Storage Consolidation:
Simplifying Storage Infrastructures

White Paper

Tim Rasmussen
Senior Product Manager

7-10-02

Table of Contents

Abstract

Growth in business information

Information is critical to the success of nearly every kind of business, and the rate at which information is being generated is increasing, in some cases, exponentially. Of course, this information unto itself is of no value if it cannot be accessed and used to benefit the business. Business information must be made readily available to the appropriate users and applications when and where it's needed; central to that use scenario is the storage infrastructure upon which the information resides.

Legacy storage architectures

Businesses are finding that legacy storage architectures *(direct-attached server-to-storage)* do not meet the needs of larger and more demanding storage requirements. With capacity requirements alone growing at rates faster than IT resources can manage them, what must businesses do?

Storage consolidation architectures

The answer to the limitations posed by direct-attached storage architectures lies in leveraging new technologies that enable designing scalable storage consolidation architectures for information-rich businesses. In short, storage consolidation is about building storage infrastructures that meet the demands of information-rich business environments of today and scale easily for those same businesses tomorrow.

What this paper discusses

This paper discusses storage consolidation, its background, benefits, and the various architectures that are currently available. In addition, it previews some emerging technologies that could serve to augment the future functionality and capability of the chosen storage consolidation architecture.

Intended audience

This paper should first be of specific interest to those individuals who are looking for options that will solve their current storage dilemmas; and second, those C-level executives, IT professionals, and systems administrators seeking knowledge about storage consolidation.

Introduction

Access to stored information

Information means the world: Datalink chose this slogan because it highlights the importance of information to business success. Just storing the information is not enough; users and applications needing to interact with the information must be able to access their information reliably and in a timely manner.

Challenges of servers with direct-attached storage

Until recently, open systems servers such as Sun, HP, IBM, and NT used a strategy of direct-attached storage (DAS) to provide capacity to the applications running on each server. Typically, this consisted of one or more disk drives connected via a SCSI interface located internally to the server or connected externally in a stand-alone enclosure. This strategy worked fine when capacity requirements were relatively low. As the information age matured, however, not only has the need for capacity increased dramatically, but also new requirements for data sharing, high performance, high availability, and cost control have emerged, challenging the DAS strategy. Simply stated, direct-attached storage cannot meet those challenges, but a properly designed storage consolidation infrastructure can.

Questions about storage consolidation

This paper will answer these questions about storage consolidation:

1. What is storage consolidation?
2. Does it meet the demands posed by information-rich business environments, where data sharing and efficient, reliable, and cost effective access to mission-critical information are prerequisites?
3. What are the architectures?
4. Where does one start?
5. What are the emerging technologies that support storage consolidation?

Read on to get answers to those questions.

What is storage consolidation?

Definition

Storage consolidation is the concept of centralizing and sharing storage resources among numerous application servers. Storage consolidation architectures are designed to address the limitations associated with direct-attached storage (DAS) and solve the growing data demands dictated by information-rich business environments. Figure 1 shows an example of a typical direct-attached storage architecture with a one-to-one relationship between each server and its direct-attached disk storage subsystem.

