# Geospatial Multistate Archive and Preservation Partnership (GeoMAPP)

## GeoMAPP Storage Primer


December 31, 2011

Version: 1.0

# GeoMAPP Storage Primer
# Contents

# GeoMAPP Storage Primer

Establishing a storage repository is an essential component in designing a geospatial archiving solution, as the archival organization's mission is to preserve, in most cases, indefinitely, the digital geospatial resources that are transferred into the care of the archival organization. As the GeoMAPP partners began exploring storage solutions for their digital repositories, we knew that the storage system needed to be:

- secure and stable,
- robust enough to potentially handle many numbers of files
- able to scale up to support an every-increasing collection of numerous and large datasets as they are added to the archival repository.

Starting at the beginning, the GeoMAPP partners had to educate themselves on terms and technologies in order to be able to be part of the conversation with Information Technology staff as the repositories were built and configured.

Storage alternatives continue to evolve, and offer a wide range of technological and architectural options in designing the archival organization's storage architecture. Utilizing the knowledge gained by the GeoMAPP partners, this paper provides a survey of several storage concepts and technologies and provides a general orientation for archivists looking to understand some basics about storage. It is not intended to recommend a particular storage solution.

**Note: in addition to the storage itself, the architecture also relies on different hardware components for connections and speed. An appendix detailing this piece is at the back of the document for your reference.**

## Evolution of Storage

Figure 1 offers an illustrative depiction of the evolution of storage:

1. Beginning with storage internally located in the computing device, and
2. Progressing to moving the storage device outside of the computing device enclosure but still directly connected to the computer (server), and
3. Next, transitioning to connecting the storage device to the network rather than a particular server - providing access to storage to users and applications across the network, no longer requiring them to be directly connected to a server hosting the storage, and
4. Further progressing to a dedicated network of storage devices.

While there has been an evolution towards more complex storage ecosystems, viability, utility and economic value of the earlier storage approaches has not been eliminated. Individual storage implementations for archival organizations may use any one of these storage alternatives, largely depending upon budget, IT staff resources, IT staff skill sets, and available physical space.

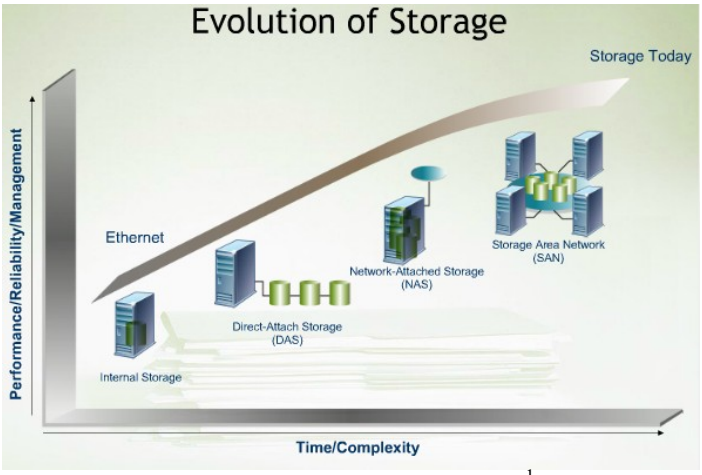



Figure 1: Evolution of Storage[1]

## Storage Categories and Accessibility

There are three general models of storage accessibility including online, nearline, and offline. Online generally provides the most immediate access to data (e.g. the hard disk attached to a computer (online) vs. accessing a file that requires pulling a tape out of the offsite storage (offline). On-line is generally considered the most expensive per storage capacity measure (e.g. $/byte). The access alternatives are generally evaluated based on the trade-off of access responsiveness and cost. The following summarizes the general characteristics of each of the storage accessibility options.

***Online***

- Storage that is immediately accessible to the end user.
- Usually allows for random access of data from the medium.
- Is usually implemented using disk-based technologies, or more recently solid-state storage technologies.
- Usually accessible to an end user directly through their computer as a disk or share drive.
- Usually used as the basis of day-to-day data access.

***Nearline***

- The user may need to wait for the media to be mounted into the drive device (e.g. optical drive or tape drive housed in a tape library or optical library, respectively).
- Optical media allows for random access, tape media requires sequential access.
- Is often implemented using optical-based technology, possibly in an auto-loader optical drive or tape library.
- Usually accessible to an end user through their computer.
- Retrieval time is usually slightly longer than on-line. Users will experience a brief delay in accessing the data (seconds to minutes).
- May be used to access important, infrequently accessed materials.
- Can hold significant amounts of data.[2]

---

[1] Siriam Ranganathachari. "Storage Primer." http://www.slideshare.net/sriramr/storage-primer-presentation

[2] For Example: (1) "In a typical autoloader or tape library, for example, an entry-level solution typically can access up to 24 physical cartridges, giving the data center nearline access of 38.4 TB compressed data." David Reine. "Making Use of Virtual Tape in a D2D2T Environment." Clipper Notes, May 28, 2008. http://www.clipper.com/research/TCG2008028.pdf
(2)"Our tape solution will consist of a scalable library with support for over 6,000 LTO cartridges *nearline,* i.e., within the library with a compressed capacity of up to 10PB (petabytes) and support for more tape drives than this application will every need." David Reine, Mike Kahn. "Disk and Tape Square Off Again - Tape Remains King of the Hill with LTO-4." Clipper. Feb. 13, 2008. http://www.spectralogic.com/index.cfm?fuseaction=home.displayFile&DocID=2118

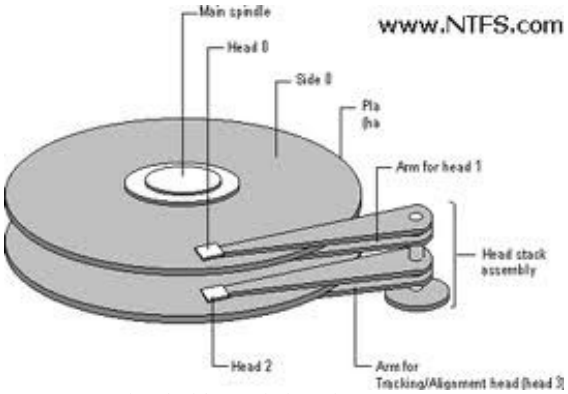
***Offline (typically)***

- Requires a longer wait time to access storage (minutes to hours to days).
- Medium is often tape-based, but could use any "portable" type of storage device.
- Access is sequential. (To retrieve a file on a tape, the tape must be read sequentially up to the point the file is located.)
- Requires heavy reliance on human resources to find and load the tape into the tape drive, retrieve the file from the tape, and deliver it to the end user.
- Often used as the model for backups of data assets; stored offsite; used as data disaster recovery method.

