Network Storage - The Basics

This note will attempt to answer questions such as "What is Network Storage?" "What is SCSI and How does it Work?" and "Isn't RAID or NAS good enough for my environment?"

Evolution of Network Storage

Network storage enables us to store data in such a way that clients on different systems, possibly running different operating systems, can access the storage. Data storage has evolved through a number of phases. The evolution has been driven by new needs demanding more and more widely available storage and by significant increases in storage capacity and network speed.

In the mainframe environments of the 70's, data was stored on physically separate hardware from the actual processing unit. The stored data was still only accessible through the mainframe system. Then along came mini-computers and the storage picture remained pretty much the same; the mini's processor was in one box with its memory, and the disk drives were in different physical boxes.

With PC servers, storage devices initially were in the same enclosure that housed the CPU, memory, and peripherals. Eventually, SCSI systems emerged where the storage could be directly cabled to the PC enclosure via specialized adaptors and cables.

Each of these three methods of using storage made sense in its time. The problem came as we needed larger and larger amounts of storage; for example what happens to a PC server when we need more storage drives than will fit inside the box. From these ideas, network storage was born. There are many network storage technologies available to enable the movement of data storage boxes outside the CPU system box. Some of the most often used techniques are:

Direct Attached Storage -- DAS

DAS describes a server or workstation where all the storage devices are directly attached to the host system. A familiar example of DAS is an IDE drive inside a desktop computer. Because there are so many computer systems out there, DAS is the most common method of storing data for computer systems. These systems are simple and with today's larger drives, can easily hold several terabytes of RAID storage, which can also be provided to other clients. But if the server goes down or is turned off, that storage is not available to anyone.

Using RAID Systems to Improve Storage

Redundant Arrays of Inexpensive Disks (RAID) systems were developed to provide storage with various forms of fault tolerance. RAID fault tolerance is used for both storage inside the PC enclosure and in external storage enclosures. RAID supports a set of levels to provide its fault tolerance. RAID-0 is called Striping and provides more spindles, so that reading is faster. RAID-0 does not provide parity checks for fault tolerance, but the later standard, RAID-5, does.

With RAID-5, the parity is spread over all the drives in the array, and the storage can remain intact even with the failure of any one drive in the array. With RAID-5, you “lose” one drive for the parity operations; thus if you put 4 1TB drives into a RAID-5 array, the array would only provide 3 TB of storage. This RAID-5 array would be said to have a storage efficiency of 75%.

The RAID-1 standard is usually called mirroring or duplexing. Good systems management practice is to separate the Operating System storage from the data storage. System managers often install the operating system storage as RAID-1, where the data is copied simultaneously or mirrored to two drives, which are kept synchronized by the RAID-1 controller. If one of the two mirrored drives fails, the second drive can easily be setup to become the system drive. In Microsoft servers, a system manager simply has to change one line in the system boot file to point at the other mirror, and reboot the system. This technique provides excellent fault tolerance for system drives, but comes at the cost of providing an entire drive to the mirror. Thus, we say that RAID-1 storage is 50% efficient.

Taking all this into account, we often find even SAS storage composed of two RAID arrays, a RAID-1 array for the operating system (OS) and a RAID-5 array for the data. With modern 1TB Serial Attached storage drives and a large CPU case, a good system manager might be able to provide a 200 GB RAID-1 array for the OS and a 5 TB RAID for the data. Of course, even this system still has the disadvantage that if the server is down or turned off, the data it supplies to its clients is not usable., but most consider this a good choice.

These initial standards have led to others, usually known as RAID-N. So far, we have seen RAID-2, RAID-3, RAID-4, RAID-5, RAID-6, RAID-10, RAID-50, and RAID-60.

See my 23 page article, http://tjscott.net/storage/raid.explained.pdf , which describes RAID levels and their pros and cons in detail.

Storage Evolution and Standards: SCSI Arrives

Hard disk storage really started in 1957 when IBM introduced its 24-platter 5GB hard drive. In the mid-80's, fueled by the explosive growth of PC's, the move toward PC based networking gained pace. Local Area Networks were getting cobbled together using products such as IBM's LAN Manager and Novell Netware. Part of that development was to be able to deliver a high performance data storage standard. The Small Computer Systems Interface (SCSI) was the solution. The first SCSI standard, SCSI-1, was introduced in 1986, when network storage technology was still in its infancy.

SCSI was slow to catch on. During the 80's, storage devices of any type were expensive. The additional performance benefits that SCSI offered over the standards of the time were often sacrificed for budgetary concerns. In modern storage systems, people are willing to pay the extra for a SCSI system and its benefits.

As 2000 unfolded, there are two competitors - SCSI and ATA (IDE). The ATA interface is dominant in desktop systems, which almost exclusively offer ATA hard disk support rather than SCSI. SCSI storage devices are more expensive than their IDE counterparts. In addition, a SCSI controller or host adaptor, is needed to support SCSI devices, which adds cost and complexity. The extra host adaptor raises potential installation and configuration issues. This has discouraged many from using SCSI as a solution for desktop systems.

