

 SearchStorage.com All-in-One Guide

Advanced Storage

Chapter 3: SAN

SAN deployments can be complicated — an administrator must configure the hardware elements to achieve a mix of high performance and high reliability. Then, the storage resources must be organized and allocated to the various users or applications across the organization. If you're comfortable with the basics of storage networks, this guide looks at some of the details involved in components and architecture, connectivity and protocols and management issues.



Advanced Storage Chapter 3: SAN

Table of Contents

[Overview](#)

[SAN components and architecture Expert: HBAs — Are dual HBAs worth the investment?](#)

[SAN components and architecture Expert: Multipathing during boot](#)

[SAN components and architecture Expert: GBIC connection to single or multimode fiber](#)

[SAN components and architecture Expert: Troubleshooting clustered server issues](#)

[SAN components and architecture Expert: Set up persistent binding on an HBA](#)

[Connectivity and protocols Expert: Fibre Channel Arbitrated Loop or switched disk?](#)

[Connectivity and protocols Expert: Fiber optics over distance](#)

[Connectivity and protocols Expert: Wavelength vs. dark fiber for SAN](#)

[Connectivity and protocols Expert: How to minimize I/O latency](#)

[Connectivity and protocols: The slow move toward faster Fibre](#)

[Connectivity and protocols Expert: FC-AL port protocol for storage](#)

[SAN management: Testing iSCSI SAN performance with Iometer](#)

[SAN management: SAN fine tuning — Get the most from your storage](#)

[SAN management Expert: LUN expansion with Microsoft cluster server](#)

[SAN management Expert: Testing iSCSI SAN performance with Iometer](#)

[SAN management Expert: Windows can't read from the bootdisk](#)

Overview

By Stephen J. Bigelow, Features Writer

August 30, 2006 | SearchStorage.com

A [storage area network](#) (SAN) is vital to the enterprise. SANs concentrate storage into one ubiquitous and specialized network. But SAN deployments can be complicated — an administrator must configure the hardware elements to achieve a mix of high performance and high reliability. Then the storage resources must be organized and allocated to the various users or applications across the organization. Now that you've learned a bit about the [basics of storage networks](#), let's look at some of the details involved in components and architecture, connectivity and protocols, and management issues.

SAN components and architecture

Although the cabling used for many SAN implementations is tough and reliable, there are some important rules that can ease deployment and subsequent troubleshooting. Whether optical or copper, cables should always be labeled in a clear and consistent manner — it doesn't matter what labeling scheme you adopt, but it should be followed faithfully. Once cabling is installed or replaced, make the time to denote it on SAN (and physical plant) documentation so that others can follow the SAN later. Poor installation practices also include tight runs, excessive bends (crimps) and inadequate protection across floors and other exposed areas. This is particularly true for optical cable and leads to premature failure. It is vitally important for administrators or other IT staff to oversee the work performed by professional cable installers to verify that reasonable installation practices are followed.

Normally, a single [host bus adapter](#) (HBA) is enough to establish a connection to each SAN server or storage device, but traffic congestion and connection reliability are always a concern. For example, a single HBA link handles all of the traffic to a storage device, so excess traffic can congest the HBA and reduce performance. Also, if an HBA fails, communication is cut off and the affected portion of the SAN can become unavailable. Storage administrators overcome this by implementing multiple HBAs at the storage device. Multiple HBAs can improve better traffic handling through [load balancing](#), as well as [redundancy](#) in the event of a fault. Software tools are used to define the load balancing and [failover](#) behaviors of the HBAs.

Today, [switches](#) are taking on more functionality in the network, supporting management features and intelligent functions, such as [switch-level storage virtualization](#) and data migration. The main benefit of this development is heterogeneity — since switch traffic is largely independent of the particular manufacturers' servers or storage devices on the SAN, a switch can often oversee a much larger suite of systems without interoperability concerns. If you're considering [iSCSI](#) SAN technology in your enterprise, opt for Ethernet (IP) switch devices, rather than Fibre Channel (FC) switches.