## Disk and Hard Drives

***Hard Disk***

The hard disk is often the primary computer storage medium that record and store data, as illustrated in Figure 2. It consists of a spindle of glass or aluminum platters coated with a magnetic coating, that record and store data, as illustrated in Figure 2. As the data is stored magnetically, it persists when power is turned off. The hard disk is usually permanently enclosed in a hard disk drive. Computers have an internal hard disk drive, or disks may be attached externally to a computing device , and external disk drives can be is enclosed in the hard drive, which reads and writes the data to the disk. Most modern disks are connected to the host system through either Serial ATA (SATA) or Serial Attached SCSI (SAS) technology.[3]




reprinted with permission of www.NTF.com

Figure 2: Parts of a Hard Drive[4]

***Solid State Disk (SSD)***

A solid state disk uses solid-state random access or flash memory rather than magnetic or optical mechanisms. It does not utilize moving parts to read/write data, and thus is quite durable. It also offers higher performance than rotating magnetic disks, as there is no mechanical seek and rotation time to acquire the data. However, it is more expensive per Gigabyte (GB) than traditional disk storage technologies.[5]

***Redundant Array of Inexpensive/Independent Disks (RAID)***

RAID is a drive mechanism that uses two or more magnetic disks to enhance performance and/or reliability (usually by the way of back up technology within the same system) that often appear to the user as a single disk device. [6] There are several RAID configuration variants, each providing different capabilities and levels of performance and/or data security.

---

[3] Matthew McKenzie. "57 Storage Terms You Need to Know." Information Week, SMB. June 2010. http://www.informationweek.com/whitepaper/Hardware/Peripherals/57-storage-terms-you-need-to-know-wp1278003128543

[4] NTFS.com."Hard Disk Drive Basics." File Storage Hardware and Disk Organization. of http://www.ntfs.com/hard-disk-basics.htm

[5] Matthew McKenzie. "57 Storage Terms You Need to Know." Information Week, SMB. June 2010. http://www.informationweek.com/whitepaper/Hardware/Peripherals/57-storage-terms-you-need-to-know-wp1278003128543

[6] Nick Gold. "Systems Storage Primer." http://chesapro.com/articles/articleID/4

**Note:** Industry best practices recommend that when procuring drives to build the RAID array that drives be selected from different production batches to minimize redundancy of production-related hardware defects.

RAID Levels:[7]

| | |
|---|---|
| RAID 0 | Combines two or more hard drives that are seen by the user as a single volume or drive. Uses striping to write data across the multiple drives, thus improving write performance,<br>Read performance is improved over a single drive through the availability of multiple read heads.<br>**Benefit:** Better performance vs. a single drive.<br>**Storage Space:** The total space available is the sum of the space on each of the drives.<br>**Risk:** If one of the drives is lost, the entire volume of data is lost. |
| RAID 1 | A pair of drives that are mirrored, so every data bit is written redundantly to both drives.<br>**Storage Space:** Requires a minimum of 2 drives. The total space available across all of the drives divided by 2.<br>**Security Benefit:** If one drive encounters a physical failure, the failed drive can be removed and replaced, and the data continues to be accessible from the mirror drive.<br>**Note:** RAID 1 is NOT considered a backup, it is simply a more robust storage medium. |
| RAID 3 | Data is striped across each of the drives, with each byte being written to a different drive, providing improved write/read performance. Includes a dedicated drive in the RAID set that stores parity data.<br>If any of the data bits are lost they can be recovered by comparing the data bits to the parity data.<br>**Storage Space:** Requires a minimum of 4 drives. The disk used for the parity data is not available for general data use.<br>**Security Benefit:** If one of the data drives is lost, the data can be restored from the parity data. If the parity drive is lost, it can be restored from the data. |
| RAID 5 | Data is striped across 3 or more drives, with each block being written to a different drive, providing improved write/read performance. The parity data is also distributed across the drives.<br>**Storage Space:** Requires a minimum of 3 drives. The space used to store the parity data will not be available for data.<br>**Security Benefit:** If one of the drives is lost, the data can be restored. The system will experience reduced performance until the time the drive and data are restored.<br>**Risk:** The data will be at risk until the failed drive is restored. |
| RAID 6 | Data is striped across each of the drives, with each block being written to a different drive. The parity is written redundantly and also distributed across the drives.<br>**Storage Space:** Requires a minimum of 4 drives. The space used to store two copies of the parity data will not be available for data storage.<br>**Security Benefit**: Supports the loss of two drives. Adds additional data security during the time that a failed drive is being rebuilt. |
| RAID 10 | Combines mirroring and striping to provide both performance and security. The drives are first set up as mirrored pairs (RAID 1), and then striped (RAID 0).<br>**Storage Space:** Requires a minimum of 4 drives. Space available is half the total drive space.<br>**Security Benefit:** provides the most security and performance. |
| RAID 50 | Combines distributed parity (RAID 5) with mirroring (RAID 0).<br>**Storage Space:** Best implemented across 2 RAID 5 arrays, with data striping across the array. Space available is half the total drive space less the space required for the parity data.<br>**Security Benefit:** can sustain one to four drive failures while maintaining the data integrity - if each failed disk is in a different RAID 5 array.<br>**Note:** This will be one of the more costly alternatives, but provides the highest level of performance and security. |

[7] ProWare. "RAID Levels." see http://www.iscsi-raid.com/raid0.html for nice animations on the different RAID levels.

***Optical Media***

This media refers to discs that are written and read by a laser (magnetic or phase change are the common imprinting technologies). Examples include CDs and DVDs, which come in a variety of access modes such as read-write (RW), and write-once read many (WORM). Optical media typically offers slower access times than hard disks.

Optical Media can exist as single instance or can be part of a disc library. An optical library is a "a set of optical volumes and the optical disk drives associated with those volumes. The volumes within an optical library are said to be library resident optical volumes." Typically, these are accessed via a Jukebox and typically, are considered as a nearline storage solution. , but the media is more durable as data storage is not based on a magnetic charge, and it is suggested to have a shelf-life of 30years, and can withstand a million rewrites.[8] Magneto-Optical drives originated by Sony, originally offered 650MB per 5 ¼" platter when introduced in 1988, and have progressed in capacity to 2001, and support 9.1 GB per platter. The WORM option is especially useful for archival applications, where it is important to ensure data authenticity and protect against deliberate or accidental overwrites.[9]

***Optical Disk Library***

IBM describes an optical library as "a set of optical volumes and the optical disk drives associated with those volumes. The volumes within an optical library are said to be library resident optical volumes. Optical volumes can also be located outside of the optical library. These volumes are referred to as shelf-resident optical volumes."[10] Optical disk libraries are often used as the basis of a nearline storage solution. As an example, Alan Freedman reported in 1998 that a 150-disk DVD-RAM jukebox could be used to support nearline storage with an 800 GB capacity based on 150 disk, and should cost about $15,000.[11] The high end Plasmon G-638 Optical Library supports 638 slots for media and up to 12 drives for reading and writing data. It has a potential capacity of 38TB using 60GB Ultra Density Optical (UDO-2) media, and specifies a 6.4 second media exchange time, and costs $152,275. Media ranges from $325/5-pack to $425/5-pack.[12]

***Just a Bunch of Disks (JBOD)***

JBOD is a collection of disk drives that can be joined together to look like a single volume or virtual disk.[13] It provides no additional security through data redundancy or parity bits, or performance benefits, such as striping, as is provided by RAID. It may be useful to join several smaller disks into a larger virtual disk. JBOD is sometimes referred to as "non-RAID."

