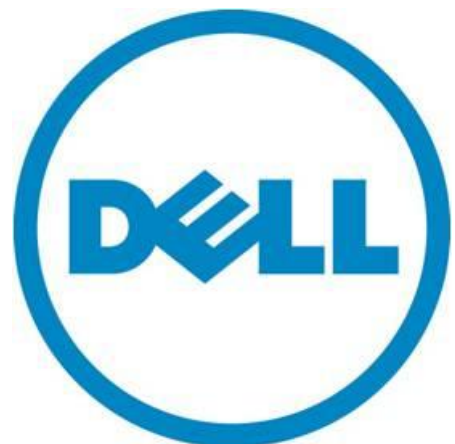


# Dell PowerVault™ MD Series Storage Arrays: IP SAN Best Practices

---

A Dell Technical White Paper



THIS WHITE PAPER IS FOR INFORMATIONAL PURPOSES ONLY, AND MAY CONTAIN TYPOGRAPHICAL ERRORS AND TECHNICAL INACCURACIES. THE CONTENT IS PROVIDED AS IS, WITHOUT EXPRESS OR IMPLIED WARRANTIES OF ANY KIND.

© 2010 Dell Inc. All rights reserved. Reproduction of this material in any manner whatsoever without the express written permission of Dell Inc. is strictly forbidden. For more information, contact Dell.

*Dell*, the *DELL* logo, and the *DELL* badge, *PowerConnect*, and *PowerVault* are trademarks of Dell Inc. *Symantec* and the *SYMANTEC* logo are trademarks or registered trademarks of Symantec Corporation or its affiliates in the US and other countries. *Microsoft*, *Windows*, *Windows Server*, and *Active Directory* are either trademarks or registered trademarks of Microsoft Corporation in the United States and/or other countries. Other trademarks and trade names may be used in this document to refer to either the entities claiming the marks and names or their products. Dell Inc. disclaims any proprietary interest in trademarks and trade names other than its own.

March 2011

## Contents

Introduction .....	2
iSCSI Overview .....	2
IP SAN Design.....	2
Best Practices – Implementation.....	3
Redundancy.....	3
Security.....	4
IP SAN Network Infrastructure .....	5
General Network Practices.....	5
Ether Flow Control.....	5
Unicast Storm Control.....	6
Jumbo Frames .....	6
IP SAN Optimization .....	6
SUMMARY.....	10

## Figures

Figure 1. Fully Redundant MD32X0i Configuration .....	3
Figure 2. Fully Redundant MD36X0i Configuration .....	4
Figure 3. MD32X0i Controller Configuration.....	7
Figure 4. MD36X0i Controller Configuration.....	7
Figure 5. MD32X0i in a Network .....	8
Figure 6. MD36X0i in a Network .....	9

## Introduction

This document provides guidance for optimizing an IP storage area network (SAN) environment utilizing the Dell PowerVault™ MD32X0i (1 GB iSCSI technology) and MD36X0i (10 GB iSCSI technology) series of storage arrays. The best practices in this document are recommendations for providing a fault tolerant, high performance environment in which to maximize the SAN capabilities of the MD32X0i and MD36X0i series. The recommendations may be applied according to the requirements of the environment in which the installed storage array or arrays are utilized, and not all best practices may be applicable to all installations. The best practices in this paper are focused on Dell Inc. technology-based solutions.

## iSCSI Overview

iSCSI is a block-level storage protocol that lets users create a storage network using Ethernet. iSCSI uses Ethernet as a transport for data from servers to storage devices or SANs. Because iSCSI uses Ethernet, it does not suffer from some of the complexity and distance limitations that encumber other storage protocols.

The iSCSI protocol puts standard SCSI commands into TCP and sends those SCSI commands over standard Ethernet. An iSCSI SAN consists of servers with an iSCSI host bus adapter (HBA) or network interface card (NIC), disk arrays, and tape libraries. Unlike other SAN technologies, iSCSI uses standard Ethernet switches, routers, and cables, and the same Ethernet protocol deployed for communications traffic on LANs (TCP/IP). It can take advantage of the same type of switching, routing, and cabling technology used for a LAN.