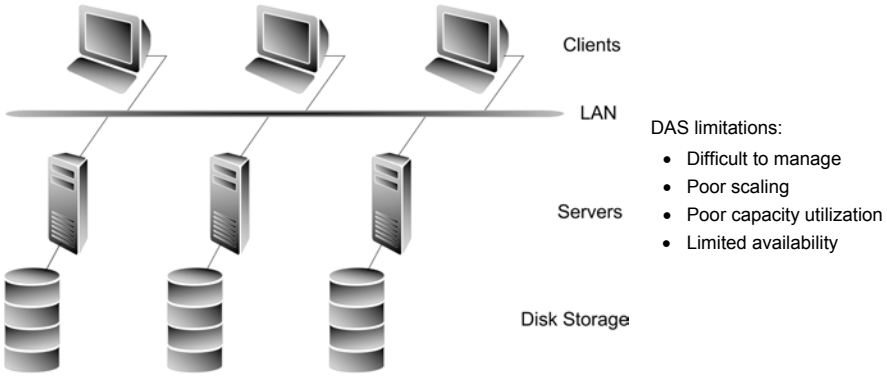

Figure 1: Direct-Attached Storage Architecture

Storage consolidation history

Storage consolidation is not a new concept. In open systems environments, storage consolidation was first introduced with RAID subsystems in the early 1990s. Although the primary objective of early RAID subsystems was to protect against data losses due to single disk drive failures, a byproduct was to pool multiple disk drives to create a single-logical disk drive consisting of several physical drives. As RAID subsystems evolved, they began to support heterogeneous server connections and provided high capacity, availability, and performance.

Continued on next page

Storage area networks

One of the most significant technological advancements in the storage industry is the relatively recent introduction of Storage Area Networks (SANs). SANs enable enterprise storage consolidation by providing high-speed connectivity between heterogeneous servers and a storage subsystem. SAN connectivity allows many servers to share capacity residing on a single storage subsystem, such as a RAID. Figure 2 is an example of a single-fabric SAN storage consolidation architecture. Although not a prerequisite, most enterprise SANs are based on fibre channel (FC) switches or directors. Other transport options will be discussed later in this paper.

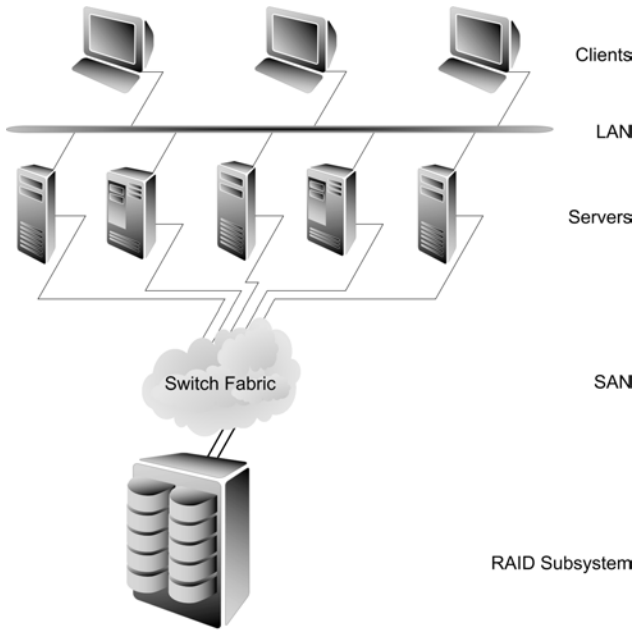

Figure 2: Single-Fabric SAN Storage Consolidation Architecture

Why invest in storage consolidation?

Storage consolidation simplifies the storage infrastructure

The primary reason businesses deploy storage consolidation is to simplify their current storage infrastructures. Massive data growth over the past several years has forced businesses to scale their storage infrastructures to accommodate capacity, performance, and high-availability requirements. Most businesses have reacted to those requirements by adding additional storage subsystems as needed. In traditional direct-attached storage environments, that reaction has resulted in numerous storage subsystems, often from several manufacturers. Given that scenario, customers inevitably experience storage infrastructures that are difficult to manage, ineffective at scaling, costly, and most importantly, do not keep pace with the growing quality of service (QoS) requirements posed by application servers. QoS requirements for servers include capacity, performance, and availability.

Storage consolidation key benefits

Storage consolidation architectures solve many of the problems businesses face in environments where data growth is outpacing the capabilities of traditional direct-attached storage solutions. The following are key benefits derived from storage consolidation relative to management efficiency, cost savings, and the appropriate level of QoS for application servers.

- Centralized management – a single interface can be used to manage the entire storage pool.
- Increased management efficiency – storage administrators can manage 10 to 20 times the capacity per person than previously possible in DAS environments.
- Storage investment spread across multiple servers – a properly designed storage consolidation architecture could support literally hundreds of heterogeneous servers simultaneously sharing the capacity of one or more storage subsystems.
- High capacity utilization – capacity can be allocated to specific servers on an as-needed basis, drawing from a pool of storage. Storage consolidation architectures can achieve upwards of 80 percent capacity utilization, compared to 25 to 30 percent utilization commonly found in DAS environments.
- Reduced administration costs – fewer storage administrators are required to manage a growing storage environment, resulting in direct cost savings when calculating the total cost of ownership (TCO).
- Meet quality of service requirements – Because storage consolidation architectures justify using industry-leading technologies, high availability, performance, and capacity can be designed into those architectures to ensure that application server requirements are met.

