

## Trends in Enterprise Storage Technology

Abhinav Joshi & William J. Bender  
Project Performance Corporation  
2121 Crystal Drive, Suite 701  
Arlington, VA 22202  
703-920-0033

### Overview

Growing needs in both enterprise data storage and data protection have accelerated the enterprise storage spending worldwide. The value of enterprise storage lies in the intellectual property contained therein. When data is stored and organized in a way that enables an organization to synthesize new information, it drives all the major business processes, applications and ultimately revenue. Business critical applications such as Enterprise Resource Planning, Supply Chain Management, Customer Relationship Management, Knowledge Management, and Inventory Management add to the bottom line by using knowledge stored in the company's data. Companies are moving from Direct Attached Storage (DAS) to networked storage with the adoption of Fiber Channel (FC) Storage Area Network (SAN) technologies. The major benefits associated with this move are: higher availability, scalability, minimal interference with LAN traffic, increased management efficiency and utilization levels of about 90%, resulting in lower Total Cost of Ownership (TCO) and higher Return on Investment (ROI). As the economy improves from the recession of 2000, and the demand for storing increasingly large volumes of data, companies are beginning to spend on IT infrastructure and storage.

In this paper we discuss the current and the emerging technology trends in enterprise storage and backup.

### Introduction

Prior to 9/11/01, many organizations placed storage and backups at the bottom of the IT priority list. Only organizations in the World Trade Center that had robust data backup systems in place managed to survive. Today, storage, backup, and disaster recovery is a high priority for every organization. Of the two levels of storage, i.e. live storage and backup, backup is increasingly becoming more and more critical.

The hardware components for any infrastructure system are broken down into several subsystems:

- Client systems to be backed up and requiring restore services

- Server systems where databases and data warehouses reside
- Mainframes and backup servers for mirroring, storing and caching data
- Storage devices - either onsite or offsite
- Multi-platform network environment: Windows NT/2000, Windows 95/98/Me, OpenVMS, NetWare, Tandem, Linux, UNIX platforms etc.

## Current Trends in Technology

Storage technology has seen evolution in all aspects, ranging from interfaces and transport protocols to the way the data is served in today's networked environment. The most commonly used storage and backup systems today are:

- **Tape arrays and libraries:** These are the most widely used solutions when it comes to data backup. Tape arrays are the entry-level solutions whereas tape libraries form the higher end in tape-based backup. Tape arrays and tape libraries use the same kind of data backup technology and the only difference is in their capacities. Tape arrays are usually used for storage in the case of special applications or for subsystems such as a section in an enterprise. Tape libraries are usually used for backing up large volumes of data such as data in a storage network. They are the most popular solutions for larger application. Tapes offer the following advantages:
  1. Tapes offer the lowest costs per gigabyte of storage
  2. Current tape technology is extremely reliable
  3. Tapes can take overwrites amounting up to thousands of load/unload cycles
  4. Tapes are easy to use, store and manage
  5. The tape backup automation system is easy to use
  6. Hardware from multiple vendors is easily integrated in an existing system
- **Disk Based Backup System:** Disk based data protection technology is commonly used in addition to tape libraries. The data is immediately sent to a less expensive disk and subsequently copied to tape and sent offsite. This reduces the backup window, facilitates faster data recovery, and minimizes cost associated with distributed backups. Advantages of this approach include:
  1. Immediately offsite replication of critical data
  2. Recovery can be performed much faster from a disc copy
  3. Better reliability and consistency
  4. Longer data lifespan