SAN connectivity and protocols

SAN connectivity is gaining bandwidth, allowing more users to access burgeoning volumes of data. Fibre channel connectivity is particularly noteworthy, building on traditional 1 gigabyte per second (Gbps) and 2 Gbps speeds by adding 4 Gbps support available today, and even implementing some 10 Gbps ports. However, 10 Gbps FC is still in limited deployment and only found in high-end SAN devices like McData Corp.'s Intrepid series or the MDS 9513 director-class switch from Cisco Systems Inc. Storage switches typically support common protocols like SCSI [FC Protocol](#) (FCP) on open systems and [FICON](#) for IBM mainframes. An increasing number of storage switch products are including Ethernet/IP ports to handle iSCSI, iFCP and FCIP protocols.

As SANs grow, the number of switches deployed in the SAN also increases, and this can eventually lead to performance degradation due to interswitch latency (passing traffic switch-to-switch). Switches normally interconnect using dedicated interswitch link (ISL) ports. Reduce ISL latency by choosing switches with fast ISL ports, or trunk multiple ISL ports together for improved performance and redundancy.

iSCSI SAN deployments are gaining ground — largely due to the simplicity and low cost of Ethernet technology, as well as the ready availability of [Gigabit Ethernet](#) (GigE) components. If GigE connectivity is not available, iSCSI performance is often bolstered by the use of network interface cards (NIC) with TCP/IP offload engine (TOE) features. iSCSI HBAs are available from major component makers, such as Adaptec Inc., Intel Corp. and QLogic Corp. With [10 GigE](#) in the near future, iSCSI may eventually become a serious contender for enterprise SAN deployments against 10 Gbps FC. iSCSI software and operating system (OS) support are also improving, with initiator (client) and target (server) software is readily available for most enterprise OSs, including Windows, AIX, NetWare and Linux.

Whether you're using FC or IP connectivity in the SAN, it's important to consider the connection speeds at every link between a server and storage device. As SAN deployments grow and change, administrators sometimes forget that all data in a given path will only be as fast as the slowest link. Investing in a 10 Gbps switch may not be worthwhile if the main storage system is only communicating at 2 Gbps. Auto-negotiation can also be an issue if one link fails to shift its speed properly to accommodate another data rate. Many administrators choose to eliminate potential connectivity problems by configuring the speed of each link manually.

SAN management

As storage capacity spirals upward, IT staff has remained virtually unchanged; and has even been reduced in some cases. An administrator that might have managed 3 terabytes (TB) just a few years ago is now often responsible for 15 TB or more. This trend has put a new emphasis on SAN management — especially in areas of process control and automation.

[RAID](#) platforms often include management tools, but the goal is to provide administrators with maximum flexibility, while incurring an absolute minimum of downtime. For example, RAID arrays typically required downtime to add disks or to change the RAID group size, and the entire RAID group would have to be

rebuilt when changing RAID levels (e.g., migrating from RAID-5 to RAID-6). Today, look for RAID platforms and management tools that can support these features on the fly.

One of the biggest growth areas in SAN management is [storage resource management](#) (SRM) tools. SRM tools can analyze and report on available storage systems and utilization, and improved analytical features can even help to ease trouble areas. For example, Softek Storage Solutions Corp.'s Performance Tuner first establishes a performance baseline across the SAN and then proactively alerts IT staff when performance falls outside of the norm — offering suggestions for improvement.

But the most common push for SRM tools is in heterogeneity and automation. CA's BrightStor product supports over 100 storage arrays, SAN switches, tape libraries and applications. This kind of interoperability is important in order to centralize management functions. BrightStor also supports 500 million files and centralizes backup functions across a variety of popular products like ARCserve, Tivoli Storage Manager, Legato Networker, and Veritas NetBackup. Automated provisioning features also save considerable time for storage administrators, and emerging chargeback capability ensures that enterprise SAN users pay for only the storage capacity that they are utilizing — a crucial element in [tiered storage](#) strategies.

SAN components and architecture Expert: HBAs — Are dual HBAs worth the investment?