Storage consolidation architectures

Three storage consolidation architectures

Storage consolidation can be achieved via three specific architectures:

- Enterprise RAID storage consolidation
- SAN storage consolidation
- Network-attached storage (NAS) storage consolidation

Block and file level transfers

All three architectures use either block level or file level transfers between the server and the storage subsystem.

- Block level access, used by SANs and enterprise RAID storage consolidation architectures, requires the operating system of the application server to communicate directly with the storage subsystem.
- File level access, used by NAS devices to transfer data over a standard communications network, uses standardized protocols such as NFS *(Network File System)* or CIFS *(Common Internet File System)*.

Enterprise RAID storage consolidation architectures

An enterprise RAID storage consolidation architecture can support heterogeneous server environments with up to 64 direct-attached host interface connections. Figure 3 shows an example of an enterprise RAID storage consolidation architecture.

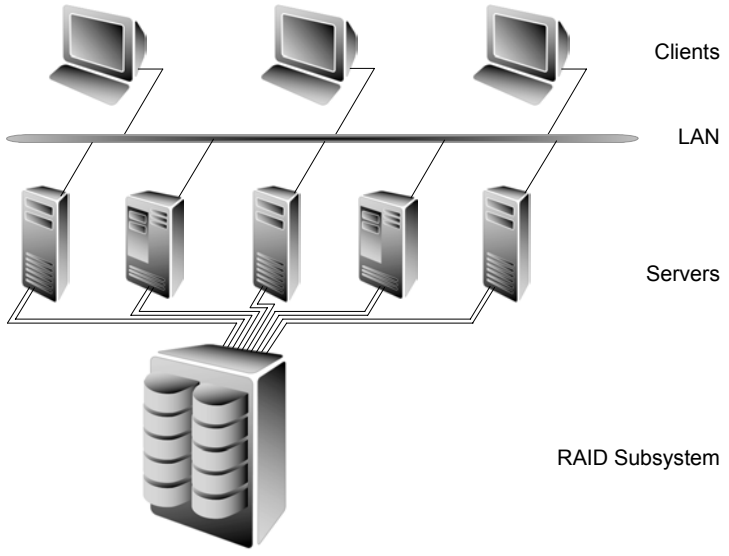

Figure 3: Enterprise RAID Storage Consolidation Architecture

Continued on next page

Storage consolidation architectures, *continued*

SAN storage consolidation architectures

SAN Storage Consolidation architectures leverage enterprise RAID subsystems by connecting them to a dedicated SAN, which is typically fibre channel based. With the RAID storage subsystem connected to the SAN, many more servers can use its capacity than the number of physical connections would normally allow. Figure 4 shows an example of a dual-fabric SAN storage consolidation architecture.

