, customer dissatisfaction, and lost revenue. As high-availability technology continues to advance, both the cost and complexity associated with clustering for availability and failover have been reduced significantly, making these configurations feasible for small businesses as well as large enterprise-class data centers.

For data to be continuously available, IT systems must maintain uninterrupted data access during both planned and unplanned downtime. Host-based clustering, in conjunction with high availability at the storage level, can greatly enhance IT service levels, and represents a crucial step toward ensuring data availability and minimizing downtime problems. Allowing multiple servers (for example, a two-node cluster) to share access to data on a single storage system helps provide application availability even if one of the servers becomes unavailable.

The Dell PowerEdge Cluster SE600W solution and Dell PowerVault MD3000 storage array are designed to allow enterprises of all sizes to build highly available systems in a cost-effective way. Deploying this solution enables enterprises to achieve highly available service levels in a cost-effective way.

**Dell PowerEdge Cluster SE600W**

Dell PowerEdge clusters are based on Microsoft Cluster Service software and designed to keep applications and services available during any single failure within the cluster. When Microsoft Cluster Service detects a failure, it automatically moves cluster resources from the failed cluster node to a healthy one and restarts the applications. PowerEdge clusters support cluster-aware applications such as Microsoft SQL Server™, Microsoft Exchange Server, and Oracle® Database with Oracle Fail Safe software.

In addition to application-level availability, enterprises should consider redundancy in the server-to-storage I/O path. Because a failure of any component along this path (such as a server, adapter, controller, cable, or disk drive) jeopardizes system availability, the storage system’s ability to maintain data access during such a failure is a key part of a highly available storage deployment.

The PowerEdge Cluster SE600W is the latest addition to the Dell family of feature-rich clustering solutions, and the first based on Serial Attached SCSI (SAS) technology. It currently supports two-node clustering based on the components shown in Figure 1. The supported PowerEdge servers provide key high-availability features such as error-correcting code memory; software or hardware RAID for the internal drives; hot-swappable drives, power supplies, fans, and PCI slots; optional dual host bus adapters (HBAs); and optional redundant paths to the storage systems.
---|---
Servers | Dell PowerEdge models 1800, 1850, 2800, 2850, 6800, 6850, 1950, 2900, 2950, 2970, and 6950
Storage | Dell PowerVault MD3000

**Figure 1. Supported components for the Dell PowerEdge Cluster SE600W**

The PowerEdge Cluster SE600W offers three configurations:

- **Entry-level configuration:** This configuration offers only one path from each server to the storage (see Figure 2). If a component in one path fails, the server loses access to the storage; when this happens, Microsoft Cluster Service moves resources from the affected server to the other server and restarts the applications.

- **Redundant configuration with dual-port HBAs:** This configuration provides higher availability than the entry-level configuration, incorporating two ports on each SAS 5/E HBA to create two paths from each server to the storage (see Figure 3). If an HBA fails, Microsoft Cluster Service handles application migration the same way it does in the entry-level configuration. If another storage component fails, I/Os are transparently rerouted to the alternate path without affecting applications or clients.

- **Fully redundant configuration:** This configuration provides the highest availability of the three configurations, incorporating two SAS 5/E HBAs and two paths from each server to the storage (see Figure 4). Administrators can configure I/Os to use any path for load balancing. If a storage path fails, I/Os are transparently rerouted to the alternate path without affecting applications or clients.

**Dell PowerVault MD3000**

The Dell PowerVault MD3000 storage array, with two high-performance, active/active external RAID controllers and a mirrored cache, can be a key part of designing a cost-effective, highly available cluster system. It is based on SAS technology and offers high-availability features such as hot spares; hot-swappable drives, power supplies, cooling fans, and storage controllers; redundant storage controllers; redundant back-end paths with dual-port SAS drives; and optional premium features such as snapshots and virtual disk copy.

To appreciate the potential value of clustering with PowerVault MD3000 storage, enterprises should first understand the *host-based RAID* method of clustering. This method uses a direct connection from a storage device to each server, and the RAID controller itself resides on a PCI RAID card installed in the server. In an *external RAID* configuration, like that of the PowerVault MD3000, the RAID controller resides outside the server in the external storage system itself.