Greg Schulz

June 14, 2006

We have a 150 server dual-fabric [storage area network](#) (SAN) currently in place. That will grow to 350+ servers before year-end. We have 10 [terabytes](#) (TB) of HDS arrays, soon to be 20 TB, going through two Cisco directors. Some of our servers have dual [HBAs](#), but many do not. Management is balking at spending the extra funds to go with dual HBAs for all servers, basically saying to live with single HBAs for the time being. What are the pros and cons of this approach, as you see it? We can already see concerns with regard to things like future microcode upgrades, cache upgrades, unplanned outages, etc. What would you recommend (if money were no object)? Thanks in advance.

A question for your management is what value they place on a specific application or server's availability and uptime. If a particular application or server can be offline and unavailable for some period of time for planned and unplanned outages, then a single HBA might be applicable. On the other hand, if the value of application availability exceeds the costs of an HBA; you need to work with your management so they will understand that the price of an HBA can be less than the cost for downtime. When possible, look for ways to build in redundancy including dual HBAs or network [NICs](#) unless there is no business reason or benefit to do so. It is important to consider the potential consequences to your customers and applications users.

Without knowing more details about your particular environment, applications, performance, physical layout, network and availability requirements, I hesitate to make any specific recommendation. However, generally

speaking, I would look to see if some level of [server consolidation](#) using [server virtualization](#) is appropriate to reduce the number of physical servers, and thus, the number of HBAs and associated SAN ports. However, this would be dependent on the application's performance requirements as well as application and business needs for dedicated servers. With server consolidation comes the risk of putting too many eggs (or applications) in a basket (server), so be mindful of high availability and server failover capabilities.

What you may find is that some servers could be consolidated onto a larger, higher performance server with redundant 2 gigabit (Gb) or 4 Gb [Fibre Channel](#) (FC) adapters to reduce the total number of servers, and thus, adapters. Another consideration to explore is attaching some of the servers using iSCSI to access FC storage via your Cisco directors or third-party [iSCSI-to-FC bridge/gateway/router](#) eliminating the need for FC HBAs on your servers. Of course, you would need to have available Ethernet ports or add additional NICs to your servers if needed, thus there would be a cost tradeoff to consider.

SAN components and architecture Expert: Multipathing during boot

Arun Taneja
May 5, 2005

When booting from SAN, how can I provide multiple paths via HBA's? If I boot from the SAN and my HBA fails, connectivity to my OS fails. Is automatic failover to my second HBA possible to maintain connectivity to the operating system on the SAN?

The short answer is: No. Multipath is not supported during boot, since it depends on special multipathing drivers to load, which happens during boot. So, if a path fails during boot, before these drivers have loaded, the boot will probably fail. A server, however, may be configured (this varies by OS) to boot from a number of locations in a sequence. If one boot device fails, booting would happen from the second source.

SAN components and architecture Expert: GBIC connection to single or multimode fiber

Greg Schulz
September 9, 2005

Is it possible to connect a SX GBIC to both single-mode fiber and multimode fiber? What will be the effect of LX GBIC to both single-mode fiber and multimode fiber?

Without using some type of mode conditioner or conversion device, single-mode fiber (SMF) transceivers (e.g. [GBIC](#) and SFP) need to attach to SMF cabling and other SMF transceivers. The same holds true for multimode fiber (MMF) transceivers and devices. Short-range or short-distance transceivers such as GBICs

and SFPs (also known as SX) utilize MMF cabling unless some type of mode conditioner or conversion device is employed to attach to SMF. Long-distance transceivers (also known as LX) typically utilize SMF cabling.

The fundamental difference lies in the type of cabling and optical transceivers. SX transceivers utilize lower-powered, lower-cost optics and lower-cost, shorter-distance MMF cabling while LX utilizes higher-powered, more accurate optics for longer distance with more expensive SMF cabling.