Because iSCSI uses SCSI commands and relies only on Ethernet to transport the SCSI commands, operating systems view iSCSI-connected devices as SCSI devices and are largely unaware of whether the SCSI device resides across the room or across town.

Most components inside these iSCSI devices are very familiar to network professionals, including RAID controllers and SAS and Nearline SAS drives. The only added feature is the iSCSI protocol, which can be run on standard NICs, in software or on specialized iSCSI silicon, or HBAs that off-load the TCP/IP and iSCSI protocol.

iSCSI is built using two of the most widely adopted protocols for storage (SCSI) and networking (TCP). Both technologies have undergone years of research, development and integration. IP networks also provide the utmost in manageability, interoperability and cost effectiveness.

## IP SAN Design

For an IP SAN, the network infrastructure consists of one or more network switches or equivalent network equipment (routers, switches, etc.). For the purpose of this document, it is assumed that the network has at least one switching or routing device. An IP SAN therefore consists of one or more hosts, connected to one or more storage arrays through an IP network, utilizing at least one switch in the network infrastructure.

There are several factors to keep in mind when designing an IP SAN. The importance of these factors depends on the specific implementation of the IP SAN. These factors include but are not limited to:

- **Redundancy:** If data availability is required at all times, consider a fault tolerant IP SAN.
- **Security:** Depending on your IP-SAN implementation, different security mechanisms can be taken into consideration. This includes dedicated networks, CHAP, array passwords, etc.
- **Network Infrastructure:** Components of the network infrastructure like 1Gb or 10Gb infrastructure, NICs, HBAs, switches, cabling, routing, etc. can affect IP SAN performance and maintainability.
- **Optimization:** Depending on the application, various elements of your IP SAN can be tuned for improved performance. Some of these include the ability to use hardware offload engines, jumbo frames, etc.

## Best Practices – Implementation

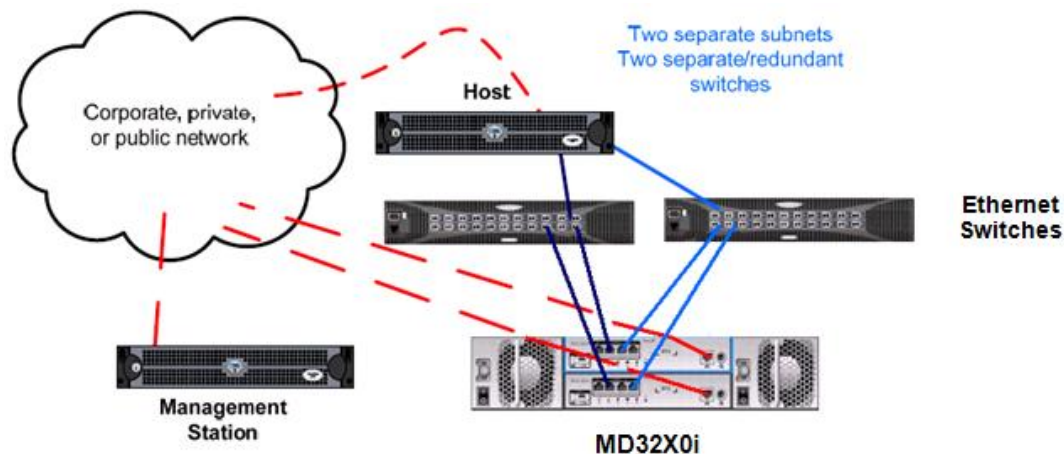
There are many ways to implement an IP SAN based on need, available resources, and intended application. For instance, one important but easily overlooked item that can improve the manageability of your IP SAN implementation is to assign a consistent and representative naming scheme to the storage arrays. This is especially useful if the SAN has more than one storage array attached. The blink array feature of the MD Storage Manager can be used to correctly identify each array physically.