## Storage Constructs

There are a few basic disk configuration or disk formatting constructs that might be helpful to understand when evaluating storage solutions.

***Block***

A "block" refers to the file system construct representing the smallest allocatable section of space in a file system. There is a performance/usage trade-off in constructing file systems to use block size configurations. Using larger block sizes can improve disk access performance, as it reduces the number of block items that must be retrieved thus requiring less processing by the system. This is especially useful if dealing with large files. However, if the system is configured or used to write small files to the disk the remainder of the block not used is unavailable and becomes lost space.

***File System***

The file system provides the organizing and management basis for storing and accessing files stored on a disk device. The file system is usually tightly coupled with the operating system, which may provide additional file

---

[8] Alan Freedman. "A Storage Primer." 1998. http://www.govtech.com/magazines/gt/A-Storage-Primer.html?page=3

[9] Data Archive Corp. "About M-O (Magneto-Optical). http://www.dataarchivecorp.com/about_magneto_optical.htm

[10] IBM. "Optical Libraries." http://publib.boulder.ibm.com/infocenter/zos/v1r11/index.jsp?topic=/com.ibm.zos.r11.idao200/o2030.htm

[11] Alan Freedman. "A Storage Primer." 1998.

[12] Data Archive Corp. "Plasmon G-638 Optical Library. http://www.dataarchivecorp.com/udo-plasmon_g-638.htm

[13] "No-RAID drive architectures." Wikipedia. http://en.wikipedia.org/wiki/Non-RAID_drive_architectures

management services such as ownership, access control, and file manipulation commands such as viewing directory contents, moving, renaming, and deleting files. Note: different file systems impose different maximum sizes on both the total file system as well as individual files. For example, the ext3 file system, commonly used for Linux can contain a maximum of 232 blocks, where a block can range from 1 KiB, 2 KiB, 4 KiB or 8 KiB. In the 8KiB configuration, an individual file is limited to approximately 2 TiB, while the file system is limited to approximately 32 TiB.[14] Oracle released the ZFS file system, developed by Jeff Bonwick from Sun, and can store up to 256 quadrillion zettabytes (ZB), where a zettabyte is $2^{70}$ bytes.[15] Some file systems will also offer additional security features such as journaling, which keeps a log of changes to the file system as they are being made.

***Logical Unit Number (LUN)***

A logical unit represents a virtual segment of a disk or a set of disks. The LUN assigns a unique identifier to each logical unit, allowing the operating system or storage software to track, manage, and if necessary move it to a new physical location.[16] A logical unit can span physical devices (e.g. used in RAID disk systems). A file system can be configured on a logical unit.

***Partition***

Partitioning is the virtual subdivision of a physical hard disk. A physical disk may contain a single partition that represents the drive's full capacity, or it may be divided into smaller segments that function like physically separate drives. Different partitions may use different file systems; part of a drive, for example, may be formatted using the Windows NTFS, while the remainder is formatted for Linux using the ext3 file system.[17]

## Tape Technologies

Even as the capacity of disk drives has steadily increased and the cost of drives has decreased, tape systems remain a viable and attractive component for the overall storage architecture. When provided with an uninterrupted high-speed data stream, tape still provides the greatest data transfer performance of all media types, which is especially suitable to support backup needs.[18] Tape is commonly referred to as "offline" as the data is not usually immediately accessible. However, tape technology provides benefits of media portability allowing tape cartridges can be stored in an offsite facility, and longevity.

**Note:** Refer to the USGS "Archive and Records Management Offline - Fiscal Year 2010 Archive Media Trade Study" [19] for a comprehensive study that compares several offline digital archive storage technologies. The most recent study was published for fiscal year 2010. Earlier reports from fiscal year 2004, 2006, and 2008 have also been prepared. This provides an excellent side-by-side comparison of several tape technologies including feature/function, performance, and comparative cost information.

***Linear Tape-Open (LTO)***

As of 2006, LTO held an 82% marketing share, and has since seen the discontinuance of competitors DLT and SAIT.[20] (Note: LTO is included here due to its significant market share. Its inclusion here is for informational purposes to illustrate expanding tape capacity, and does not denote any sort of endorsement or recommendation by the GeoMAPP project.) LTO is a standard tape form factor and goes by the commercial name of Ultrium, and was formed in 1997 by HP, IBM, and Quantum, to provide an open format specification for tape storage options.

---

[14] "ext3." Wikipedia. http://en.wikipedia.org/wiki/Ext3
[15] "ZFS." Wikipedia. http://en.wikipedia.org/wiki/ZFS
[16] Matthew McKenzie. "57 Storage Terms You Need to Know." Information Week, SMB. June 2010. http://www.informationweek.com/whitepaper/Hardware/Peripherals/57-storage-terms-you-need-to-know-wp1278003128543
[17] Ibid.
[18] Quantum. "D2D2T Backup Architectures and the Impact of Data De-duplication." 2008. http://salestools.quantum.com/getDocPRetriever.cfm?ext=.pdf&type_mime=application/pdf&filename=576635.pdf&doc_id=28383
[19] USGS. Archive and Records Management Offline - Fiscal year 2010 Offline Archive Media Trade Study. 2010. http://eros.usgs.gov/government/records/media/FY10MediaTradeStudy.pdf
[20] USGS. Archive and Records Management Offline - Fiscal year 2010 Offline Archive Media Trade Study. 2010. http://eros.usgs.gov/government/records/media/FY10MediaTradeStudy.pdf

The first version was released in 2000 and had a capacity of 100 GB of uncompressed data, and 200GB of compressed data compressed, per cartridge. LTO-(Generation) 5 was released in 2010 and will support a native capacity of 1.5 TB per cartridge with a maximum data rate of 140 MB/s. Announced in May 2010, LTO-8 will support a native capacity of 12.8 TB, and 32.8 TB compressed capacity per tape with a native data rate of 472 MB/sec and a compressed data rate of up to 1180 MB/sec. They have also introduced some advanced storage management features such as WORM (LTO-3), drive-level encryption (LTO-4), and partitioning (LTO-5).[21]