SCSI drive systems are not only faster than ATA, they are also more efficient in terms of reading and writing. SCSI drives access the disk data in such a way that the physical drive mechanics don't have to work so hard to read or write data. This ultimately means that SCSI drives last longer than ATA drives, a fact indicated by higher Mean Time Between Failure (MTBF) figures for SCSI drives. ATA also has the built-in limitation of four devices, and no support for external devices. In other words, ATA can work fine for a desktop, but cannot provide a foundation for a storage area network.

SCSI Standards

Since its introduction in 1986, SCSI technology has been developed, enhanced, improved and reworked. Progress is a good thing, though the vast array of SCSI standards can be both confusing and intimidating. Even though the standards are clearly defined, there are so many of them it becomes difficult to remember what means what. Although there are many SCSI standards, there are two basic types which run throughout – narrow, which uses an 8-bit bus for data transfer, and wide, which uses a 16-bit bus. The SCSI waters are further muddied by different types of signaling and termination.

SCSI-1 accommodated a maximum transfer speed of 5MB/sec, at 5Mhz on an 8-bit bus which means that it was a 'narrow' implementation. SCSI-1 was succeeded by SCSI-2, which was available in two versions: Fast SCSI-2 on an 8-bit bus, and Fast Wide SCSI-2, on the 16-bit bus. Then SCSI-3, the current standard, which has also defined a number of different “SCSI standards”. These include Ultra SCSI, Ultra Wide SCSI, Ultra2 SCSI, Wide Ultra2 SCSI.

In 2007, we have Ultra3 SCSI or Ultra160, which supports speeds up to 160MB/sec, and Ultra320, which supports 320MB/sec speeds. Ultra160 supports an 80Mhz clock speed and a 16-bit data bus. This is twice the data bandwidth and 16 times the bus speed of SCSI1. Ultra320 supports a maximum data transfer speed of 320Mbps.

For a complete rundown on the various characteristics of SCSI standards check out SCSI Trade Associate Website at http://www.scsita.org.

SCSI ID's and LUNs

SCSI devices are attached to cable to form a chain of devices. The cable, whatever physical form it takes, is referred to as the SCSI bus. Each device on the bus must be assigned a unique ID. These numeric ID's identify the device on the bus. Narrow SCSI standards allow for a maximum of 8 devices on the bus, while the wide standards can accommodate up to 16. Both internal and external devices must have a valid SCSI ID with no exceptions. It should also be noted that in addition to each storage device using a SCSI ID, the SCSI interface card, or host adapter, also uses one of the SCSI ID's. That means on a narrow SCSI bus there are 7 device ID's available, and 15 on a Wide SCSI bus.

Because the number of available SCSI ID's can be a limitation, it is possible to break a SCSI ID down further into what are called Logical Unit Numbers (LUNs). LUNs make it possible for a number of devices to share a single SCSI ID. A good example of where LUNs might come in useful is that of a CD-ROM jukebox where there are 5 CD-ROM devices in a single external housing. The housing itself can be assigned a single SCSI ID, while the devices in it adopt different LUNs. Another feature is called Sub Logical Unit Numbers (SLUNs). SLUNs allow each LUN to be split down into further devices.

SCSI Signaling

There are three SCSI signals that can be used. These are: High-voltage differential (HVD), Low-voltage differential (LVD) and Single Ended (SE).

HVD has been around since the earliest SCSI specifications and found popularity primarily due to the fact that HVD offers significant signal integrity, which allows for the utilization of longer cables without data loss or corruption. Ultra2 SCSI using HVD signaling can use cables up to 25 meters in length before the signal begins to corrupt. Ultra2 SCSI using LVD can only use 12-meter cables.

However, HVD signaling is not the method of choice. Why? Because HVD uses two wires for each signal to increase signal integrity, making HVD expensive to implement and at the same time increasing its power requirements. And second, in a practical situation, running a SCSI cable 25 meters is rarely necessary.

Like HVD, SE has been around since the early SCSI standards. SE sends data on a single wire and, as is far more susceptible to electronic interference. It also limits the cable lengths to 6 meters instead of 25 meters for HVD. Thus, SE signaling was cheaper and adequate for most SCSI implementations and was originally the most common signaling method.

SCSI technology is stagnant. As the SCSI reached speeds of 40 MHz and higher, SE couldn't cope, as its maximal length had to halve with each double in bus speed. Fast SCSI using SE, for example, had a cable length of 3 meters, and Ultra SCSI dropped the cable length to 1.5 meters. As a result, SE signaling is not defined for SCSI standards beyond Ultra SCSI.

LVD signaling is used for the faster SCSI systems. LVD uses two wires for each signal; however, the lower voltage requirements reduce cost and allow power requirements to be controlled. There is a bonus here: SE devices and LVD devices can coexist on the same SCSI bus. HVD, on the other hand, is completely incompatible with the other two signaling methods. But all is not wonderful! If not all devices on the SCSI bus are LVD capable, the bus will operate with SE signaling. Not only will the speeds be limited, but the cable lengths must also conform to the SE specifications. Today's modern implementations use LVD signaling.