Some implementation guidelines are described below. However, note that these are general guidelines and may not benefit all applications.

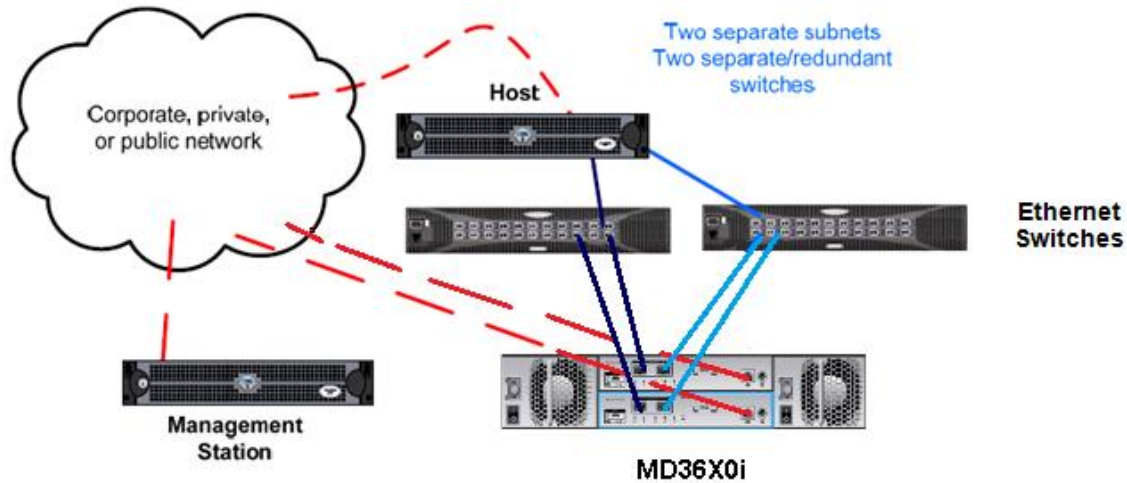
### Redundancy

“Redundancy” refers to a second set of hardware and communication paths. If one piece of hardware on one path breaks down there is a second path that can be utilized. In an IP SAN this can be achieved with a second controller in the array and by using two different switches for the iSCSI network. The following diagrams demonstrate such a configuration.

Figure 1. Fully Redundant MD32X0i Configuration



**Figure 2. Fully Redundant MD36X0i Configuration**



**Physical Network Infrastructure:** A fully redundant IP SAN is characterized by multiple physical independent iSCSI data paths between the hosts and the array. Each data path should be on a separate subnet.

- **iSCSI configuration:** In the case of an iSCSI target like the MD32X0i and MD36X0i, it is recommended that you establish multiple sessions to the storage subsystem from each host. It is also recommended that you set up one session per port from each of the network cards to each RAID controller module. This method allows one session to restart if a link goes down while not affecting any of the other links.
- **RAID:** Choose an appropriate RAID level based on your applications. RAID 1 or RAID 5 provides some level of redundancy, which will be useful in the case of failed physical disks. Each RAID level works best with certain applications and this should be taken into consideration when configuring the MD32X0i and MD36X0i.
- **Power:** Each redundant component of the data path should be on a separate power source. This ensures that even if one component fails due to a power issue, the alternative path will continue to work. In the same way, two power supplies of the MD32X0i and MD36X0i should be connected to separate power sources.

## Security

The optimal way to ensure data security on an IP SAN is to implement an isolated, physically independent network for iSCSI data traffic. Besides providing better security, a dedicated storage traffic network avoids the congestion of the non-storage traffic network.

**VLAN:** If physically isolated iSCSI networks are not feasible, then VLANs can help to separate iSCSI traffic from the general traffic in the network. It is recommended that you turn on VLAN tagging. The MD32X0i and MD36X0i series of storage arrays support VLAN tagging. A port can either transmit all tagged IP packets or all non-tagged IP packets.

