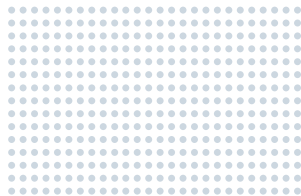


Deploying Dell PowerVault MD3000 Storage in Oracle Database 10g Cluster Environments

BY MAHMOUD AHMADIAN
CHETHAN KUMAR



The Dell™ PowerVault™ MD3000 storage array offers cost-effective enterprise storage that can scale to help meet future storage and application server requirements. This article discusses the benefits possible when deploying the PowerVault MD3000 along with Oracle® Database 10g with Oracle Real Application Clusters and Dell PowerEdge™ servers.



Using cost-effective storage with enterprise-class performance, manageability, and scalability for databases can be important to many enterprises, particularly small and medium-size ones. The Dell PowerVault MD3000 offers such a storage array: easy to configure and manage, the PowerVault MD3000 is expandable to provide room for growth and includes features such as redundant storage paths, path failover, load balancing, and clustering support. This article discusses the potential benefits of using PowerVault MD3000 storage with Oracle Database 10g Release 2 (R2) and Oracle Real Application Clusters (RAC), and compares this storage array's performance with typical Fibre Channel–based storage.

The PowerVault MD3000 has two RAID controllers with redundant power supplies and cooling fan modules, along with two input ports for host connections and one port for expansion. Other key features are designed to offer data protection, flexible storage partitioning, multiple management options, expansion scalability, and virtual disk snapshots and virtual disk copy.

Related Categories:

Application servers

Characterization

Dell PowerEdge servers

Dell PowerVault storage

Microsoft Windows Server 2003 x64 Editions

Oracle

Performance

RAID

Serial Attached SCSI (SAS)

Visit www.dell.com/powersolutions for the complete category index.

Understanding Dell PowerVault MD3000 features

The Dell PowerVault MD3000 uses Serial Attached SCSI (SAS) technology, which is well suited for critical applications and provides several features and benefits not available with ATA and SCSI. One example is the point-to-point topology of SAS, which enables dedicated and scalable throughput between devices. In addition, its full duplex communication of 300 MB/sec is almost twice that of Ultra320 SCSI because of the half-duplex nature of parallel SCSI. Because of their fast throughput, SAS disks can be used as external physical disks in mass storage devices requiring high I/O access. Serial connectivity also helps increase RAID volume creation and rebuild performance, and thin SAS cables help simplify cable and temperature management.

Data protection

The PowerVault MD3000 offers RAID support to help provide fault tolerance and protect data. It supports the following RAID levels:

- **RAID-0:** This level provides the fastest performance of the supported RAID levels, but no redundancy. It is typically used for noncritical data.
- **RAID-1 and RAID-10:** These levels provide fast performance and the highest data availability of the supported RAID levels. They are typically used for accounting, payroll, and financial applications. In the Dell Modular Disk Storage Manager software, RAID-10 is automatically used when four or more physical disks are selected.
- **RAID-5:** This level is typically used in multiuser environments with a high proportion of read activity and a small typical I/O size, such as file, database, Web, e-mail, news, and intranet servers.

The PowerVault MD3000 also enables administrators to configure a separate physical disk as a hot spare to provide additional fault tolerance. This disk does not contain any data, but acts as a backup in case a physical disk fails in a RAID-1, RAID-10, or RAID-5 virtual disk.

Flexible storage partitioning

The PowerVault MD3000 provides storage partitioning capabilities for virtual disk management. A storage partition is a logical entity consisting of one or more virtual disks that can be accessed by a single host or shared among hosts that are part of a host group. Partitions can be useful when specific hosts must access specific virtual disks in the storage array or when hosts with different operating systems are attached to the same storage array. In the latter case, administrators must create a separate storage partition for each host type.

Multiple management options

The PowerVault MD3000 offers both in-band and out-of-band management, which differ in the type of connection they use for sending commands and receiving event updates. With in-band management, commands, events, and data all travel through host-to-controller SAS interface cables. With out-of-band management, data is separated from commands and events: data travels through host-to-controller SAS interface cables, while commands and events travel through Ethernet cables. Dell best practices recommend using both types of management.

An Ethernet port on each RAID controller in the PowerVault MD3000 provides the out-of-band management interface. When using out-of-band management, administrators must set the network configuration for each RAID controller module, including its IP address, subnet mask, and gateway. If administrators are using a Dynamic Host Configuration Protocol (DHCP) server, they can enable automatic network configuration; if not, they must configure network settings manually.