Among the issues associated with host-based RAID clustering, one of the most significant is related to the cache and performance. The cache is high-speed memory designed to reduce the time required to read or write data. In host-based RAID configurations, the cache resides on the RAID adapter card located in the server. To help ensure that in the event of a server failure the unwritten cache data is written to disk, vendors disable the use of the cache on RAID adapters to maintain cache coherency for the cluster. Unfortunately, doing...
so can cause a significant degradation in cluster performance.

When using external RAID, the cache resides on the external controller within the storage system, and cache coherency can be maintained without disabling the cache functionality. For this reason, external systems are preferable for a clustered configuration. Storing the cache separately from the server helps eliminate the risks associated with a server failure in a host-based RAID cluster. As an external storage system, the PowerVault MD3000 is ideal for highly available two-node clustering applications on PowerEdge servers when performance and access to data are critical.

**Cluster features and advantages**

The PowerEdge Cluster SE600W and PowerVault MD3000 are designed to build upon the reliability of parallel SCSI while addressing that technology’s performance, reliability, and scalability limitations. Figure 5 summarizes the cluster features of the SAS-based PowerEdge Cluster SE600W compared with the parallel SCSI–based PowerEdge Cluster SE500W.

**Performance**

The PowerVault MD3000 can provide better performance than PowerVault 22xS storage because of its increased data transfer rate over the SAS link and its enabled write cache. While the maximum data transfer rate for Ultra320 SCSI is 320 Mbps (and the complications of the shared bus architecture make increased speeds unlikely), SAS currently offers speeds of up to 3 Gbps, and the SCSI Trade Association plans for it to eventually support up to 12 Gbps.

In a PowerEdge Cluster SE500W, the write cache is disabled in the RAID controller to maintain cache consistency, because the RAID intelligence is part of the controller residing on the host. In a PowerEdge Cluster SE600W, the write cache is enabled in the RAID controller, because the RAID intelligence is part of the controller residing on the external storage device. Each storage controller has 512 MB of cache, which helps significantly improve overall system performance.

**Reliability**

When deployed in an environment using parallel SCSI and a shared bus architecture, a misbehaving device can bring down the entire system. The PowerVault MD3000 offers a point-to-point architecture in which such devices can be easily isolated without affecting the entire system. In addition, the fully redundant PowerEdge Cluster SE600W configuration provides two paths from the servers to the storage, and includes dual-port SAS drives to allow data access from either port from one of the two storage controllers. As a result, if a component in a storage path fails, the multipath software can automatically reroute the I/Os to the alternate path without affecting the applications or clients.

**Scalability**

Parallel SCSI can support up to 16 devices on the shared bus, while a SAS domain can support up to 16,384 devices without performance degradation. The PowerEdge Cluster SE600W can support up to 65 SAS hard drives by adding PowerVault MD1000 disk expansion enclosures to the PowerVault MD3000 using a daisy-chain topology (see Figure 6).
Highly available cluster solution

The Dell PowerEdge Cluster SE600W and PowerVault MD3000 take advantage of SAS technology to offer enhanced performance, reliability, and scalability, and include additional features not found in parallel SCSI–based PowerEdge Cluster SE500W configurations. This cost-effective clustering solution can benefit enterprises of all sizes that require high availability for critical applications such as databases, messaging systems, and file, print, and Web servers.

Nam Nguyen is a senior consultant in the High-Availability Cluster Development Group at Dell, and the lead engineer for SAS, Internet SCSI (iSCSI), and Fibre Channel Dell PowerEdge Cluster products. His current interests include business continuity, clustering, and storage technologies. He has a B.S. and an M.S. in Electrical Engineering from the University of Texas at Austin.

Daniel Moges is a systems engineer adviser in the High-Availability Cluster Development Group at Dell. His responsibilities include developing SCSI-, iSCSI-, and Fibre Channel–based high-availability clustering products; his current interests related to high-availability clustering include enterprise storage technologies and database systems. Daniel has a B.S. in Electrical Engineering from the University of Illinois at Urbana-Champaign.

Shobana M. is an engineering analyst in the High-Availability Cluster Development Group at Dell. Her interests include the development of SCSI- and Fibre Channel–based clustering solutions and application performance. She has a B.E. in Computer Science and Engineering.

Travis Vigil is a senior product marketing consultant for Dell PowerVault storage. He has a B.S. from Stanford University and an M.B.A. from Northwestern University’s Kellogg School of Management.