VLAN must be enabled throughout the entire iSCSI SAN from the NICs, switches, and iSCSI ports; otherwise, behavior may be inconsistent. To simplify troubleshooting initial deployments, make sure those NICs, switches, and MD32X0i and MD36X0i series of storage arrays are fully operational before enabling the VLAN feature solution-wide.

**CHAP:** To have secure access between your host and array, target and mutual CHAP authentication should be enabled on the host(s) and storage array(s). Follow standard CHAP password guidelines for best security.

It is highly recommended that you set a password on all devices with your IP SAN, and use a strong password that meets standard IT guidelines.

## IP SAN Network Infrastructure

### *General Network Practices*

Make sure the category rating for the cables used are gigabit Ethernet compliant (CAT5e, CAT6). Design the network to have the least amount of hops between the array(s) and the host(s). This will greatly reduce failure points, simplify manageability, and reduce latency and complexity of the network architecture (particularly in the area of redundancy). Managed switches are recommended because they provide advanced features to help you optimize and maintain the network for your application. It is recommended that you use auto-negotiation only, since gigabit Ethernet networks are designed to have auto-negotiation enabled. If a particular application requires a specific speed/duplex mode, change the advertisement options of the switch.

**Spanning-Tree Protocol:** It is recommended that you disable spanning-tree protocol (STP) on the switch ports that connect end nodes (iSCSI initiators and storage array network interfaces). If you decide to enable STP on those switch ports, then turn on the STP FastPort feature on the ports in order to allow immediate transition of the ports into forwarding state. PortFast immediately transitions the port into STP forwarding mode upon linkup. The port still participates in STP, so if the port is to be a part of the loop, the port eventually transitions into STP blocking mode.

- PowerConnect Switches default to RSTP (Rapid Spanning Tree Protocol), an evolution in STP that provides for faster Spanning tree convergence and is preferable to STP.
- The use of Spanning-Tree for a single-cable connection between switches or the use of trunking for multiple-cable connections between switches is encouraged.

**TCP Congestion avoidance:** TCP Congestion Avoidance is an end to end flow control protocol that limits the amount of data sent between a TCP sender and transmitter. This protocol uses a sliding window to size the data being sent to the TCP receiver. It starts with a small segment size and increases with each packed segment sent until a segment is dropped. Once it is dropped, TCP starts the process over again.

### *Ether Flow Control*

Dell recommends that you enable Flow Control on the switch ports that handle iSCSI traffic. In addition, if a server is using a software iSCSI initiator and NIC combination to handle iSCSI traffic, you must also enable Flow Control on the NICs to obtain the performance benefit. On many networks, there can be an imbalance in the network traffic between the devices that send network traffic and the devices that receive the traffic. This is often the case in SAN configurations in which many hosts (initiators) are communicating with storage devices. If senders transmit data simultaneously, they may

exceed the throughput capacity of the receiver. When this occurs, the receiver may drop packets, forcing senders to retransmit the data after a delay. Although this will not result in any loss of data, latency will increase because of the retransmissions, and I/O performance will degrade.

The Flow Control default for PowerConnect switches is “off”. The MD32X0i and MD36X0i Series of storage arrays will auto-configure to the switch when Flow Control is turned on.

### ***Unicast Storm Control***

A *traffic storm* occurs when a large outpouring of packets creates excessive network traffic that degrades network performance. Many switches have traffic storm control features that prevent ports from being disrupted by broadcast, multicast, or unicast traffic storms on physical interfaces. These features typically work by discarding network packets when the traffic on an interface reaches a percentage of the overall load (usually 80 percent, by default).

Because iSCSI traffic is unicast traffic and can typically utilize the entire link, it is recommended that you disable unicast storm control on switches that handle iSCSI traffic. However, the use of broadcast and multicast storm control is encouraged. See your switch documentation for information on disabling unicast storm control.

