Architecting the High Performance Storage Network

Jim Metzler
Ashton, Metzler & Associates

Table of Contents

1.0 Executive Summary

Driven by dramatic improvements in both performance and cost, networked storage is rapidly gaining momentum in the enterprise IT market. For example, companies that deploy a Storage Area Network (SAN) can expect to reduce the cost of storage by greater than fifty percent when compared with the cost of traditional storage solutions. However, the technologies that enable the efficient and effective deployment of storage are rapidly evolving. Enterprise IT organizations that deploy a SAN need to guarantee that they will continue to leverage the benefits of this deployment, even as the technologies change. In order to provide this guarantee, these organizations must develop a storage architecture. This architecture needs to properly position storage, computing, and the storage network. The architecture needs to also articulate best practices in SAN design.

This is the second paper in a three part series that is intended to address the key challenges facing the enterprise IT organization as they struggle to deploy SANs that will be able to both protect their initial investment and still leverage the ongoing improvements in storage related technologies. The first paper in the series provided background on both the movement to deploy SANs, as well as some of the key challenges relating to SAN deployment. This paper will provide insight into the existing and emerging best practices in SAN design. The third paper in the series will discuss some of the technological innovation that is impacting the deployment of SANs.

In order to understand the likely evolution of SANs, it is helpful to take a look back at how the LAN and the WAN evolved. That follows because many industry analysts, including Steve Duplessie of the Enterprise Storage Group, believe that the evolution of storage networking will bear some similarities to the evolution of the broader networking market. For example, in an article that was posted September 9, 2002 at Storage Networking World Online, Steve compares the integration of SAN islands to when we first linked LAN islands. The article also compares the current movement to deploy both the enterprise as well as the Wide-Area SAN, to when we deployed both the WAN and the enterprise WAN. Steve concludes the article by drawing a parallel between when we deployed a NOC in order to administer and monitor policies and best practices in the LAN/WAN, to what will happen relative to SAN operations.

This paper will specifically address the following questions:

1. What are the best practices in LAN design?
2. What are the key characteristics of a SAN architecture?
3. What issues are created by the Many to Few nature of a SAN?
4. What are the current best practices in SAN design?
5. What are the emerging best practices in SAN design?

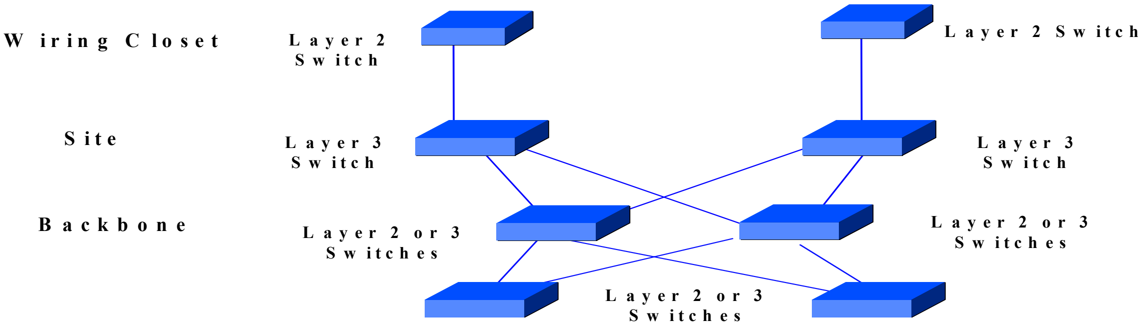
2.0 Best Practices in LAN Design

This section of the document will briefly describe how the best practices in LAN design have evolved over the last decade. The motivation for doing this is the realization that the design of a SAN is somewhat similar to the design of a LAN. As such, an understanding of how the best practices in LAN design have evolved can facilitate an understanding of how the best practices in SAN design need to evolve.

In the early to mid 1990s, LANs were built using hubs and routers. At that time, the best practices in LAN design included the use of highly over-subscribed links. For example, it was common in that time frame to have tens of devices share a single 10 Mbps Ethernet segment. To quantify the degree of over-subscription, assume that there were twenty devices attached to a 10 Mbps Ethernet segment, and further assume that each of these devices could transmit at 5 Mbps. If all of the end devices were all transmitting at the same time, they could generate an aggregate of 100 Mbps of traffic. Because of this, the 10 Mbps Ethernet segment is over-subscribed by a factor of 10 to 1 (10:1).

If the traffic on the highly over-subscribed LAN segment was light, this design did not negatively impact performance. However, when there was moderate to heavy traffic on the LAN, this design resulted in a badly congested, and hence a poorly performing network.