Keep in mind that when you introduce a mode conditioner or mode conversion device, you will introduce additional degradation into the optical signal (e.g. db loss). SX and MMF should enable distances depending on speed (e.g. 1Gbit, 2Gbit, 4Gbit, 10Gbit) of the interface to a couple of hundred meters (actual distance will vary with speed). Using SMF and LX transceivers and depending upon [optical cabling](#), repeaters and other technology including CWDM, WDM and [DWDM](#) distances of 10s to 100s of kilometers can be supported. Avoid simply looking at distance and attainable [bandwidth](#) supported and also look at effective latency and effective bandwidth over the supported distance.

SAN components and architecture Expert: Troubleshooting clustered server issues

Marc Staimer

December 1, 2005

I have a two-node clustered server. Both servers have two host bus adapters (HBAs) connected to two fabric switches and to a redundant SAN controller. My problem is only one path has connectivity to storage device, HBA 1 of server 1 and HBA 1 of server 2 has no connectivity to storage device. Both fabric switches are working fine. What could be wrong, can you help me?

Since I do not know enough about your environment to give you definitive answers, here are some things you can look at. Let's start with some procedural questions. Can your Fibre Channel ([FC](#)) [switches](#) see/discover [HBA 1](#) in both [servers](#)? If not, that would explain your problem. What multipathing software are you running on your servers? (MPIO, PowerPath, DMP, etc.) Is it compatible with your target storage? If you do not have multipathing software, that could explain your problem. If you do, and it is not compatible with your target storage, that also could explain your problem. How are your HBAs and target storage ports [zoned](#)? If HBA 1 of both servers are not correctly zoned to the right target storage ports, that too will cause your problem.

There is currently only one software package on the market that can instantly diagnose your problem. It is called SANScreen from Onaro. It is a bit pricey and typically only cost justifies for SANs larger than 128 ports.

You can also contact your server, HBA, multipath software, FC switch or target storage vendor for troubleshooting help.

SAN components and architecture Expert: Set up persistent binding on an HBA

Chris Poelker
April 5, 2004

How can I set up persistent binding on an [HBA](#) for a [Unix](#) host?

That is a function of the driver for the specific [HBA](#) and the SD.conf file. Use the GUI that came with the driver install software to set persistent bindings.

For JNI adapters, the utility is EZFibre. Use the target to [WWN](#) binding feature for the driver to set the WWN in the /kernel/drv/sd.conf file. You can set the proper bindings in the fcaw.conf file for JNI adapters like this:

```
# Configuration flag def_wwn_binding
# Type: string; default: "$xxxxxxxxxxxxxxxx" (means WWN is "static don't care")
# Sets the 16 digit hexadecimal default wwn binding for every target/lun
# instance which does not explicitly define one.
#     - A "$" preceding the string indicates static binding enabled
#     - A "x" in place of a digit indicates "don't care" for that digit
# *See technote for details on wwn bindings
def_wwn_binding = "xxxxxxxxxxxxxxxx";
For the SD.conf file, follow this format:
name="sd" class="scsi" target=1 lun=0; wwn="2000004568018769";
name="sd" class="scsi" target=1 lun=1;
name="sd" class="scsi" target=1 lun=2;
```

Connectivity and protocols Expert: Fibre Channel Arbitrated Loop or switched disk?

Greg Schulz
July 13, 2006

We are considering buying high-end storage from two different vendors, one uses a one uses Fibre Channel Arbitrated Loop technology and one uses a switched-disk interface. Of course, each vendor is saying that their technology is better. We'd like to hear from independent source and expert on this dispute.

Traditional [Fibre Channel](#) Arbitrated Loop (FCAL) has been used on the back-end of storage systems for attachment of disk drives and disk expansion shelves to controllers. Similar to the way the host or server

side attachment ports changed over to switched Fibre Channel several years ago, the new standard for storage systems is to leverage a switched or semi-switched (switched loop) connection.

The advantage to using a switched interface is simpler point-to-point connectivity and in some cases, potential for better performance and reliability. The lower cost of switching technology and embedded switch chips has given rise to the shift to switched bunch of disk (SBOD) for storage.