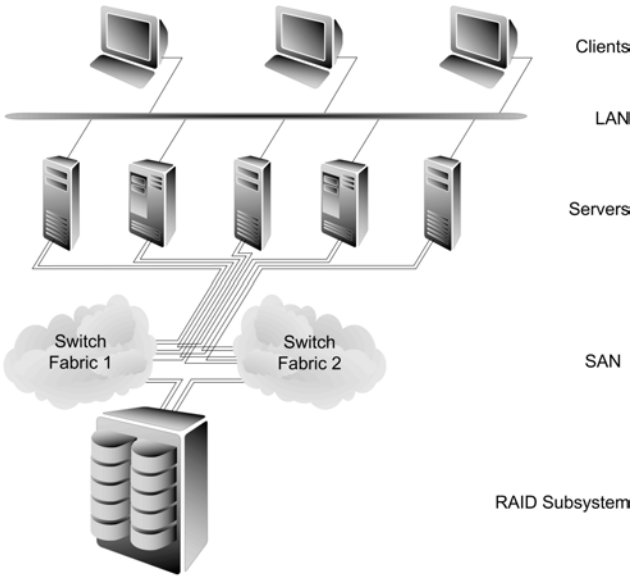

Figure 4: Dual-Fabric SAN Storage Consolidation Architecture

NAS storage consolidation

NAS storage consolidation architectures share many of the same benefits provided by enterprise RAID and SAN storage consolidation architectures, but use standard IP based communications networks to provide file level access to stored information. Figure 5 shows an example of a NAS storage consolidation architecture.

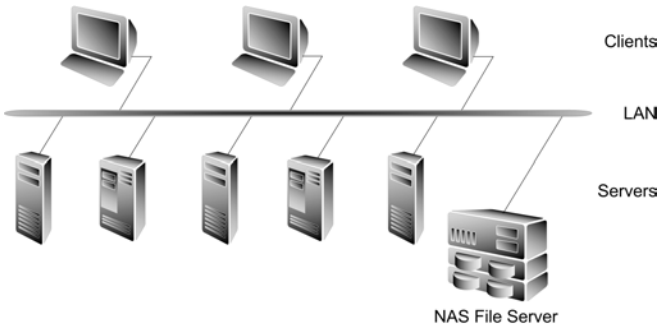

Figure 5: NAS Storage Consolidation Architecture

Continued on next page

Storage consolidation architectures, *continued*

Loss of access to information concern	One primary concern with storage consolidation is that all business mission-critical information could be stored on a single storage subsystem; in effect, putting all the "eggs in one basket." If that storage subsystem would fail, access to that information would be lost. This is an obvious and valid concern, and one that must be addressed.
How storage consolidation addresses the access concern	Storage consolidation architectures address this issue by using high end, typically enterprise-class RAID storage subsystems. Enterprise RAID subsystems are designed for maximum uptime with redundancy built in to handle any type of physical failure. Maintenance functions such as upgrades in capacity, cache, firmware, etc., can be done without system downtime. In addition, redundant paths to each host ensure that access to information is not compromised due to a cable, switch, or host bus adapter (HBA) failure. With this level of availability built into the local consolidation infrastructure, the single point of failure becomes the entire storage infrastructure itself, if a major disaster occurs.
Disasters and unplanned downtime events	Natural disasters, fires, user errors, etc. that could potentially cause data loss or interrupt access to mission-critical information can be planned for via what Datalink calls Storage Continuance Planning. Under Storage Continuance Planning, availability architectures such as clustering, and recoverability architectures such as server-to-server replication, server-to-storage mirroring, and storage-to-storage replication are options to consider when addressing potential downtime events or disasters. Storage Continuance and its architectures, however, is a separate topic and outside the scope of this white paper.

Where to start

Implementing storage consolidation solutions

Implementing customized storage consolidation solutions can be a very complicated and time-consuming endeavor, given the numerous equipment manufacturers and platform providers all claiming to have the "perfect" solution with their numerous technology offerings. Few companies have the resources or time to research, design, implement, and support a large-scale enterprise-storage consolidation project.

Implementation options

Businesses interested in implementing a storage consolidation solution have the following options:

- Purchase through distribution – meaning doing the project with in-house resources, which requires researching all available technology options. It is possible to source the individual components for less money; however, any savings could be quickly lost due to a lengthy implementation, improper design, and support issues.
- Direct relationship with a manufacturer – this can include a storage or server manufacturer. Interoperability and support are usually not a problem, however, the components comprising the architecture are usually limited to only what that manufacturer produces. In the case of server manufacturers, their expertise, outside of their specific platform, tends to be limited.
- Storage architect – as an independent provider of networked storage infrastructures, Datalink's expertise is building storage infrastructures based on business needs, using the best technologies the storage industry has to offer. Datalink is involved throughout the lifecycle of the storage consolidation project, from analysis and design through implementation and support.