## 5. Interoperability

- **Optical and magneto optical disk arrays:** Optical disk based backup solutions still have to make a major impact in the data backup arena. These usually consist of read/write optical disk libraries used to store data. Most of these solutions, like optical and magneto optic jukeboxes, are usually limited to specific areas like medical and legal applications.
- **Hierarchical Storage Management (HSM):** This kind of backup hierarchy makes use of many types of storage devices. It is basically a policy-based management of file backup and archiving. In a HSM, the user is not aware if the data is coming from a slower device like a tape array or from a faster device. HSMs usually make use of a combination of different storage devices such as RAID devices, tape, optical disk arrays, etc. At this point, it is important to stress the fact that HSM is more of a data management strategy than a backup option.
- **Direct Attached Storage (DAS):** It is the most popular form of storage for small organizations because of its simplicity and low initial investment. The term DAS includes all forms of storage directly attached to the server. This includes individual hard disks as well as external JBOD. DAS will show low or negative growth over the next few years because it is difficult to scale, creates isolated pools of disk resource, and is relatively difficult to manage from a centralized control station.
- **Network Attached Storage (NAS):** NAS provides access to storage over the network. NAS devices contain a thin server that provides file services to other hosts on the LAN using network file access methods such as CIFS or NFS. NAS will see the maximum growth in the next few years. NAS devices work very well in heterogeneous environments where network clients may be PCs, Macs or Unix workstations. Most NAS devices will offer file services over the Windows-centric SMB (Server Message Block) and CIFS (Common Internet File System) protocol, the Unix favorite NFS (Network File System), or the near-universal HTTP. NAS devices have seen much use in Web application environments, in which presentation, logic, and storage elements have evolved to be clearly differentiated. They are often used as storage for email and images, and other static data to be served over the Web. Web-based email providers are big users of NAS devices. NAS devices have also recently seen use as database

storage. With new emerging protocols such as DAFS (Direct Access File System), traditionally I/O intensive applications like databases are also moving to NAS. NAS is directly connected to LAN and hence consumes large amounts of LAN bandwidth. In addition, special backup methods need to be used for backup and disaster recovery.

- **Storage Area Network (SAN):** Storage Area Network is expected to account for a large percentage of installed storage capacity within a few years. Currently SAN solutions are considered to be very expensive, very specialized systems that cost large amounts of money to implement and maintain. However, vendors are working on significantly reducing their costs. It provides plug-and-play scalability as well as high performance, and the ability to share tape and disk resources between multiple servers. Managers can respond to growing storage needs by simply adding more arrays or tape libraries. SAN is advantageous in terms of performance and utilization.

## Emerging Technology Trends

Recently, the storage industry has seen some major developments in the field of Tiered Storage, Storage Over IP, NAS/SAN Gateways, Storage Resource Management, Storage Virtualization, and Serial ATA drives. We will discuss each one of these in details, as follows:

- **Tiered Storage:** The fundamentals of tiered storage are to establish multiple classes of storage with different tradeoffs in terms of costs, performance, and availability in order to optimize the placement of data. The first tier represents online data, which is most valuable to an enterprise on a daily basis. It is stored on primary storage, where speed, performance and reliability are greatest. In tier two, the end user is willing to sacrifice some performance and availability in order to benefit from a lower cost per terabyte. Typical examples of tier-two storage include low-cost NAS, storage devices with inexpensive ATA disks or near-online optical storage libraries. Finally, in tier three data that is retained for long periods of time, data is stored on off-line tape media, which has the lowest price per terabyte and the response times are also usually the slowest. The big challenge associated with tiered storage is the automated data migration to the various storage classes.
- **Storage Over IP (SoIP):** SoIP extends the scope of storage resources beyond the storage network, across LAN, MAN or WAN. Storage over IP is a framework for the integration of legacy and future storage devices in a

universal IP-based network. A key benefit of SoIP is that it does not require complete restructuring of the existing storage subsystems. Legacy SCSI and Fiber Channel devices can be integrated into the IP-based storage network. NAS (Network Attached Storage) is an excellent example of the benefits associated with Storage over IP. Protocols like iSCSI (Internet SCSI) enable block-level access to storage devices, not just high-level access as with NAS devices using protocols like CIFS and NFS. Block-level access via SoIP protocols enables continued use to the existing backup and disaster recovery applications, as well as to exploit new concepts like storage virtualization.

The three primary SoIP protocols are iSCSI, iFCP (Internet Fiber Channel Protocol) and FCIP (Fiber Channel Over IP). Although all three differ greatly, they provide a common basic service, i.e. block-level access to storage subsystems over an IP network. This gives them all of the advantages of scalability, availability, ease of management and leverage of legacy systems.

**- iSCSI**

iSCSI protocol allows for block-level storage over common IP networks. It takes advantage of the existing IP network infrastructure and IP skill set. It offers a connectivity option to connect application servers to the storage consolidation infrastructure, like Fiber Channel. It loads a SCSI intercept driver on the host, which captures a SCSI command and put it into IP packets.

**- FCIP**

FCIP concentrates purely on Fiber Channel SANs and groups of devices across the enterprise. It is for an enterprise that uses multiple small Fiber Channel SANs in different locations. These would then be disconnected from each other, that is, not part of a single storage network. FCIP is a protocol that brings these Fiber Channel SANs together under one roof across IP networks.