Connectivity and protocols Expert: Fiber optics over distance

Greg Schulz

June 12, 2006

We were having a discussion the other day about using a LX [GBIC](#) to push a signal farther than the typical SX limitation of 550 m over a multimode fiber. I know we can do it, but the point I was trying to make is that the GBIC regulates whether the signal propagation is multimode (multiple paths along the fiber) or single mode (single path along the fiber) and not the actual fiber that refracts the light. The way I understand it, the multimode fiber allows the signal to reflect at the correct angle to reach the end of the tunnel. Is my assumption correct, or does the fiber actually regulate the propagation of the light?

To achieve long distances (LX) using fiber-optic technologies for [Fibre Channel](#) (FC), [Ethernet](#), [SONET/SDH](#), [DWDM](#) or other networks requires several components. The components needed for long distance include a LX GBIC (aka SFP) plus single-mode (SMF) cabling. To achieve ultra-long distances, higher power LX GBICs or DWDM and possibly repeater technologies will be needed. If you have single-mode fiber, but you do not have a powerful enough LX GBIC, your distance will be limited. The fiber-optic cable is an important component, for example good quality cable fewer connections will result in less db (signal) loss or light degradation that in turn would limit your distance. The GBIC or DWDM type device is essential for generating and pushing the light wave across the fiber-optic cable. For long distances, you need the combination of SMF optic cabling and an SMF LX GBIC.

Make sure you verify what distance the LX GBIC that you plan to use is capable of supporting as the distances can vary from 10 km to 80 km. You can learn more about fiber optics over distance in chapter 5, "Fiber optic essentials," in my book "Resilient Storage Networks" (Elsevier).

Connectivity and protocols Expert: Wavelength vs. dark fiber for SAN

Christopher Poelker
January 12, 2006

Which do you consider the best option for connection of a SAN network: wavelength products from the local LEC or dark fiber? Please explain the pros and cons of each.

The answer is "it depends" on what you are trying to do.

The local cable plant (all the devices connected to a fabric at one site) should be using 50u (u = micron) multimode cabling (850 nanometer wavelength), which allows a distance of up to 300 meters using 2 gigabit (Gb) lasers (GBICS), and 500 meters using 1 Gb lasers. So, if all you are trying to do is connect servers to [SAN switches](#) and SAN switches to storage, then your best option is to use standard multimode SAN cables as described above.

[Dark fiber](#) is a term used to describe 9u single-mode cabling. Single-mode cables have a narrower core, which limits light defraction and absorption during transmission, which means greater distances can be traveled before db loss is an issue. Single-mode transmission also uses a 1300 nanometer (nm) wavelength source, which also allows for greater distance. Single-mode cables can be run up to 10 kilometers before repeaters are required. So, if you are trying to connect two SAN islands together over greater distances, than dark fiber (9u single-mode 1300 nm) is the better choice.

If there is currently no dark fiber owned by your company between buildings though, installing or leasing dark fiber can be VERY expensive. It can cost the same as leasing an OC-48 connection from a Telco. If you own the fiber though, you can connect up dense wave division multiplexor (DWDM) equipment at either end, and actually slice up your dark fiber into multiple wavelengths (usually 32 or more) that will allow for massive bi-directional bandwidth between sites.

If you don't need all that bandwidth and just want to connect two SAN networks together for, say, data replication between sites for disaster recovery, then using a standard leased [IP](#) connection could also work for you. All you would require is an FC to IP bridge at each location. The bridge can run either the [FC-IP](#) or iFCP protocol to tunnel your FC frames across the IP network to the other location. I would recommend a leased T3 or above for the IP network, although a T1 may do if the amount of data is low, and the bridge also does compression.

Connectivity and protocols Expert: How to minimize I/O latency

Greg Schulz
January 4, 2006

How do you minimize the latency during communication between the storage device and the system? In our case it has to happen through the router. Is there any alternative?

To minimize the latency between a [server](#) and a storage device you want to make sure that there are no bottlenecks, congestion or other points that would add latency and impact performance.