Implementation outcomes

When the storage consolidation project is completed, the storage infrastructure should deliver the following:

- Meet current and future business needs
- Scale easily and non-disruptively
- Provide ease of management
- Fit the budget

Datalink, as an independent storage architect, has the resources and expertise to help your business achieve those outcomes.

Emerging technologies

Emerging technologies

New technologies are emerging that may provide additional functionality and cost justification for storage consolidation infrastructures. Among these new technologies, four are worth noting:

- iSCSI protocol
- Virtualization
- Storage resource management (SRM) software
- SAN/NAS convergence products

iSCSI, what it is and how it works

iSCSI is a protocol that offers another connectivity option to connect application servers to the storage consolidation infrastructure. This technology works by loading a SCSI intercept driver on the host, which captures a SCSI command and encapsulates it into IP packets. The SCSI command is then routed through the IP communications network rather than a dedicated physical cable, as with fibre channel. The main benefit of this is servers that could not otherwise connect to the storage subsystem *(due to distance limitations or cost to connect)* can now share the storage subsystem. In addition, no new technologies are necessary because the interface uses a standard Ethernet connection. Figure 6 shows an example SAN storage consolidation architecture with an iSCSI router, which provides direct block-level transfer to client servers outside the fibre channel SAN.

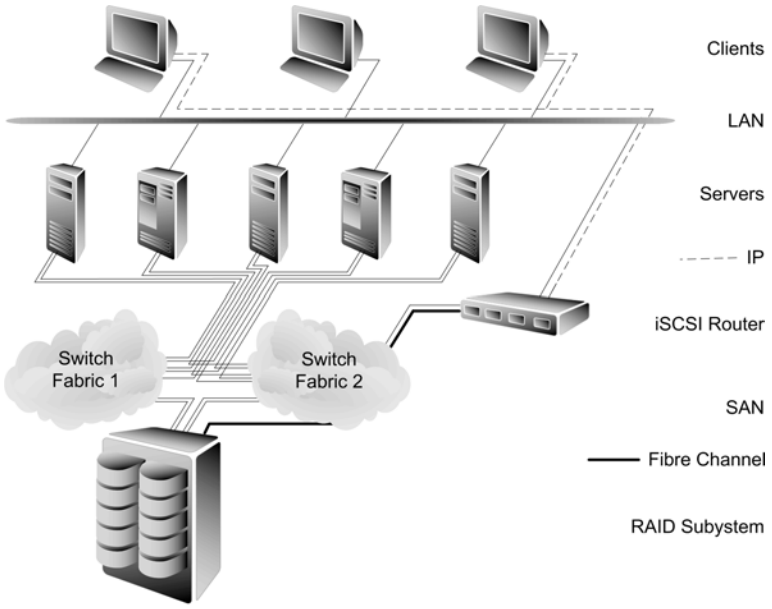

Figure 6: iSCSI Connectivity to SAN Storage Consolidation Architecture

Continued on next page

Emerging technologies, *continued*

Drawbacks with iSCSI technology

The drawbacks with iSCSI technology are poor performance relative to fibre channel, new technology, and limited support. Expect to see performance advancements in iSCSI technology in the coming year. Dedicated TCP/IP offload engine (TO) network interface cards will address many of the performance issues. Once the iSCSI protocol becomes standardized, large-storage manufacturers will most likely begin offering iSCSI interfaces directly on their storage subsystems, eliminating the need for routers to convert from fibre channel to IP.

Virtualization, what it is and how it works

From a marketing perspective, virtualization is one of the most liberally used terms in the storage industry—it is defined differently by different manufacturers. A recent trend from many manufacturers, offering virtualization architectures, is to position their products as network storage infrastructures based on virtualization technology. Fundamentally, virtualization can be defined as creating a logical representation of physical devices, a so-called abstraction layer. This means that physical disk drives or RAID subsystems can be logically combined to form one large pool of storage. By moving the intelligence from the storage subsystem to the SAN, many powerful capabilities can be leveraged, such as:

- Mirroring across heterogeneous storage
- Remote replication to heterogeneous storage
- Utilizing legacy storage subsystems in SANs
- Leveraging less expensive storage subsystems
- Increased management efficiency
- Dynamic LUN expansion and assignment

Virtualization benefits

One of the main benefits of virtualization is increased management efficiency because all storage capacity is managed by one common interface. As a corollary, routine storage administration functions such as assigning LUNs to hosts become much easier and less time consuming. In addition, other management capabilities such as performance monitoring and device managing are being incorporated to further improve management efficiency.