Expansion scalability

Administrators can add up to two Dell PowerVault MD1000 SAS enclosures to the PowerVault MD3000, allowing a total of up to 45 physical disk drives. Using an average SAS disk size of 146 GB, this setup provides up to 6.6 TB of raw disk space (or up to 3.3 TB of fault-tolerant disk space in a RAID-10 configuration), which exceeds the typical requirements of small and medium-size businesses. RAID-10 virtual disks configured on the PowerVault MD3000 can survive double disk failures, adding another layer of protection against downtime.

Virtual disk snapshots and virtual disk copy

The PowerVault MD3000 provides virtual disk snapshots and virtual disk copy as add-on premium features.

Virtual disk snapshots. A virtual disk snapshot is a point-in-time image of a virtual disk in a storage array. It is not an actual virtual disk containing data; rather, it is a reference to the data that was contained

on a virtual disk at a specific time. A snapshot is the logical equivalent of a complete physical copy, but it can be created more quickly and uses less disk space than a physical copy.

The virtual disk on which the snapshot is based, called the source virtual disk, must be a standard virtual disk in the storage array. Typically, a snapshot is created so that an application can access the snapshot and read the data while the source virtual disk remains online and accessible (although no I/O requests are permitted on the source virtual disk while the snapshot is being created).

A virtual disk snapshot repository containing metadata and copy-on-write data is automatically created along with a snapshot. Because only data that has changed since the time of the snapshot is stored in this repository, it uses less disk space than a full physical copy. After the repository is created, I/O write requests to the source virtual disk can resume. Then, before a data block on the source virtual disk is modified, the contents of the block to be modified are copied to the repository for safekeeping. Because the repository stores copies of the original data in these data blocks, further changes to these data blocks write only to the source virtual disk.

Virtual disk copy. A virtual disk copy is a copy pair that creates a copy of a virtual disk, with the source and target virtual disks on the same storage array. The source virtual disk can be a standard virtual disk, a virtual disk snapshot, or the source virtual disk of a snapshot. The target virtual disk can be a standard virtual disk or the source virtual disk of a failed or disabled snapshot.

A source virtual disk accepts host I/O read activity and stores the data until it is copied to the target virtual disk. When a virtual disk copy is started, all data is copied to the target virtual disk, and the source virtual disk permissions are set to read-only until the virtual disk copy is complete. After the virtual disk copy is complete, the source virtual disk becomes available to host applications for write requests. To help prevent any data corruption, best practices recommend not accessing a source virtual disk that is participating in a virtual disk copy while the copy is in progress.

Virtual disk copy enables administrators to do the following:

- **Copy data to larger-capacity physical disks:** As the storage requirements for a virtual disk change, administrators can copy data to a virtual disk in a disk group in the same storage array that uses drives with larger capacity than the original drives.
- **Restore snapshot data to source virtual disks:** Administrators can restore the data from a virtual disk snapshot and then copy the restored data to the original source virtual disk.
- **Create backup copies:** Administrators can create a backup of a virtual disk by copying data from one virtual disk to another in the same storage array, minimizing the time that the source virtual disk is unavailable to host write activity. They can use the target virtual disk as a backup for the source virtual disk, as a resource for system testing, or as a means for copying data to another device, such as a tape drive.