### ***Jumbo Frames***

Dell recommends that you enable jumbo frames on the switch ports that handle iSCSI traffic. In addition, if a host is using a software iSCSI initiator and NIC combination to handle iSCSI traffic, you must also enable jumbo frames on the NICs to obtain the performance benefit (or reduced CPU overhead) and ensure consistent behavior.

Jumbo frames must be enabled throughout the entire iSCSI SAN from the NICS, switches, and array ports; otherwise, behavior may be inconsistent. To simplify troubleshooting initial deployments, make sure that NICs, switches, and MD32X0i and MD36X0i storage arrays are fully operational before enabling jumbo frames.

## **IP SAN Optimization**

When designing your IP SAN you must consider various factors in your network and the actual application you are using. There are some general guidelines for designing your IP SAN. To maximize the data throughput of your storage arrays, all data ports need to be utilized. If your application is IO intensive, utilizing iSCSI offload NICs is recommended. Consider manually balancing your virtual disk ownership so that no single controller is processing an excessive amount of I/O relative to the other controller.

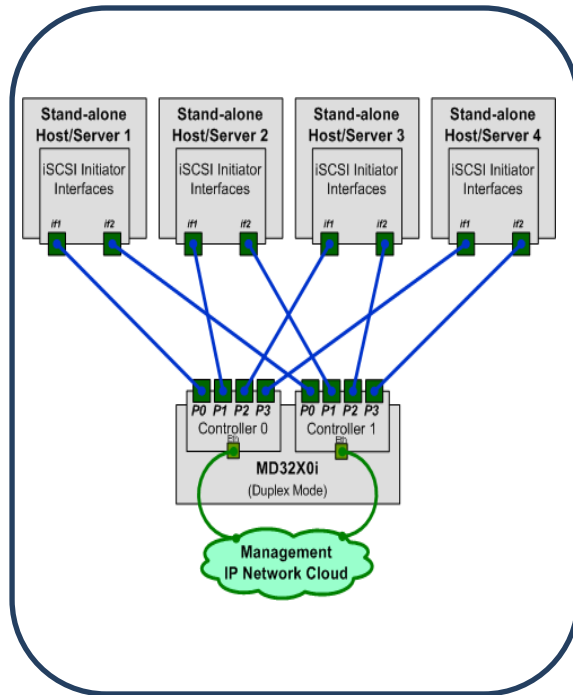
The MD32X0i and MD36X0i series of storage arrays support active/active controllers, with each controller able to simultaneously process IO. The asymmetric design of the controllers means that a virtual disk (LUN) is owned by a controller and all IO access to the virtual disk is only possible through the owning controller. To take advantage of both the controllers for IO access, virtual disks can be distributed among the controllers. Virtual disk ownership can be modified to balance IO access, in order to balance utilization of both controllers. If a host configured for redundant access loses IO access to a virtual disk through its owning controller, the failover drive will execute ownership transfer from one controller to the other and resume IO access through the new owning controller.

The following figures illustrate the active/active asymmetric architecture of the MD32X0i and MD36X0i. The configurations consist of two virtual disks (Virtual Disk 0 and Virtual Disk 1), with Virtual Disk 0

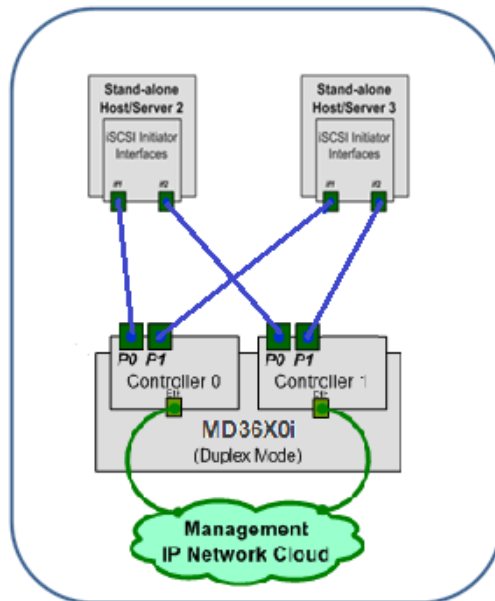


owned by Controller 0 and Virtual Disk 1 owned by Controller 1. Virtual Disk 0 is assigned to Host 1 and Virtual Disk 1 assigned to Host 2.