Not knowing the details of your environment, I will assume that by "system" you are referring to a server of some sort. I will also make an assumption that since you mention a [router](#), that there is thus some sort of network exists between your server and storage device.

The more items that sit between an application on a server and the storage device will contribute to overall latency. These items include software layers ([file systems](#), volume managers and network and device drivers), network layers (protocols, interfaces, [switches](#) and routers, cabling, distance, [bandwidth](#), effective latency) and the interfaces on the storage device. So, the fewer the layers, the less overhead and lower the latency should be.

Do you currently have a latency performance issue with the router or are you concerned about a potential challenge? If you are encountering performance problems, test the performance without the router and verify that the router and applicable networks are adequately configured and have the most current firmware and software.

Connectivity and protocols: The slow move toward faster Fibre

Alan Earls
February 9, 2005

At issue: Hungry for faster Fibre Channel? Four Gbps will have to hold you for now.

[Fibre Channel](#) vendors have been talking up a move to 4 Gbps, with possibilities of 8 Gbps or 10 Gbps shimmering on the horizon, but some analysts remain skeptical.

At a recent 4 Gbps plugfest conducted at University of New Hampshire InterOperability Laboratory, Engenio Information Technologies announced its native 4 Gbps disk array technology as the industry's first to be tested for compliance with the latest Fibre Channel Industry Association 4 Gbps protocol. The array has

also been tested for compatibility with current 2 Gbps and legacy 1 Gbps devices, according to the vendor, in order to ensure full backward compatibility.

Looking forward, the Fibre Channel Industry Association (FCIA) has announced ratification of an 8 Gbps Fibre Channel roadmap designed to take it from an "inside-the-box" storage device interconnect into switched [SAN](#) fabrics.

Four Gbps is now starting to show up in subsystems, says Steve Duplessie, principal analyst at Enterprise Strategy Group, Milford, Mass. "I expect to see Engenio announce 4 Gbps throughout their product line as soon as this quarter," he says.

Duplessie points out that "QLogic has had 4 Gbit HBAs for a while, and now all the switch guys are publicly on board."

On the other hand, Duplessie says "10 Gbps is a long way away" for most of the industry and "8 Gbps will most likely never happen," because 10 Gbps is in the works.

Data Mobility Group analyst John Webster says the adoption of 4 Gbps will hinge on price versus performance. If the industry prices it aggressively, it may begin to attract buyers. As for even faster Fibre Channel, Webster sees 10 Gbps, when it becomes available, as primarily a "backbone enabler."

While vendors like McData have already begun to dangle the potential of 10 Gbps in front of buyers of director class switches, Webster says, "The message I'm getting from industry is that it's probably going to be most effectively used by the largest of the large enterprises." As with 4 Gbps, adoption of 10 Gbps is purely a price/performance issue and, admits Webster, it may be a case of the industry actually getting ahead of the real needs of most of its customers.

On the other hand, Webster is bullish on 10 Gbps Ethernet. "I believe there is a feeling among at least some CIOs that they simply don't understand why they are supporting two networks — one for storage and one for the rest of IT," he says. And, as they look at the possibility of building out their IP networks with 10 Gbps, "[iSCSI](#) becomes a potentially disruptive technology."

"Even if it doesn't have all the functionality of Fibre Channel, with 10 Gbps Ethernet people could decide that iSCSI is a good enough for them," he adds.

Connectivity and protocols Expert: FC-AL port protocol for storage

Christopher Poelker
June 3, 2005

I have a technical EMC Clariion question: When connecting a server directly to a CX product, is the FC protocol strictly FC-AL (Fibre Channel Arbitrated Loop)? In other words, does CX only support FC-AL in a direct connection not point-to-point (by direct connection I mean that there is no FC switch between the server HBA and the Clariion host port)?

To the best of my knowledge, that is correct for the CX, and most everyone else's storage. If there is no [switch](#) in the mix, most storage [arrays](#) would need to be set to FC-AL as the port protocol. The switch usually supplies all the FLOGI (Fabric Login) support for the FC-SW connections, including WWN registration. FC-AL is the easiest way to go for non-fabric connections.