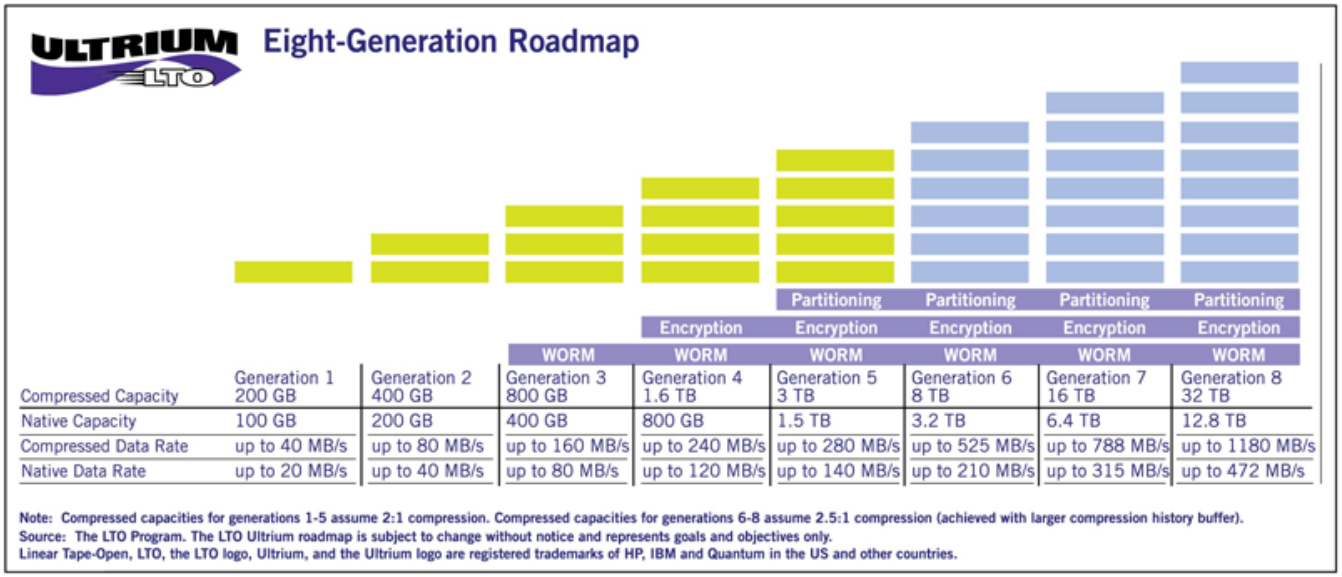



Figure 3: Ultrium LTO Eight-Generation Roadmap[22]

The tape drive technology has expanded to support multiple read/write elements per head, and has incorporated reliability features such as a read/verify element that will immediately verify the data written to the tape. If an error is detected, the block is rewritten farther down the tape.[23]

***Other Tape Media***

There are several other tape media technologies that have been developed including the ½ inch reel tapes created by 3M in 1972 that supported densities of 800-6250 bits per inch (bpi). The workstation era in the 1980s and 1990s, saw quarter-inch cartridge (QIC) tapes, but there were more than 120 standards, and not all QIC drives were compatible with all standards. Based on the CD-quality digital audio tape (DAT) format, HP and Sony defined the Digital Data Standard (4 mm tape), transforming an audio-based standard into a data standard. The 8 mm format originated from the video industry, and is generally used to transfer high-quality color images to tape for storage and retrieval. As an example of how vendors can supplement the technology with additional capabilities, IBM's TotalStorage Enterprise Tape System supports a RAID-like striping capability applied to tape.[24]

***Autoloaders and Tape Libraries***

These reduce the manual labor involved in swapping tapes during a tape-writing operation. An autoloader is a machine that has one tape drive but holds numerous tapes. A tape library includes at least two tape drives, and holds numerous tapes. Enterprise-level tape libraries may include other advanced tape management features such

[21] Ultrium. "LTO Ultrium Generations." http://www.lto.org/technology/generations.html
[22] Ultrium. "Ultrium LTO Eight-Generation Roadmap." http://www.lto.org/technology/roadmap.html
[23] Ultriuim. "LTO Technology Primer." http://www.lto.org/technology/primer3.html
[24] IBM. "IBM System Storage Tape Library Guide for Open Systems." 2011. http://www.redbooks.ibm.com/redbooks/pdfs/sg245946.pdf

as barcode scanners, robotic mechanisms for automatically retrieving and loading tapes, and redundant hardware components to minimize the impact of component failures.[25]

***Virtual Tape Libraries (VTL)***

A virtual tape library is a high-performance storage (e.g. disk) and server combination that emulates a tape storage device. Backup software has been designed to write to tape and tape library types of devices. However, with the falling cost of disk storage, and the increased demands to recover data quickly, some organizations are turning to disk-based devices for their backup medium. As the virtual tape library emulates a tape-based system, backup software can be configured to back up the data to a disk masquerading as a tape device. Virtual tape libraries are also providing additional value through additional features such as encryption, compression, and data deduplication. A virtual tape library may not necessarily negate the need for tape-based backups, as, e.g. tapes have the advantage of being able to be stored offsite, but may, instead be inserted as an intermediary between the data sources and the tape system in a disk-to-disk-to-tape (D2D2T) type of configuration. Leveraging disk's better performance, the backup can be quickly "staged" to the intermediate virtual tape library, and then the backup can be transferred from the virtual tape library to the tape system, reducing the backup window on the data source systems, and possibly necessitating less downtime on the source systems in order to perform the backup operation. In addition, recoveries from recent backups can be more quickly restored as they may still be cached on the disk-based virtual library system.[26]

## Hierarchical Media Architectures

Traditionally, backups were written to tape devices to support disaster recover (DR) situations. However, with the growth in daily data production, the need to shrink backup windows (which may require storage resources to be unavailable), and growing needs for long-term archival storage, multi-tiered storage architectures are being developed.

disk => to => tape (D2T)

Traditional backup architecture creating a backup image on tape of the data and/or files that reside on the disk.

Applications: backup, archiving

disk => to => optical => to => tape

An architecture configuration where an optical disk device is used as an intermediary on behalf of tape device. Files may be migrated to the intermediate device either based on policy or usage (e.g. time since last accessed) and then transferred to a tape device. The optical intermediary provides quicker access to files that are older, but still periodically accessed.

Applications: backup, archiving

disk => to => disk => to => tape (D2D2T)

A more recent multi-tiered storage architecture configuration where a disk is used as an intermediary storage device. In the VTL scenario, the intermediate disk will frequently look like a tape device to the primary disk, and the primary server will think it's just creating a traditional backup. Alternatively, an intermediate disk can be simply used as a staging device to "buffer" files targeted to be written to tape.