**Figure 3. MD32X0i Controller Configuration**



**Figure 4. MD36X0i Controller Configuration**



Virtual disk ownership defined by the asymmetric architecture ensures that Host 1 accesses Virtual Disk 0 through Controller 0 and Host 2 accesses Virtual Disk 1 through Controller 1.

**Bandwidth Aggregation:** With the MD32X0i and MD36X0i you can connect two Ethernet ports from one host to one controller, and the bandwidth will be aggregated. Set up the MD32X0i and MD36X0i iSCSI driver with a Round Robin Queue to aggregate all the packets being sent to that controller, placing them on each link and thereby doubling the available bandwidth.

Figure 5. MD32X0i in a Network

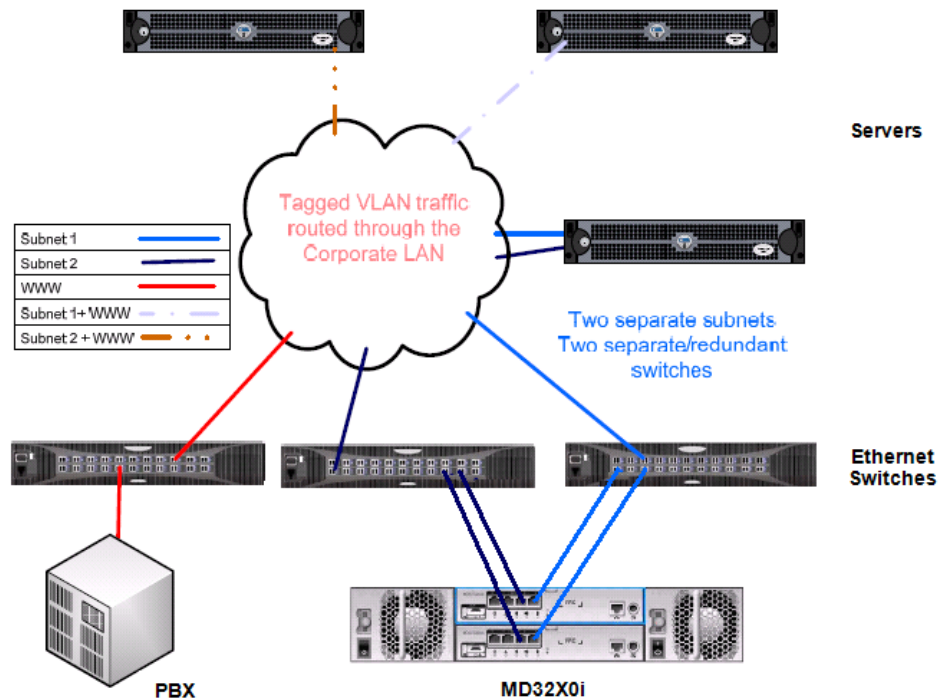
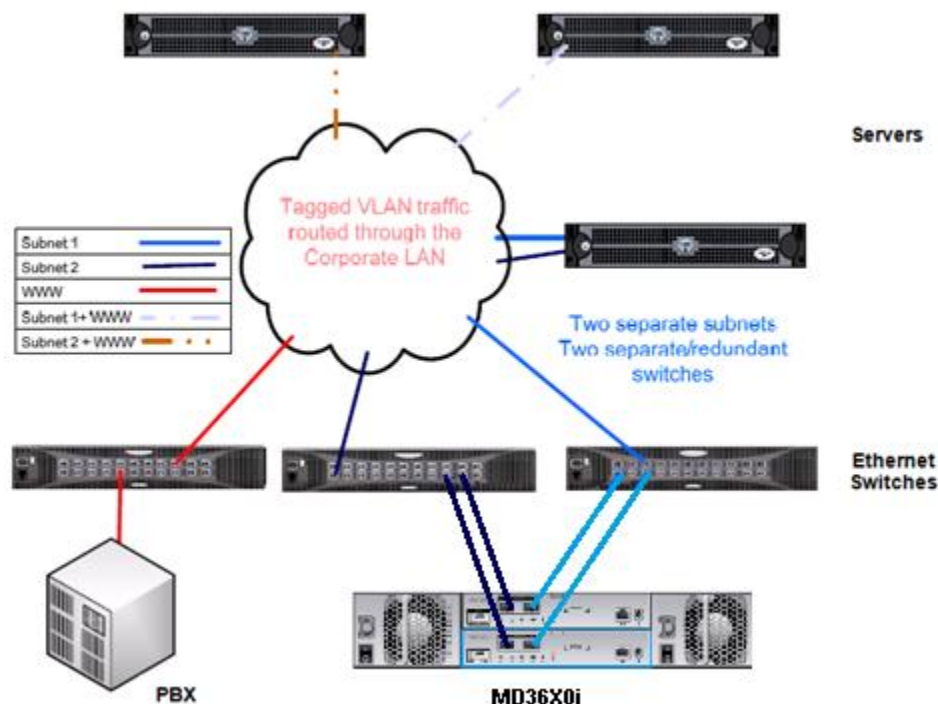