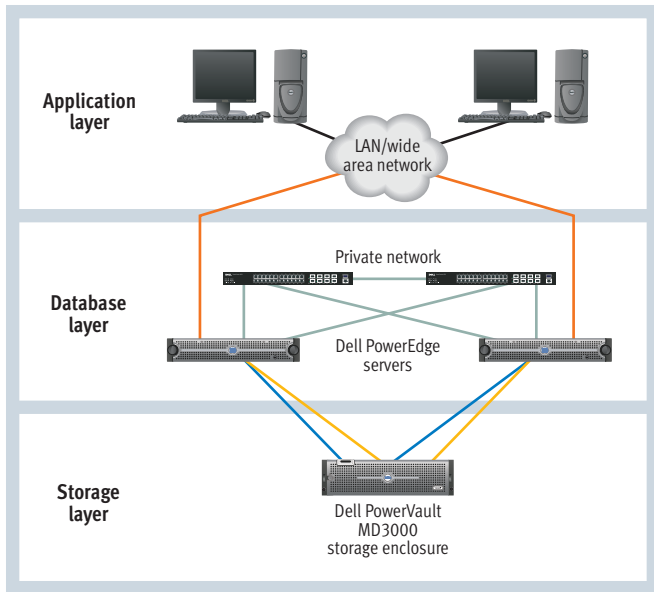


Figure 1. Database cluster configuration using Dell PowerVault MD3000 storage

- Recover from backup copies:** Administrators can use the Edit Host-to-Virtual Disk Mappings feature in Dell Modular Disk Storage Manager to help recover data from a backup virtual disk. The Mappings option enables un-mapping the source virtual disk from its host and then mapping the backup virtual disk to the same host.

Using Dell PowerVault MD3000 storage with Oracle Database 10g and Oracle Real Application Clusters

Oracle Database 10g Standard Edition is well suited for medium-size enterprise environments, and it includes Oracle RAC capabilities to help protect against hardware failures and provide the flexibility to scale up hardware resources. It is also easy to install and configure, and includes its own cluster and storage management capabilities. Because Standard Edition is built from the same code base as Enterprise Edition, it can easily be upgraded to Enterprise Edition if necessary, thus providing scalability and helping reduce total cost of ownership.

Enterprises using Oracle Database 10g with Oracle RAC can take advantage of the PowerVault MD3000 for database storage as part of a validated two-node Dell PowerEdge server cluster configuration (see Figure 1). Each node in this configuration can be a PowerEdge 1850, PowerEdge 2800, PowerEdge 2850, or PowerEdge 6850 server with dual-core Intel® Xeon® processors; a PowerEdge 1950, PowerEdge 2900, or PowerEdge 2950 server with dual- or quad-core Intel Xeon processors; or a PowerEdge 6950 server with up to four AMD Opteron™ processors per node. Both Oracle RAC nodes share the PowerVault MD3000 as common storage connected through dual SAS storage paths. The PowerVault MD3000 includes dual storage processors, and each node is connected to both processors. This configuration is supported for systems running

the Microsoft® Windows Server® 2003 Standard x64 Edition and Enterprise x64 Edition operating systems with Service Pack 2 (SP2).

High-availability clustering

In the validated Dell/Oracle configuration, the PowerVault MD3000 can connect to the cluster nodes using a single path through two SAS 5/E HBAs on each node. Using a dual-HBA configuration can help increase availability.

Both RAID controllers on the PowerVault MD3000 forward I/O requests to their respective virtual disks, and if one of the controllers fails, the requests are rerouted through the other RAID controller. If an HBA port on a cluster node fails, I/O paths fail over to the other HBA in the node, and virtual disks fail over to the RAID controller in the storage. However, the advantage of a dual-HBA configuration is that administrators can quickly change the physical connection from the failed port to the other functional port on the HBA—allowing the cluster nodes to continue using all four paths until the SAS controller is replaced. Multipath drivers such as Microsoft Multipath I/O installed on host systems that access the storage array provide I/O path failover capabilities. Thus, in case an HBA, I/O path, or RAID controller fails, multipath drivers can automatically fail over the virtual disks to other available paths or RAID controllers, helping ensure disk accessibility for the cluster nodes.

Each node runs a separate instance of the Oracle Database 10g Automatic Storage Management feature and Oracle database services, independently fulfilling client requests. Redundancy for the HBAs, network interface cards, and RAID controllers as well as software components help provide high availability. Multiple path management, load balancing on multiple paths, path failover software, and volume management are included on the Dell PowerVault MD3000 Resource CD.

Performance testing

In October 2006, Dell engineers set up a two-node Dell PowerEdge 2850 server cluster with PowerVault MD3000 external storage running Oracle Database 10g with Oracle RAC (see Figure 2). To evaluate cluster performance, they measured transactions per second under different loads by using Benchmark Factory from Quest, which provides an industry-standard TPC-C benchmark for testing online transaction processing (OLTP) databases. They then ran the same tests on this cluster configuration using 2 Gbps Fibre Channel–based storage and 4 Gbps Fibre Channel–based storage in place of the PowerVault MD3000.