SAN management: Testing iSCSI SAN performance with Iometer

Rick Cook
July 26, 2006

iSCSI SAN tuning tip: [Iometer](#), a performance test tool for tuning Windows environments, can also be used to tune performance in iSCSI SANs.

Iometer lets you simulate a [storage area network's](#) (SAN) performance by setting many of the parameters and running tests. Repeated runs with different parameters such as block sizes will let you tune your system.

Since Iometer tests many parameters of the storage network as well as the storage array, you need to isolate [iSCSI](#) performance from the [array](#) performance when running tests. The best way to do this is to set the parameters to optimize storage array performance, so any variations that show up in the tests are the result of the iSCSI network.

To that end, set the parameters to use the maximum number of physical disks hit as well as the maximum number of [LUNs](#). However, since block size critically affects iSCSI throughput, use realistic sizes for the block data.

In fact, the two main factors affecting iSCSI performance are host performance and block size. While block size affects performance on any SAN, iSCSI SANs are particularly affected by the host resources.

The reason is that iSCSI requires a lot of processing power to wrap and unwrap SCSI packets in TCP/IP shells. Even with [host bus adapters](#) (HBAs) with [TCP/IP offload engines](#) (TOE) to handle the bulk of the processing, host processor speed and the amount of available memory can limit the performance of the iSCSI SAN. The Iometer tests will help identify such bottlenecks. Adding more memory or an upgraded processor to the host will often substantially improve SAN performance.

Of course, [block](#) size is an important performance parameter on any kind of SAN. Ideally, you want to match the block size to the size of the data in a typical read or write operation. While this can improve performance with certain kinds of data (notably backups and video editing where block sizes of 32 to 64K or even more are often useful), Microsoft chose the default 4K block size carefully. There may be no particular performance gain in changing it. However, a series of tests with Iometer will tell you more precisely what those gains might be.

SAN management: SAN fine tuning — Get the most from your storage

Jerome M. Wendt

June 28, 2006

What you will learn from this tip: The way key [storage area network](#) (SAN) components, principally [host bus adapters](#) (HBAs) and [switches](#), are configured will determine overall SAN performance. If you know what to look for and how to make adjustments, performance issues can be greatly reduced.

SANs are among the most mission-critical enterprise networks, so organizations tend to overbuy and over-allocate SAN capacity to avert performance problems. Buying more capacity might work sometimes, but it isn't always an option. Storage administrators need to find ways to wring more value out of their SANs, and performance tuning is the place to start.

All SAN hardware vendors offer tools that configure, monitor and collect performance data for their devices. For example, Emulex Corp.'s HBAnyware allows administrators to discover HBAs and collect data about their status, attributes and performance. Brocade Communications Systems Inc.'s optional Advanced Performance Monitoring tool collects SilkWorm director performance data, which can be viewed in its Fabric Manager or exported to Microsoft Excel. EMC Corp.'s Navisphere Analyzer, an add-on to its Navisphere Management Suite, allows Clariion users to gather performance data and identify trends. Yet the key to SAN performance tuning is knowing what device settings to monitor — for HBAs, Fibre Channel (FC) directors, [arrays](#) or [tape libraries](#) — and when their parameters need adjusting.

Use this checklist as a guide for performance tuning:

HBAs

- Leave the FC ports at autonegotiate
- Avoid putting more HBAs on a server than the bus throughput can support
- Turn on the [I/O](#) coalesce parameter in high-performance environments
- Install the latest HBA firmware and driver
- Use multipathing software that supports both load balancing and path [failover](#)

FC switches

- Keep the highest performing directors at the core of the SAN
- Connect storage devices and the highest performing applications to the core
- Benchmark the performance on oversubscribed ports
- Leave the FC ports at autonegotiate for host and storage connections

Of course, SAN tuning is a work in progress. You need to constantly monitor and tinker with your storage fabric on a regular basis to get the best performance. By taking simple steps to tune SAN components before they're installed in the SAN, and knowing which ones to analyze first when performance hiccups occur, the number of performance issues will be greatly reduced down the road.