Figure 6. MD36X0i in a Network



Examine your network architecture to ensure there is no bottleneck between the host and array. Ways to optimize your IP SAN include using separate switches to physically isolate the iSCSI data traffic, and using VLANs with FastPort turned on.

**Layer 2 Optimization:** When setting up the VLAN through your network, VLAN tagging can be helpful in routing the iSCSI data traffic through your network. You can then set priority within the VLAN, but you have to look at all traffic to determine priorities. If, for example, your VOIP traffic runs through the same VLAN, you must ensure that voice quality is not hurt, and you need to look at general internet traffic versus iSCSI and VOIP.

**Layer 3 Optimization:** Differentiated Services (DiffServ) provides a good method for managing your traffic. Some switches have a proprietary implementation called Quality of Service (QoS). DiffServ uses the Differentiated Services Code Point (DSCP) to distinguish between service levels of each IP connection. These service level agreements are on a Per-Hop Basis (PHB); within the internal corporate network, traffic flows can be predictable, but once a WAN link leaves the company the service agreements are no longer valid. There are four levels normally used with DiffServ.

- *Default PHB*—Typically best-effort traffic.
- *Expedited Forwarding (EF)* PHB—Low-loss, low-latency traffic.
- *Assured Forwarding (AF)*—Behavior group.
- *Class Selector* PHBs—Defined to maintain backward compatibility with the IP Precedence field.

In order to choose the service level, examine the needs of the applications connected to the array. For instance, if hosts are set to iSCSI boot, or are using virtualization to “hide” the array, and the guest OS boots off a C: drive on the array, select expedited forwarding, since the data must get there and if there is much delay, the host will lock up. On the other hand, if you want all your traffic coming in from the world wide web, set to the lowest possible class of assured forwarding, so it does not affect your critical data.

## **SUMMARY**

An IP SAN is a flexible, easy to deploy and use storage solution for businesses of all sizes. By following the practices recommended in this white paper and using regular IT best practices you can have a highly reliable, flexible data storage solution. Remember to design and build your corporate network with the IP SAN in mind; as your data needs grow so will your data traffic. Follow the recommendations in this white paper and you will be in a much better position to deal with those changes.