Continued on next page

Emerging technologies, *continued*

Storage resource management

Storage Resource Management (SRM), although not a new technology, is becoming common. As the amount of stored information continues to grow, just adding capacity is not always the answer. Understanding and managing the actual data ensures that only appropriate information resides on the storage consolidation infrastructure. Reducing the amount of redundant, unused, or unnecessary information can dramatically reduce the amount of capacity required, which means less to manage.

SRM categories

SRM solutions traditionally have been categorized into two groups.

- The first looks at the storage environment from a file level perspective. In this category, software agents are loaded onto all servers where the administrator wants to collect data. A central-management server gathers information from each client agent and performs monitoring functions. Some SRM packages incorporate automated policy-based management, enabling actions such as deleting or migrating files, matching predefined characteristics.
- The second type of SRM software focuses on the storage infrastructure and deals with hardware components such as switches, HBAs, and storage subsystems. This type of software monitors and manages the infrastructure, but does not necessarily deal with the type of information being stored.

Future SRM software products

As SRM software evolves, solutions incorporating both types of functionality should emerge allowing storage administrators to not only monitor and manage their hardware components, but also to ensure that the information stored meets the company's standards for storing data.

Continued on next page

Emerging technologies, *continued*

SAN/NAS convergence technology

Until recently, networked storage consolidation architectures fell into one of two categories, SAN or NAS. Due to the wide range of application requirements, many customers chose to implement both architectures. As both technologies evolve, architectures are becoming available that incorporate both SAN and NAS functionality. Figure 7 shows an example of a SAN storage consolidation architecture with a NAS appliance sharing the RAID subsystem.

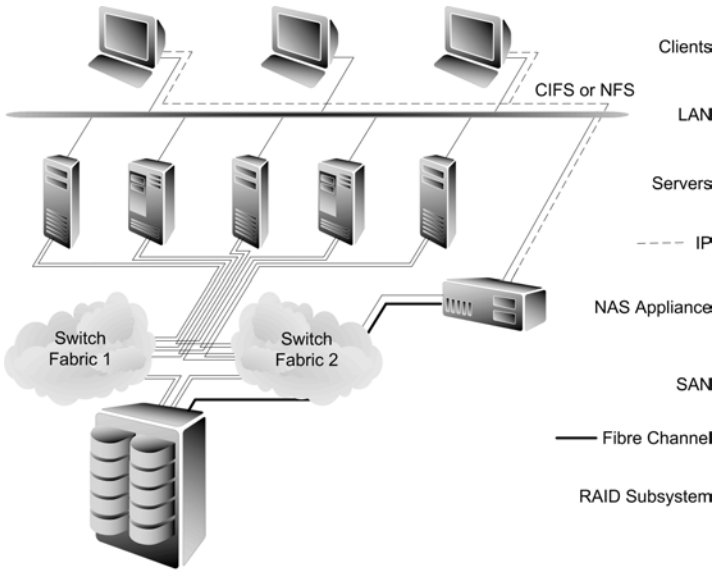

Figure 7: SAN/NAS Convergence Storage Consolidation Architecture

SAN/NAS convergence advantage

The advantage of this architecture is that storage administrators now have complete flexibility to provide capacity to application servers and clients based on the most beneficial file access method, block or file level. Either choice will leverage the single storage consolidation architecture, contributing to very efficient management.

Some enterprise RAID subsystems are beginning to incorporate built-in NAS appliance functionality, providing the choice of file or block level access. This design provides further efficiency and ease of management.

Summary

The information growth challenge

With information being the lifeblood of businesses today, numerous applications are used to enhance the ability of businesses to develop, design, produce, and deliver core products and services to market faster and more effectively than the competition. As a result, information generated by applications such as data warehousing, data mining, decision support, CRM, ERP, web servers, databases, and email is considered mission critical. As the massive volume of information continues to grow, businesses need answers to the query, what can be done to effectively manage this growth?