Applications: backup, archiving

***Related Topic: Hierarchical Storage Management Software***

Special storage management software manages the migration of files from one tier in the storage architecture to the next for backup purposes. Hierarchical storage management software can be used to configure policies that define the basis for when files should be migrated from one level of the storage hierarchy to the next. Policies can also be developed that stipulate criteria that may, for example, retain more frequently used data assets on the

[25] Ibid.
[26] Ibid.

intermediate, faster-accessible device. Hierarchical storage management software may even direct how the data is written to the physical device to optimize its storage and access availability.

## *Storage Networking Technologies*

Storage networking technologies provide an alternative to the storage device(s) that is directly attached to the computer (direct attached storage (DAS)), by putting them out onto the network. The primary benefits of storage networking technologies are:

- shared access by multiple users (a DAS is accessible by only the person(s) and/or applications logged into that computer) , and
- centralized management (a central pool of storage is managed, rather than individual DAS devices on each computer).

There are two primary storage networking technologies,

1. Network Attached Storage (NAS), where the storage is attached to an Ethernet-based local area network, and
2. Storage Area Network (SAN), where the storage is attached to a Fibre Channel network.

### *Network Attached Storage (NAS)*

Network attached storage is a specialized file server, sometimes referred to as an "appliance," that connects simply to the local area Ethernet network using a standard network connection and cabling (instead of connecting directly to a computer, as in the direct attached storage (DAS) device), as seen in Figure 4. The NAS provides centralized shared storage for an organization, and the storage can be made accessible to anyone on the network. The device usually consists of storage (hard disk SATA drive(s) or USB storage) and a simplified operating system customized for managing file input/out requests. NAS devices are relatively easy to provision. In some cases it may be as simple as plugging the device into the network, powering on the device, and configuring a few basic options through the administrative interface. The NAS may support a variety of file systems and network resource sharing protocols (e.g. NFS (UNIX), SMB/CIFS (Windows), NCP (Netware)). Once connected to the network, the NAS often appears to users simply as "share" drive that they can connect to. However, as the NAS is communicating over the Ethernet network, file requests will add to the network traffic, and local area network traffic may be erratic. Also, NAS support to serve files for application servers may be limited.

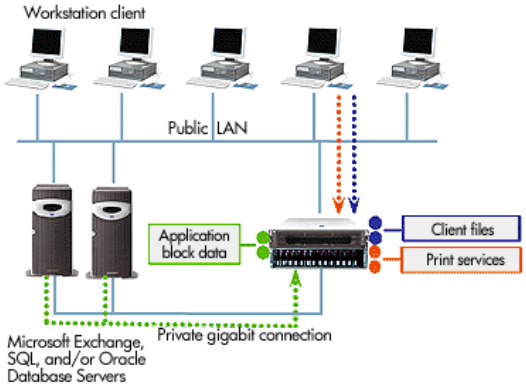



Figure 4: Network Attached Storage[27]

There are a variety of levels of NAS, ranging from:

1. The low end devices that provide very simple management features and do not support storage expansion,
2. Devices that support storage expansion through integration with iSCSI devices,
3. Devices that support more advanced storage features such as RAID (mirroring, parity), and integrating with the corporate directory service (e.g. Active Directory).[28]

[27] Advanced Network Solutions. "Storage Area Networks (SAN) / Network Attached Storage (NAS)." http://www.ansnetworks.com/storage-area-networks/

A NAS also provides the benefits of a centralized file server for both storage management and backup. As the files are stored on one central system, rather than distributed across the network on each individual's local direct-attached storage device, the IT department need only backup the centralized NAS storage ... though the availability of the NAS does not necessarily prevent or preclude users from storing files on their local disk devices. The NAS also provides a low cost, low barrier to entry as the Information Technology management staff is usually already familiar with managing Ethernet-based local area network devices, and does not require the specialized management skills or the complex technical infrastructure required by a Storage Area Network. NAS also often provide a web-based administrative interface allowing administrators to administer the NAS through any computer in the network.

***Storage Area Network (SAN)***

A storage area network is a more advanced storage networking model. A SAN is a high speed network that consists of a communication infrastructure and a management layer. The communications infrastructure provides the physical connections, and the management layer organizes the connections, storage elements, and computer systems so that data transfer is secure and robust.[29] A SAN connects a collection of storage devices including all of the technologies previously discussed and equipment. SAN is usually deployed over a Fibre Channel network, to create a shared pool of storage on the enterprise network.[30] Typically, a SAN is complex to install, configure and manage, and often requires consulting services and special administrative training to manage on a day-to-day basis. While SANs are often built on a dedicated Fibre Channel network infrastructure, they can also use iSCSI or Fibre Channel Over IP.

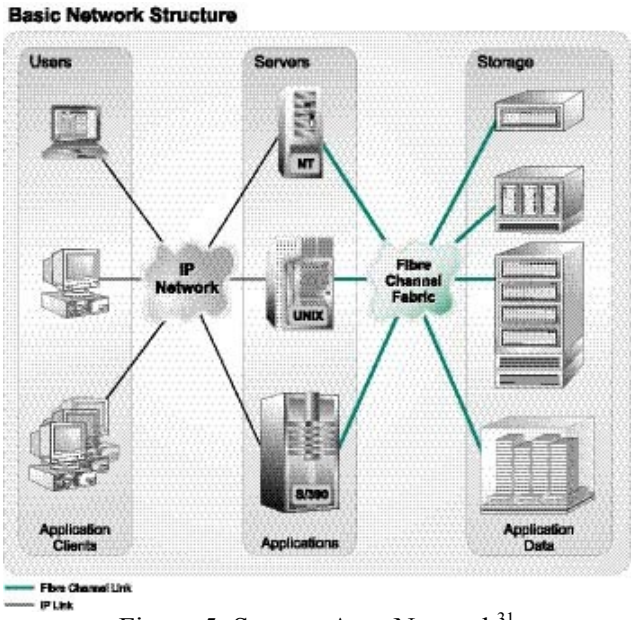



Figure 5: Storage Area Network[31]

In addition to advanced storage management features, such as redundancy for failover, mirroring, traffic re-routing, server clustering to localize or constrain storage access, and backup services, storage capacity can be added as the organization needs it.[32] SANs are used to connect storage arrays and tape libraries to multiple

---

[28] Samara Lynn. "How to buy a NAS." PC Mag.com. 2011. http://www.pcmag.com/article2/0,2817,2354173,00.asp#fbid=UfDNNp0tmph

[29] Matthew McKenzie. "57 Storage Terms You Need to Know." Information Week, SMB. June 2010. http://www.informationweek.com/whitepaper/Hardware/Peripherals/57-storage-terms-you-need-to-know-wp1278003128543

[30] Ibid.