SCSI Termination

Termination on the SCSI bus must be done properly, or the entire bus may have random problems. A SCSI bus can only have two termination points -- one at each end of the physical bus. The purpose of termination is to prevent the data signals that reach the end of the bus from reflecting back down the bus and then affecting other signals.

There are two basic types of termination, passive and active. Passive termination, the method used by the lower performance SCSI standards, uses a group of resistors to "soak up" the signals on the SCSI bus. A more advanced system, active termination, uses a group of voltage regulators to deal with the signals. This results in a more defined level of control, which is preferred and in some cases required on some of the higher performance SCSI standards.

SCSI termination is set on the host adapter for internal devices. With external devices, termination must be removed from the SCSI host adapter and placed on the last device in the chain.

SCSI Connectors

SCSI supports both internal and external devices with an array of connector options for each. The original regular density 50-pin SCSI connectors were designed to be used with SCSI-1 devices. The SCSI-2 standard uses a new high-density internal connector, which comes in a 50-pin and a 68-pin version. Today's SCSI implementations normally use the 68-pin internal connectors.

When it comes to external SCSI connectors, there are a range of possibilities which can be grouped into 2 distinct categories, D-shell connectors and Centronic connectors.

A D-shell connector was first used with the SCSI-1 standard, and its original 50-pin connector was replaced with the high-density shielded connectors in SCSI-2. The newer high-density D-shell connectors are available in both 50-pin and 68-pin versions. The latest external SCSI D-shell connector is referred to as the Very High Density Cable Interconnect (VHDCI) connector. The VHDCI connector is only available in a 68-pin version and is the one you are most likely going to encounter when working with modern SCSI implementations.

The Centronic connectors are 50-pin, and now SCSI Centronic connectors may still be found for connecting scanners or printers, but they have fallen out of favor for the more versatile VHDCI connectors.

SCSI Today

SCSI has done well in an industry where the only constant is change. SCSI's ability to adapt and improve help ensure that it is not destined to fade into the background. SCSI products and offerings have always been complex and confusing, and it looks like this will continue as SCSI moves into the future. For example, iSCSI will continue to make the world of SCSI one mired in compatibility issues and questions about capabilities and speeds.

In terms of application, SCSI is likely to remain a popular choice for connecting devices in small to medium sized storage applications. Even though other connectivity technologies like Fibre Channel and iSCSI will continue to chip away at SCSI in terms of market share, "traditional" SCSI is sure to be with us for some time yet.

Fibre Channel Disk Storage

Fibre channel is a technology used to interconnect storage devices allowing them to communicate at very high speeds (up to 10Gbps in future implementations). Fibre Channel is faster than SCSI and allows for devices to be connected as far away as six miles. Fiber Channel’s long distance connection capability allows devices in a SAN to be placed in the most appropriate physical location. See the document,

http://tjscott.net/storage/fibre.channel.org.review.pdf

for more information on fibre channel.

Network Attached Storage -- NAS

A modern NAS contains an enclosure for housing RAID drives, one or more Ethernet connections and a storage processor to arbitrate the storage requests. A modern NAS runs a minimalist operating system that supports TCP/IP and provides storage that looks like it exists on commonly used operating systems.

Modern NAS devices support both the CIFS and NFS file systems. Windows machines see NAS storage as normal Windows CIFS, giving them network shares or mapped drives. Linux/unix machines all have built-in support for NFS mappings, and see the NAS storage as another part of their file system. NAS devices are designed to be directly connected to IP networks, so they must be provided an IP address, subnet mask, default gateway, and DNS server address. NAS devices usually contain dual-ethernet adaptors, so they can provide network redundancy.

Network clients can access the NAS storage because they can be given the NAS's IP address, and will send requests for storage to that IP address. Thus NAS storage can be centralized for enabling security and for backing up the data.

Modern NAS devices are usually easy to expand, so if you need more storage, you can either add another NAS device or add drives in the NAS enclosure or in external enclosures. NAS devices provide fault tolerance for their shared storage, using various forms of RAID. This is quite different from the situation in a DAS environment, for in the DAS environment, when a server goes down, so does the storage it provides. NAS's are designed to run on your network as 7 by 24 appliances.

Storage Area Network (SAN)

A SAN is a network of storage devices that are connected to each other and to a server, or cluster of servers, which act as access points to the storage. In iSCSI configurations a SAN is also connected to your regular IP network. SAN's use special switches as a mechanism to connect the devices and to provide connectivity points.

SANs enable devices on different networks to communicate with each other, which offers significant advantages. With a properly configured SAN, you can back up every piece of data on your network without having to 'pollute' the standard network infrastructure with gigabytes of data. Providing this type of backup is one of the tremendous advantages of SANs.

Other developments, such as iSCSI, are available now. iSCSI is a technology that allows data to be transported to and from storage devices over an IP network. iSCSI systems serialize the data from a SCSI connection. Using iSCSI, the concept of network storage can be taken anywhere that IP can go, which is anywhere on the planet. Fibre Channel and iSCSI are a big factor in how fast people are able to afford and implement network storage solutions.

See the articles,

http://tjscott.net/storage/ms.san.final.pdf and

http://tjscott.net/storage/emc.invista.intro.pdf

for more information on SANs and SAN management.