Adding IT resources or technology to answer the challenge

One may tend to believe that continuously adding IT resources is the answer: According to International Data Corporation (IDC), the amount of capacity managed per storage administrator must increase 60% per year just to keep pace with the growth at which data is being generated. In addition, if there were enough storage administrators in the marketplace to hire, budgets could not be expanded to support that option. Others may believe that throwing technology at the problem is the answer; nothing could be farther from the truth. Randomly adding technology to address current storage problems without assessing future business needs is not only inefficient, but also not cost effective.

Storage consolidation to answer the challenge

Datalink believes that the answer to the continuously growing information challenge is storage consolidation. Storage consolidation can provide storage infrastructure optimization using three architectures: Enterprise RAID, SAN, and NAS. Storage consolidation architectures, utilizing current and emerging technologies combined with appropriate storage management tools, can increase the amount of storage managed per administrator from 10 to 20 times. In addition to addressing growing physical capacity needs, storage consolidation architectures also provide high performance and high availability to meet the quality of service requirements of all connected application servers; Therefore, the choice of architecture should be based on current and future business requirements.

Importance of partnerships

A partnership with Datalink will ensure that a properly designed and implemented storage consolidation infrastructure provides the flexibility, scalability, and efficiencies necessary to cost effectively meet the present and future information storage needs of your business.

Terms and definitions

Term	Definition
Open System	Any system that is based on publicly available standards for subsystem interaction that facilitates multi-vendor and multi-technology integration. Open Systems are portable, scaleable, and interoperable.
DAS	Direct-Attached Storage subsystems that are directly attached to servers.
SCSI	Small Computer System Interface, pronounced "scuzzy." SCSI is a hardware interface that allows for the connection of up to seven or 15 peripheral devices to a single expansion board that plugs into the computer called a SCSI host adapter or SCSI controller.
LAN	Local Area Network is a communications network that serves users within a confined geographical area. It is made up of servers, workstations, a network operating system, and a communications link.
RAID	Redundant Array of Independent Disks is a disk subsystem that provides increased performance and/or fault tolerance. Performance is improved by disk striping, which interleaves bytes or groups of bytes across multiple drives, so more than one disk is reading and writing simultaneously. Mirroring or parity achieves fault tolerance.
FC	Fibre Channel is a high-speed transmission technology that can be used as a front-end communications network, a back-end storage network, or both at the same time. Fibre Channel is a driving force in storage area networks (SANs) arena for connecting multiple hosts to dedicated storage systems. With Fibre Channel, the hosts can not only talk to the storage system via SCSI, but the hosts can talk to each other via IP over the same network. Fibre Channel supports existing peripheral interfaces and communications protocols, including SCSI, IPI, HIPPI and IP.
SAN	Storage Area Network is a back-end network connecting storage devices via peripheral channels such as SCSI, SSA, ESCON and FC. There are two ways of implementing SANs: centralized and decentralized.
NAS	Network Attached Storage is a specialized file server that connects to the network. It uses traditional LAN protocols such as Ethernet and TCP/IP and processes only file I/O requests such as NFS (UNIX) and SMB (DOS/Windows).
HBA	Host Bus Adapter is a card that handles the server's CPU I/O operations, leaving CPU cycles to the server for application processing. The ability of a HBA to manage the I/O operations *(freeing the server from that task)* is a persuasive and valuable advantage of this technology.
iSCSI	Internet SCSI protocol that characterizes methods of enabling block storage applications over TCP/IP networks.

Continued on next page

Terms and definitions, *continued*

Term	Definition
IP	Internet Protocol, the IP part of the TCP/IP communications protocol. IP implements the network layer *(layer 3)* of the protocol, which contains a network address and is used to route a message to a different network or subnet work.
LUN	Logical Unit Number is the physical number of a device in a daisy chain of drives.

Datalink Corporation
8170 Upland Circle
Chanhassen, MN 55317
Phone: 1.800.448.6314
www.datalink.com