In the mid 1990s, LAN switches began to be deployed and changed how LANs were designed and implemented. With the use of LAN switches, it became common to implement large campus LANs in a hierarchical three-tiered architecture. As shown in Figure 1, this LAN architecture corresponds to the physical topology of wiring closet, site backbone and campus backbone. As will be discussed below, one key attribute of the LAN architecture that is depicted in Figure 1 is that different levels of functionality and intelligence reside in different layers of the network. In particular, the movement to what is called a "collapsed LAN backbone" resulted in the majority of intelligence and management control residing in the backbone switches.



Standard High Availability LAN Design
Figure 1

Note that the deployment of a hierarchical three-tiered LAN architecture did not necessarily eliminate over-subscription. For example, a company could deploy the LAN architecture depicted in Figure 1 with a wiring closet switch that had sixteen 10 Mbps Ethernet ports that are used to support end user devices, and a single 100 Mbps Ethernet port that is used to provide connectivity to a site switch. If each of the end devices could transmit at 10 Mbps, then in aggregate these devices could generate 160 Mbps of traffic. As such, the 100 Mbps link between the wiring closet switch and the site switch is over-subscribed by a factor of 1.6 to 1 (1.6:1).

As mentioned, each tier of the LAN that is depicted in Figure 1 is comprised of switches that are optimized for the required functions of that tier. For example, since the switches in the wiring closet tier have to attach to a very large number of end devices, they are optimized to provide a low port cost. In order to achieve a low port cost, these wiring closet switches typically do not posses much intelligence.

The switches that are positioned in the site tier connect a number of wiring closet switches. These switches are designed to provide high availability, support for multiple protocols (i.e., Ethernet and ATM), and somewhat more sophisticated network management than is available in the wiring closet switches.

The switches that are positioned in the backbone tier connect a number of site tier switches, which themselves connect several wiring closet switches. Because of their central position in the LAN, backbone switches are designed with more of an emphasis on functionality than on low port cost. For example, a typical LAN backbone switch has a sophisticated architecture that allows it to to provide the highest levels of availability, the most throughput, support for multiple protocols, the most sophisticated network management, the highest speed interfaces, sophisticated security, and the lowest latency.

3.0 SAN Architectural Principals

It is currently common practice for an enterprise IT organization to acquire a SAN through an OEM. It is also common to have the OEM implement the storage solution and leave. This can result in the enterprise IT organization not being able to leverage the investment they made in deploying the SAN.

However, a shift is underway in terms of how SANs get deployed. On a going forward basis, there is a strong movement to deploy not just SAN islands, but rather an enterprise wide SAN. This movement will place more emphasis on the networking component of storage. One result of this movement is that the SAN itself will become more of an independent entity than it was in previous storage deployments. As such, SANs will be chosen based on their ability to provide critical functionality, and not merely accepted as part of a broader solution.

Given the trend to focus on the SAN as an independent entity, enterprise IT organizations need to have a set of principles by which they evaluate alternative SAN architectures, in a manner analogous to how LAN architectures were evaluated in the 1990s. The first paper in this series described a number of the challenges facing enterprise IT organizations as they deploy a SAN. The challenges that were identified in the first paper in this series were used to create the set of SAN architecture principles that are depicted in Table 1.

For example, the first paper explained why management is a challenge to the successful deployment of a SAN. As such, having appropriate management functionality is a key principle that enterprise IT organizations should use to evaluate alternative SAN architectures.

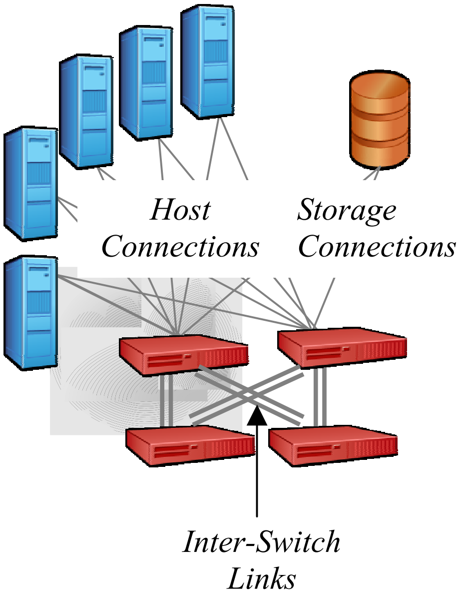
Architectural Principal	Characteristics
+ Scalability	• Have switch port density that can both accommodate as well as simplify growth • Support multi-switch topologies without sacrificing features or functionality • Provide agnostic transport for long-term backbone capability • Support the migration to high-bandwidth, wavelength-based storage networks
+ Performance	• Deliver throughput equal to the offered load • Maintain Latency Tolerances regardless of the traffic load • Create storage networks independent and exclusive of hosts and storage – Architect Storage Networks • Align bandwidth priorities with application requirements
+ Manageability	• Expose where the network affects application performance • Control how the network resources are allocated • Prescribe specific network behavior • Predict how the network will react to moves, adds & changes • Offer an intuitive network management interface
+ Availability	• Leverage a redundant, fault-tolerant H/A design • Utilize enterprise-class hardware and firmware maintenance • Implement a hierarchical network design for serviceability and uptime • Provide accountability for storage network resources as a commodity
+ Security	• Circuit-level intelligence for hardware enforcement • Open Linux Architecture (How this relates to security)