[31] Fibre Channel Industry Assoc. "Topology." http://www.fibrechannel.org/overview/fcbasics/topologies

[32] HP. "SAN Design Reference Guide." 2011. http://h20000.www2.hp.com/bc/docs/support/SupportManual/c00403562/c00403562.pdf

servers, and are used by clustered servers for failover.[33] SANs also have the benefit of allowing servers and users to access the networked storage device as if it were a direct attached storage resource.[34]

## Storage Virtualization

In a virtualized storage environment, the physical location of the data is decoupled from the naming of the data, and thus the data can reside at any addressable space. The virtualized storage environment will maintain the linkages between the data names and the physical data locations in a metadata layer. Storage virtualization offers benefits of location transparency, meaning that from the application's perspective that is requesting the file by name, it does not need to know where that storage is physically located, and from the administrator's perspective, the physical data can be moved, the internal metadata cross-references updated, and the applications continue to run totally unaffected. Storage virtualization can also offer benefits in adding new storage to the storage environment, as the storage system is no longer bound by its physical constraints.

## Disk Interface Protocols

***Small Computer System Interface (SCSCI)***
SCSI is a set of standards for physically connecting and transferring data between computers and peripheral devices, most commonly used to communicate with hard disks in a direct attached storage configuration.[35] The SCSI bus is more complex than the SATA bus, usually resulting in higher manufacturing costs. However, SCSI busses allow connection of several drives on one shared channel, whereas SATA allows one drive per channel.

***Serial ATA (SATA)***
SATA is a standard for connecting devices like optical drives and hard drives to a motherboard, and is commonly found in personal computers and mobile computing devices. There have been continued enhancements to the standard, offering greater throughput rates including the SATA 3Gbit/sec, and the most recent SATA 6Gbit/sec standard, released in 2009.[36] In August, 2011, SATA announced a new interconnect specification "designed to boost interface throughput of solid state drive (SSD) technology from 6 GB/sec to 8Gb/sec and 16Gb/s. ... The specification will define new device and motherboard connectors that will support both the new SATA Express and current SATA (3 Gbit/sec) devices."[37]

***Serial Attached SCSI (SAS)***
SAS is a computer bus for moving data between the computer and storage devices such as hard drives and tape drives. It uses a point-to-point serial protocol and replaces the parallel SCSI bus technology that appeared in the 1980s. It uses the standard SCSI command set to communicate with the storage device.[38] The first SAS deployments were targeted at servers with direct-attached storage (DAS). The SAS host controller, usually on the motherboard, can provide varying levels of RAID support providing additional data protection. SAS is the first standard specification to provide an interconnect mechanism for both SCSI and SATA. For example, a SATA 3Gb/s drive can be connected to a SAS backplane, but SAS drives may not be connected to SATA backplanes.[39]

---

[33] Jon Tate, Fabiano Lucchese, Richard Moore. "IBM: Introduction to Storage Area Networks." 2006. http://www.redbooks.ibm.com/abstracts/sg245470.html
[34] Matthew McKenzie. "Storage Primer." http://www.informationweek.com/whitepaper/Hardware/Peripherals/storage-primer-wp1278003182430?articleID=151500072
[35] "SCSI". Wikipedia. http://en.wikipedia.org/wiki/Scsi
[36] Serial ATA International Organization "Fast Just Got Faster: SATA 6 Gb/s." 2009. http://www.sata-io.org/documents/SATA-6Gbs-Fast-Just-Got-Faster.pdf
[37] Serial ATA International Organization. "SATA Group creates new high-speed storage spec. PCI Device Interface Boosts Speeds to 8 Gb/s and 16 Gb/s" System News. Aug. 18, 2011. http://sun.systemnews.com/articles/162/3/storage/24447
[38] "Serial attached SCSI". Wikipedia. http://en.wikipedia.org/wiki/Serial_attached_SCSI
[39] "The Benefits of Serial Attached SCSI (SAS)." http://www.soltechnology.com/articles/Benefits-of-Serial-Attached-SCSI.htm

## Storage Management Features—A Checklist

With the advances in today's storage solutions, storage vendors are developing integrated storage solutions that offer storage as well as management features. One advantage of SANs is that they have implemented many of these features within their storage solution. However, as the storage industry continues to evolve, it is extending the capabilities of non-SAN solutions to offer these types of features.

Many of these features are especially relevant for archivists and archival solutions as they strive to preserve digital information indefinitely. There may be some variation from vendor to vendor in what these features are called, but you should be able to inquire about the general capability. The following may serve as a helpful checklist of storage features to discuss with your storage vendor as you are evaluating storage alternatives.

✓ ***Browser-based Administrative Interface***

Storage solutions such as network-attached storage (NAS), that attach directly to a TCP/IP network will often provide a web browser-based administrative interface, allowing the administrator to administer the storage device from any computer that has a web browser and a connection to the network, as opposed to having to administer the storage device by communicating via a terminal or system directly attached to the storage system.

✓ ***Clones***

A clone is an identical copy of the data in a volume. This can be used to migrate data to a newer media. For example, archival management recommends the retirement of old media, and moving the data to newer media. With cloning, you can set a volume to "retire" in two years, and when the volume reaches its retirement date, you use the cloning capability to simply move all of the content from the retired volume to a new volume. Some cloning solutions will also perform integrity checking on the original volume being cloned as the copy is being made. Note: a clone is not intended for disaster recovery.

✓ ***Content-Addressed (or Addressable) Storage (CAS) (also known as Object Storage)***

"Content-addressed storage (CAS) is a method to store and access fixed content (data this is not expected to change), by assigning it a permanent location on a disk. Once the item has been written to the disk, it cannot be modified nor duplicated. In some cases it cannot be removed until the retention period has expired. The data is retrieved based on its content, known as its content address, rather than its storage location, and is guaranteed to be unique and unambiguous. [40] With its fixity feature, CAS is a good match for archival needs as archived content is not expected to be modified once written to the archival repository. The storage system will enforce the immutability and authenticity of the content within the storage system, thus helping to enforce the policies of the repository organization.

✓ ***Continuous Data Protection (CDP)***

"Continuous data protection, also called continuous backup, is a storage system in which all the data in an enterprise is backed up whenever any change is made. In effect, CDP creates an electronic journal of complete storage snapshots, one storage snapshot for every instant in time that data modification occurs. A major advantage of CDP is the fact that it preserves a record of every transaction that takes place in the enterprise. In addition, if the system becomes infected with a virus, or if a file becomes mutilated or corrupted, it is always possible to recover the most recent clean copy of the affected file. A CDP system with disk storage offers data recovery in a matter of seconds -- much less time than is the case with tape backups or archives. Installation of CDP hardware and programming is straightforward, simple and does not put existing data at risk."[41]

✓ ***Data Compression***

Data compression reduces the amount of physical storage required to store the information. Backup solutions (e.g. virtual tape libraries) may include data compression algorithms to maximize the usage of the backup storage device. Data compression can greatly decrease the data traffic and reduce the overall bandwidth requirements.