SAN management Expert: LUN expansion with Microsoft cluster server

Christopher Poelker

January 12, 2006

We are planning to upgrade our EMC storage box by adding additional DAE2 enclosure. Currently, there are four two node cluster are accessing the box. We are planning to increase the LUN size of each cluster. After changing the LUN size, will the cluster will work? We have Windows 2003 cluster servers.

Microsoft [cluster](#) resources require basic disks, as dynamic disks traditionally are not supported within the cluster. This is because dynamic disk resource info is kept in the local registry, and the other node would have issues during a [failover](#). If you happen to use Veritas Foundation Suite 4.2 or above on top of the Microsoft cluster, then [LUN](#) expansion is a simple matter of right clicking on the disk resource and expanding the disk to whatever size the array is presenting.

If you are using standard Windows 2003, then you can use the "DiskPart.exe" utility from the [Microsoft resource kit](#).

Use the "rescan" command at a command prompt with diskpart.exe to allow the OS to see the additional free space for the drive. Then use the "extend size = xxxx" command to extend the disk.

SAN management Expert: Testing iSCSI SAN performance with Iometer

Greg Schulz
December 23, 2005

We have just installed two McData 2640 SAN routers, one in each of our data centers, but we are seeing really bad performance over the WAN. We have set the MTU sizes to 1500, but the network guys keep seeing a 4 K negotiation, is this something the network guys can control? The connection is stable other than the retransmitted segments, the network team believes we are overloading the destination with packets before it can process them.

Without knowing more detail about your environment's network, [WAN](#), applications and workload it would be difficult to determine where the issue exists. It is possible that the source is overloading the destination somehow. However, there are some questions that come to mind to help identify where the issue exists. How are the McData 2640 [SAN routers](#) being used? For example, are they being used for remote [mirroring](#) from an EMC, HDS, HP, IBM or Sun/STK type storage device or some other function?

How are you determining that you are getting bad performance? Is this based upon latency, low [bandwidth](#), errors or some other indicator? And, what is or should be normal?

What is the effective speed of the WAN interface compared to the theoretical (quoted bandwidth)? How are the buffers configured on the destination devices, and what type of devices are they? Have you spoken with the McData technical support folks and/or the technical support folks of storage devices that are using the routers?

Lastly, can other applications get better bandwidth and utilize a larger MTU on the same WAN interface to determine if it is the routers or a combination of destination and source devices?

SAN management Expert: Windows can't read from the bootdisk

Christopher Poelker
December 28, 2005

I'm trying to set up w2k3 server to boot from SAN (HP EVA, Emulex 9802), it seems to work fine until first reboot. Message appears that says Windows "could not read from the selected bootdisk". Well, it looks like something has been read :). Any advice would be helpful.

There is a chapter in my book "Storage Area Networks for Dummies" that includes a step-by-step process for configuring [SAN](#) boot drives, and some of the "gotchas" that can cause what you are experiencing.

Things to look out for:

- 1)** Be sure to use the latest version of boot [BIOS](#) on your adapters, and follow the instructions included in the readme file that comes with the driver version for setting up the boot BIOS correctly.
- 2)** When using more than one [HBA](#) in the server, set up both HBA's for boot, and make sure your path-failover software supports it. Doing so will allow you to survive an HBA failure.
- 3)** Make sure the system BIOS is not trying to boot from any internal disks. You can verify this by disabling the internal [SCSI](#) adapters in the server, and re-trying the boot.
- 4)** Make sure the HBA topology setting is correct for your environment (FCAL or fabric)
- 5)** Make sure you use "single initiator zoning" in your fabric to the port where your boot drive is located. In other words, simply zone the HBA you are booting from to the storage port where the boot [LUN](#) is ported (do this for each HBA in the server). In my experience, If you have more than one target port in the zone for the boot HBA, the BIOS may get confused during boot, and you will experience the exact same symptom you are getting.
- 6)** Get the latest tech-tips from Microsoft on SAN booting technet.microsoft.com and search for SAN BOOT.