Servers	Two Dell PowerEdge 2850 servers with Intel Xeon processors at 3.0 GHz
Memory	8 GB for each server
OS	Microsoft Windows Server 2003 R2 Standard x64 Edition
Software	Oracle Database 10g R2 (10.2.0.2) with Oracle RAC

Figure 2. Cluster test configuration

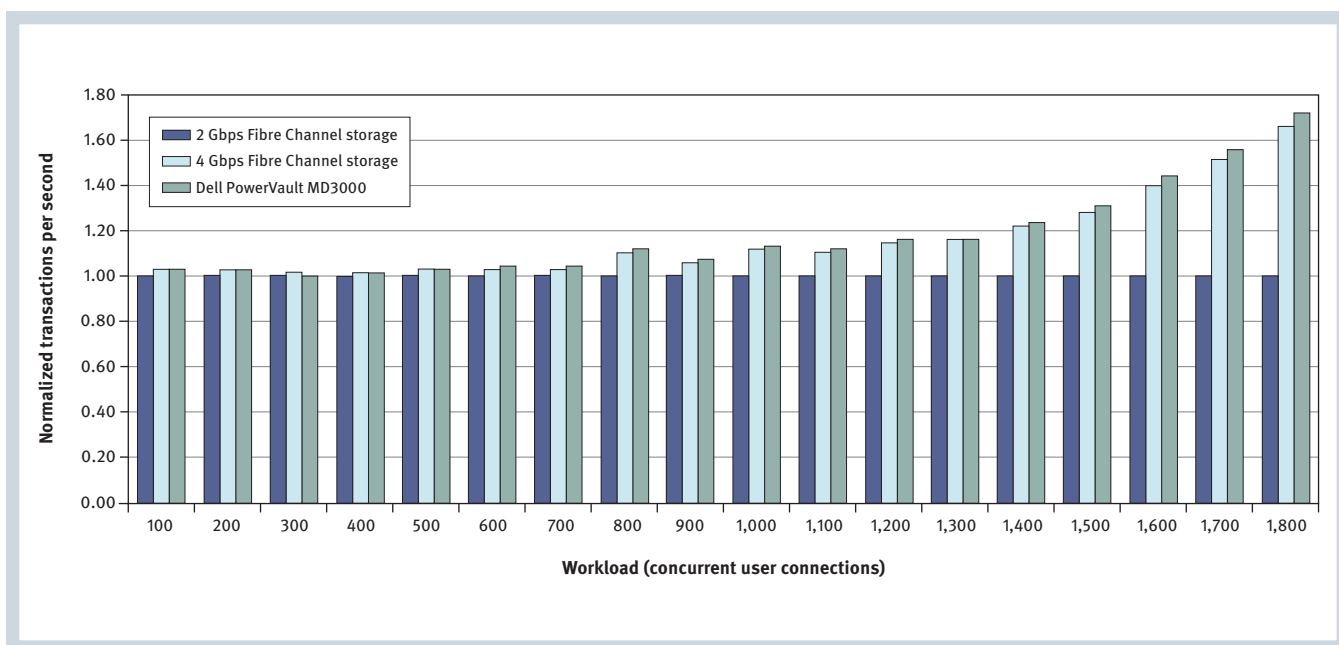


Figure 3. Performance comparison of Dell PowerVault MD3000 storage against Fibre Channel–based storage

Figure 3 shows the results, which have been normalized using the 2 Gbps Fibre Channel–based storage as a baseline for comparison. The results demonstrate that the PowerVault MD3000 performed similarly to Fibre Channel–based storage under low and moderate workloads, but exhibited better performance under heavy workloads, particularly when compared with the 2 Gbps Fibre Channel–based storage. Because the PowerVault MD3000 combines enterprise-class storage features with high performance and can typically cost less than comparable Fibre Channel implementations, it can offer better price/performance than Fibre Channel–based storage.

Deploying cost-effective database storage

The Dell PowerVault MD3000 offers enterprise-class performance, manageability, and scalability along with key features enabling data protection,

storage partitioning, and virtual disk snapshots and copies. Deploying the PowerVault MD3000 as part of a Dell PowerEdge server cluster running Oracle Database 10g with Oracle RAC can help provide highly available storage clusters in a cost-effective way, particularly in small and medium-size enterprises. [u](#)

Mahmoud Ahmadian is an engineering consultant with the Database and Applications team of the Dell Product Group. He has an M.S. in Computer Science from the University of Houston, Clear Lake.

Chethan Kumar is a systems engineer and adviser in the Database and Applications team of the Dell Product Group. He has an M.S. in Computer Science and Engineering from the University of Texas at Arlington.