---

[40] "Content-Addressed Storage (CAS)." http://searchstorage.techtarget.com/definition/content-addressed-storage

[41] "Continuous Data Protection (storage convergence). http://searchstorage.techtarget.com/definition/continuous-data-protection

✓ ***Data Deduplication***

Data deduplication is the process of replacing multiple copies of data with pointers to a single, shared copy which can dramatically reduce storage requirements.[42] It is performed at the sub-file level, block-level deduplication being the most typical, but some products will find differences at the byte level. Talk with your vendor about support and how their technology leverages deduplication.

✓ ***Data Integrity Checking***

Data integrity is a fundamental aspect of storage security and reliability, especially for archival organizations that are promoting preservation and maintaining a digital asset's authenticity over time. There are a variety of factors that can cause modification, corruption and/or errors, to data (e.g. files) written to some sort of media, or failures in the media itself. Data integrity checking is an ongoing mechanism to guarantee that the bits that have been written to the media have not been modified, altered, or corrupted. There are a variety of technologies and techniques, which can be implemented at either the software level or the hardware level. For example, the parity mechanism in RAID storage provides an integrity checking mechanism. As storage systems and storage management software are being evaluated, an important consideration is understanding the data integrity checking capabilities provided. In the absence of built-in data integrity checking capabilities, the archival organization will need to develop a method of integrity checking (e.g. capturing the baseline parity or hash, periodically checking to verify the data is unaltered) through the use of techniques such as parity bits, file checksums, or hashing. A key aspect of the data integrity checking system is what the resulting action will be when a data integrity check fails. There are a range of responses based on the technology and sophistication of the system ranging from a simple log entry to automatically recovering or restoring the damaged data.

✓ ***Encryption***

Encryption is the process of transforming information using an algorithm to make it unreadable except to those that have the decryption method and/or key.[43] It is a method of providing additional security for sensitive data and/or documents. Encryption may be included as a feature of storage systems or as a feature of storage management software.

✓ ***10 GbEthernet*** (GbE)

While 10 GbE offers substantially greater bandwidth than it's 1 Gigabit Ethernet predecessor, its expense may make it cost-prohibitive for IT shops to implement, and may require cost justification. In many cases, cheaper 1 Gigabit Ethernet technology may still be adequate to meet the needs of a company's block-based iSCSI storage, or file-based NAS storage. Two of the more common drivers to move to 10 GbE are the desire to run more virtual machines per server and to consolidate data centers, or to support high bandwidth applications such as video streaming. 1 Gigabit Ethernet can cost less than $100 per switch port, while 10 GbE is in the range of $500 per switch port. There will also be the cost associated with other modifications throughout the networking infrastructure as the organization moves to a 10 Gb Ethernet infrastructure.[44]

✓ ***Error Correction / Self-healing***

There are two general contexts for error correction: 1) when data is first being written to the media, and 2) when data is being accessed from the media. Self-healing is the ability to restore from a failure situation.[45] This may occur at a variety of levels from encountering a bad spot on media when either reading or writing data, or recovering from the failure of a complete media device, such as a hard disk drive. Digital LinearTape drives' read-after-write capability offered a mechanism that guaranteed that data was correctly written to the media, and

---

[42] Matthew McKenzie. "57 Storage Terms You Need to Know." Information Week, SMB. June 2010. http://www.informationweek.com/whitepaper/Hardware/Peripherals/57-storage-terms-you-need-to-know-wp1278003128543

[43] Wikipedia. "Encryption." http://en.wikipedia.org/wiki/Encryption

[44] "10 Gigabit Ethernet for iSCSI storage, NAS storage still hinges on cost."July 2010. http://searchstorage.techtarget.com/feature/10-Gigabit-Ethernet-for-iSCSI-storage-NAS-storage-still-hinges-on-cost

[45] Marc Staimer. Storage, Heal Thyself. Storage Magazine. June, 2009. http://www.necam.com/itpg/Docs.cfm?id=ec21d8bd-6ec1-4773-add8-e2139522f82e

could rewrite the data to a later tape position if the post-write failed. RAID offers two approaches to error correction and self-healing through the ability to rebuild a hard disk drive through either a mirror or through the parity information.[46]

✓ ***Expandability of Storage Capacity***

The ability to add additional storage to the storage system, and have the system automatically make that new storage space available. For example, lower end, lower cost NAS systems may not offer the ability to expand storage within the NAS appliance, while RAID and SAN storage systems generally allow the addition of additional disk devices within their storage array. As archival organizations accession additional collections of digital materials, the cost and ease with which storage can be added over time to the storage infrastructure is an important consideration.

✓ ***Hierarchical Storage Management***

A multi-tiered storage management solution that can move data from one media type to another, usually from faster more expensive media to slower less expensive media. The solution may even manage the placement of data upon the media to optimize use and access. A couple of typical hierarchical configuration may consist of disk to tape, or disk to optical library to tape.

✓ ***Load Balancing***

Load balancing offers storage performance improvements by distributing the operation across multiple systems. Loads can be balanced across computing CPUs, through I/O channels, or at the disk level. RAID 3 offers a good example of storage load balancing where data is striped across three different drives, taking advantage of three disks' read and write heads to perform operations in parallel. Storage Area Networks through its network "fabric" offers multiple physical connections to storage devices, providing load balancing capabilities to optimize the performance across the storage network.[47]

✓ ***Portability***

A key benefit of tape-based backups is their portability, the ability to move the tapes to another tape reader machine or to an offsite storage facility. While there are portable disks, disk-based storage systems and storage networks, except for the NAS appliances, tend to not be portable. Portability offers an additional layer of data security, as the data can be moved to a remote site, thus reducing the risk associated with storing all data at a single site or facility.

✓ ***Replication / Mirroring***

A mirror is an exact replica of the original storage device. RAID offers mirroring options within the RAID disk array, which can offer performance benefits due to multiple read heads accessing the data. Mirrors also provide fault tolerance. In the event that one mirror goes down, the other mirror remains available. Mirrors are not generally considered a disaster recovery solution.

✓ ***Remote Data Replication***

Remote data replication offers the security of geographically distributed data replication through an automated process. Instead of writing data to a local tape drive or tape library and having to physically transport the media to a remote location, the remote site is accessible through the network, and can be written to through the storage management software.

✓ ***Retention Schedules***

Some storage systems are adding more records management-like functionality, including creating retention schedules that will handle the automatic deletion of content that has reached its expiration date, or at least alert the administrator to the expiration, so that he can address it accordingly. As archival collections are often intended

---

[46] John Wilkes, Richard Golding, Carl Staelin, Tim Sullilvan. "The HP AutoRAID Hierarchical Storage System." Hewlett-Packard Laboratories. 1996. http://www.hpl.hp.com/research/ssp/papers/AutoRAID.TOCS.pdf

[47] "SAN availability and Reliability." http://searchstorage.techtarget.com/feature/SAN-availability-and-reliability

for perpetual storage, support of features such as retention schedules will not necessarily be of as much importance.