**- iFCP**

Like FCIP, iFCP deals with the interconnection of Fiber Channel devices and SANs.

With iFCP, the lower-level

Fiber Channel transport is replaced with Gigabit Ethernet and TCP/IP.

*Fiber Channels started as a way to improve beyond the current 10 Mbps Ethernet standard. It has evolved to be the default standard for storage networking, interconnecting data storage devices and servers where the high speed of access and transfer are of high importance.*

Similar to FCIP, iFCP also relies on gateways for SAN-to-SAN, device-to-SAN and device-to-device connections. Much like FCIP, iFCP allows applications that are written for native Fiber Channel SANs to work across an IP network. However, unlike FCIP, which tunnels the FC protocol in TCP/IP, iFCP actually uses the lower layers of TCP/IP and Gigabit Ethernet as transport layers.

The most important benefit associated with SoIP protocols is the linking of geographical boundaries of storage networks. iSCSI, FCIP and iFCP have their own application areas, with iSCSI being able to bring disparate devices under one umbrella and FCIP and iFCP focusing on the interconnection of Fiber Channel SANs. iSCSI, having a broader horizon of applications, will probably see more implementations with software and hardware becoming available. These protocols enable NAS devices to be able to offer storage services over IP anywhere. This is the kind of access required for storage management applications, disaster recovery, and remote backup systems. FCIP, for example, makes the interconnections of SANs transparent, which means that the applications need not even know that they are running over an IP network. Some other benefits include reduced costs, improved scalability, and flexibility.

- **NAS and SAN Gateways:** The SAN/NAS gateways typically feature Ethernet connection on one side and FC connections on the other. This technology greatly simplifies the storage environment by clubbing both SAN and NAS functionality in the storage ecosystem. As both technologies evolve, system architectures are being developed that could club both SAN and NAS functionality in the storage ecosystem.
- **Storage Resource Management (SRM):** The long-term storage management industry trend is toward increasingly broad management, automation, and provisioning. SRM addresses storage inventory, and reporting has emerged as the baseline for advanced storage management. SRM is the building block upon which more capabilities must be added in order to control increasingly complex physical and local storage environments. Storage software vendors such as EMC, McData, Veritas, etc. are competing to capture the market by providing storage consolidation through storage resource management. Adding capacity to the storage system is not the only answer to the growing problems. Understanding and managing the actual data ensures that only relevant information is stored on the storage infrastructure. Reducing the amount of redundant, unused, or unnecessary information can dramatically reduce the required capacity.

- **Storage Virtualization:** Virtualization improves storage manageability by separating the presentation of storage to the server operating system from the actual physical devices. Virtualization consists of taking several physical storage devices and joining them to appear as one logical unit to manage. This represents a collective pool from which users can request any amount of disk space. The users will thus be able to access storage without knowing where a device is, and how it is configured. It is estimated that only 5 percent of the storage in large companies in the U.S. is Fiber Channel SAN, 95 percent is still directly attached and 90 percent of SANs are homogeneous (single server type, single storage vendor, single hub/switch vendor). It is believed that a storage virtualization strategy is paramount for users facing accelerated on-line growth.
- **Serial Advanced Technology Attachment (SATA):** SATA enables organizations to capitalize on inexpensive disk capacity to drive down cost. SATA specifies a new serial interface to the standard ATA drive, featuring parallel interface connected with a short, wide ribbon cable. SATA drives use the same internal components as ATA drives, but they also support hot swapping of drives. Arrays of SATA drives will become crucial to a variety of data backup and fast recovery techniques. Instead of backing up directly to tape, storage administrators can backup from expensive SCSI disk to affordable SATA disks, and then to tape for archival purposes.

The ultimate storage solution for most large organizations will probably be a mix of all technologies complementing each other. A central SAN with sufficient storage and backup capabilities will take care of an entire organization's storage needs. This SAN could be mirrored across the globe to another site via SoIP protocol to provide the ultimate redundancy. Access to the SAN would be via servers with dedicated FC adapters, pushing data while feeding results to workstations on the LAN. These storage solutions could have data stored centrally via NAS file sharing schema. A NAS gateway to the SAN would provide authenticated access to files via SMB or NFS.