SAN Architectural Principals
Table 1

4.0 The Current Best Practices in SAN Design

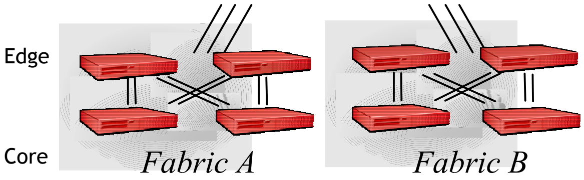
The purpose of this section is to summarize some of the current best practices in SAN design. The next section will summarize how some of these best practices need to change in order to better reflect the SAN architectural principals that were described in the preceding section.

As illustrated in Figure 2, SANs are designed around the principal of many hosts trying to gain access to relatively few storage devices. Today's best practices call for a SAN to have five times as many hosts as storage devices, and to have Interswitch Links (ISLs) that are over subscribed by a seven to one (7:1) ratio.



5:1 Ratio Servers-to-Storage
Figure 2

While it is possible to implement a SAN using a variety of topologies (i.e., star, bus, ring), the most common SAN topology today is the resilient dual fabric, core/edge topology as shown in Figure 3.



Standard Edge – Core Design
Figure 3

Figure 3 reflects how SANs are currently designed in order to provide for high availability. In particular, today's best practices in SAN design call for the use of dual fabrics. For added resilience, neither of the fabrics shown in Figure 3 would have a single point of failure. In addition, hosts and storage devices would be homed to separate fabrics, ensuring that the system would continue to function even if a given fabric was inoperative.

An important SAN design practice concerns locality. Locality refers to the degree that I/O is confined to a particular switch, or a set of switches. So, for example, if a host and a storage device that need to communicate with each other are located on the same switch, or segment of a fabric, they are said to have high locality. The converse is also true. If this host and storage devices are located on a different switch, or segment of a fabric, then they are said to have low locality.

Currently, locality is important because when hosts and storage devices are connected to the same switch (i.e., high locality) they should not experience any congestion. However, when hosts and storage devices are attached to different switches (i.e., low locality) that are themselves connected by ISLs, congestion is possible. Given that congestion leads to delay, today's best practices in SAN design call for the SAN designer to deploy hosts and storage devices in a way to maximize locality.

As described in the preceding paragraphs, locality has been and continues to be an important SAN design principle. However, an alternative approach that is sometimes used to design a SAN involves deploying a tiered architecture that is conceptually similar to the tiered LAN architecture that is depicted in Figure 1. Note that in a tiered SAN architecture, hosts and storage devices are seldom attached to the same switches.

Given the characteristics of the majority of core and edge switches that are currently on the market, today's best practices in SAN design calls for all devices to be attached to the edge switch and not to the core switch. The primary motivation for this design principle is to conserve the scarce ports on the current generation of core switches for use in connecting edge switches. As a point of reference, the first paper in this series quantifies how having a core switch with a limited number of ports severely limits the degree to which the SAN can scale.

5.0 The Emerging Best Practices in SAN Design

The set of best practices in SAN design that were described in the preceding section constitute the conventional, last 5 years of wisdom relative to the design of a SAN. These best practices were developed in response to the characteristics of the current generation of SAN switches. However, a new generation of SAN switches is being deployed. As was discussed in section 2, the deployment of LAN switches changed how LANs were designed. Analogously, the deployment of this new generation of SAN switches means that it is important to re-examine, and potentially modify, the existing set of best practices in SAN design.

One of the key areas of concern relative to storage networking is the deployment of SAN islands. In most cases, these SAN islands were the result of separate, non-coordinated initiatives to deploy storage for different applications. Typically these separate initiatives involved proprietary technologies from a number of suppliers. Because of the proprietary nature of the technologies, it was difficult to either interconnect or effectively manage these

SAN islands. As was alluded to earlier in this document, an emerging best practice in SAN design is to both avoid the deployment of SAN islands on a going forward basis, as well as to deploy a unified SAN across the entire enterprise.

As was described in the preceding section, today's typical SAN deployment involves five times as many hosts as storage devices. This type of design is typically referred to as being "Many to Few". In addition, these hosts and storage devices are all connected to a fabric using Interswitch Links (ISLs) that are over subscribed by a seven to one (7:1) ratio. This reliance on a high level of over subscription is very similar to the LAN design guidelines of the early 1990s. And, as was the case with those LANs, if the traffic on the SAN is light, this design approach will not negatively impact performance. However, when there is moderate to heavy traffic on the SAN, this approach to design will result in a badly congested, and hence a poorly performing SAN.

The combination of the Many to Few nature of SANs combined with high over-subscription of ISLs clearly results in the storage networks being a bottleneck to overall system performance. One emerging best practice in SAN design that can reduce this bottleneck is to not over-subscribe the ISLs to the same degree; i.e., use 2:1 or 3:1 over-subscription instead of 7:1 over-subscription. Another emerging best practice to reduce this bottleneck is to have the network enable intelligent control over the contended-for resources. This best practice will be described in the third paper in this series.

As was mentioned in the preceding section, one of the current best practices in SAN design calls for designing a SAN for high locality. While this intuitively makes sense, it can be quite difficult to actually implement. The difficulty comes from the implicit assumption that the SAN designer either knows which hosts will access which storage devices over time, or is willing to re-arrange the SAN as these requirements change. Another implication of the current SAN design is that it limits the number of systems that can access a particular storage resource. It also increases the cost of managing the SAN by having multiple points of configuration and management. Also SAN designers in this type of thinking must make trade-offs between servers/applications and their locality to a storage resource. This may cause an undesirable service-level that will later have to be trouble shot.

An emerging best practice in SAN design calls for the de-emphasis on designing for high locality, and the increased emphasis of the use of a tiered SAN architecture with some hosts and storage devices attached directly to the core switch. One of the key advantages of attaching hosts and storage devices directly to the core switch is the reduction in hop count that this produces. Note that this approach is analogous to implementing a collapsed backbone LAN.

A new class of switch, referred to as a backbone switch, is being deployed. This class of switch uniquely satisfies the SAN architectural principles that were detailed in section 3, and hence uniquely enables the emerging best practice in SAN design. In order to appropriately design a tiered SAN using this new class of switch, it is important to understand the key differences between this switch and the typical SAN edge or core switches. Those differences are depicted in Table 2.

Category of Switch	Availability	Manageability	Performance	Scalability
Edge	• Little if any redundancy • Little if any hot swap-ability	•Primarily CLI based interfaces •No control over the use of network resources •Manual ISL trunking	• Typically wire speed on all interfaces • Congestion is likely	• Support at most tens of interfaces • Effectively limited to Fiber Channel running over copper
Core	• Redundant, fault tolerant design • • Limited how swap-ability	• Nocontrol over the use of network resources • Both web and CLI based interfaces • Manual ISL trunking	• 20,000 Mb/second or higher • Typically wire speed on all interfaces • Guaranteed bandwidth services • Low, deterministic latency	• Support roughly a hundred interfaces • Plans to enable multiple protocols •
Backbone	• Extensive hot swap-ability	• Control over the use of Network Resources • Automatic ISL trunking	• 50,000 MB/second or higher	• Supports hundreds of interfaces • Support the migration to optical networks • Enables multiple protocols

Comparison of Switch Characteristics
Table 2

6.0 Summary

One of the challenges facing enterprise IT organizations is to ensure that they will continue to leverage the benefits of SAN deployment, even as the enabling technologies change. In order to provide this guarantee, enterprise IT organizations must implement a SAN architecture based on a fey key principles. Those principles are:

1. Scalability
2. Performance
3. Manageability
4. Availability

5. Security

The current set of best practices in the design of a SAN was developed based on the characteristics of the current generation of SAN switches. However, a new generation of SAN switches is being deployed. In conjunction with the architectural principles listed above, this new generation SAN Backbone switch is driving some changes in how to best design a SAN.

The primary changes that need to be made to the current set of best practices in SAN design include:

1. Cost effectively reduce the level of over-subscription from 7:1 down to around 2:1
2. Implement the ability to have the network intelligently control access to scarce resources
3. Implement a tiered architecture
4. Connect some mission critical hosts and storage devices directly to the next generation of core switch