✓ ***Snapshots***

Snapshots are a common industry term denoting the ability to record the state of a storage device at any given moment and preserve that snapshot as a guide for restoring the storage device in the event that it fails. It is essentially a point-in-time copy of the data. At some point in time a snapshot is initiated. The snapshot is often written to a dedicated location, and simply tracks the changing blocks on the original volume, as writes to the volume are performed. Snapshots contain the metadata that describes the data blocks that have changed since the snapshot was first created.[48]

✓ ***Thin Provisioning (TP)***

Thin provisioning is a method of optimizing efficiency with which the available storage is utilized in storage area networks (SAN). TP operates by allocating disk storage space in a flexible manner among multiple users, based on the minimum space required by each user at any given time. In the conventional storage provisioning model, also known as fat provisioning, storage space is allocated, and possibly paid for, beyond the current needs, in anticipation of the growing storage needs and increased data complexity. Potentially, large amounts of storage space are allocated, but never used.[49] Thin provisioning can be paired with dynamic storage allocation, in which users' storage allocations can be dynamically extended, as their storage needs expand. Some iSCSI storage arrays, outside of SANs, are starting to provide this capability.[50] This could be a really valuable feature for archival organizations, as they should be able to allocate just the amount of storage that they need, and then easily add additional storage capacity as their collections grow.

✓ ***WORM support***

Write-Once-Read-Many is a data writing / access feature that is of particular interest to archival organizations attempting to maintain the authenticity of preserved digital assets. One way to ensure that a digital asset remains unchanged, is to write it using a mechanism that prohibits further updates or modifications. WORM may be implemented through the media used, such as WORM CDs that can only be written to once, or can be implemented through storage management features that implement and enforce the WORM policy.

## References

Storage Magazine. http://searchstorage.techtarget.com/magazine-sections/2011/03

Fibre Channel Industry Association. http://www.fibrechannel.org/

Serial ATA International Organization. http://www.sata-io.org/

Gartner. "Eight Magic Quadrants for the Storage Industry."
http://www.gartner.com/DisplayDocument?doc_cd=114876

1. Disk Arrays
   Magic Quadrant for High-end Disk Array
   Magic Quadrant for Midrange Enterprise Disk Arrays
2. SAN Switches, Solutions, and Services
   Magic Quadrant for SAN Fibre Channel Switches
   Magic Quadrant for SAN Integrated Solution

[48] Neeta Garimella. "Understanding and exploiting technology for data protection, Part 1: Snapshot technology overvew." IBM. http://www.ibm.com/developerworks/tivoli/library/t-snaptsm1/index.html

[49] "thin provisioning (TP)." http://searchstorage.techtarget.com/definition/thin-provisioning

[50] Dell. "How Thin Provisioning with Dell EqualLogic iSCSI Storage Arrays Simplifies Management." http://www.dell.com/downloads/global/power/ps2q08-20080314-EqualLogic.pdf (Note: This is included strictly for illustration purposes. This is not an endorsement or recommendation for this product)

3. Virtual tape
   Enterprise Virtual Tape Subsystem Magic Quadrant
4. Storage Management Software
   SAN Management Software Magic Quadrant
   Initial Storage Provisioning Magic Quadrant
5. Storage Services
   Storage Services Magic Quadrant

"Raid." PC Magazine Encyclopedia.
http://www.pcmag.com/encyclopedia_term/0,2542,t=RAID&i=50148,00.asp#fbid=UfDNNp0tmph

SearchSMBStorage.com. "Primer on RAID Levels: Standard RAID Levels Defined."
http://media.techtarget.com/searchSMBStorage/downloads/Standard_RAID_levels_defined.pdf

Tate, Jon, Jim Kelly, Pauli Ramo, Leos Stehlik. "IBM TotalStorage: SAN Product, Design, and Optimization Guide." 2005. http://www.redbooks.ibm.com/redbooks/pdfs/sg246384.pdf

USGS. Archive and Records Management Offline - Fiscal year 2010 Offline Archive Media Trade Study. 2010.
http://eros.usgs.gov/government/records/media/FY10MediaTradeStudy.pdf

**APPENDIX**

## Network Storage Protocols

***Common Internet File System (CIFS)***
Formerly known as Server Message Block (SMB), Common Internet File System was co-developed by IBM and Microsoft to support file sharing in DOS, and is used as a file-sharing protocol in Microsoft Windows-based networks. CIFS can run over TCP as a transport protocol, or through the NetBIOS API which can run over UDP or several legacy protocols.[51] A protocol for accessing storage configured on NAS systems.

***Fibre Channel (FC)***
Fibre Channel is a high-speed data storage protocol commonly used in SAN networks to connect servers, network components, and shared storage resources (disks). Networking components include host bus adapters, hubs, and switches.[52] FC is a block-based protocol, transferring blocks of data over the network, as opposed to file-based protocols such as NFS that transfer files across the network.

***Fibre Channel Over Ethernet (FcOE)***
Fibre Channel Over Ethernet is an encapsulation of Fibre Channel frames over Ethernet networks. Many data centers use Ethernet for their TCP/IP networks and Fibre Channel for the SANs. [53] The benefit to data centers is cabling reduction, as IP network traffic and SAN data traffic can be consolidated to run over a single integrated network infrastructure.[54] See the EMC whitepaper "Fibre Channel Over Ethernet (FCoE)" for comparative topology diagrams.

***Internet SCSCI (iSCSI)***
Internet SCSI is a storage networking standard that enables storage systems to transfer data over local area networks, including SANs. iSCSI networking hardware may link arrays of storage resources (disks) to a company's servers.[55] iSCSI runs on top of the TCP/IP networking protocol layers.

***Network File System (NFS)***
An open standards-based network files system protocol originally developed by Sun Microsystems in 1984, and runs over Ethernet-based networks, that use TCP as a transport protocol.[56] NFS is a file-based protocol. NFS is a common storage access mechanism for NAS systems.[57]

---

[51] "Server Message Block." Wikipedia. http://en.wikipedia.org/wiki/Server_Message_Block
[52] Matthew McKenzie. "57 Storage Terms You Need to Know." Information Week, SMB. June 2010. http://www.informationweek.com/whitepaper/Hardware/Peripherals/57-storage-terms-you-need-to-know-wp1278003128543
[53] "Fibre Channel over Ethernet." Wikipedia. http://en.wikipedia.org/wiki/Fibre_Channel_over_Ethernet
[54] Mark Lippit, Erik Smith, Erik Paine, Marky Anthony De Castro, Shreedhan Nikam. "Fibre Channel over Ethernet (FCoE). Data Center Bridging (DCB) Concepts and Protocols." EMC. 2011. http://www.emc.com/collateral/hardware/technical-documentation/h6290-fibre-channel-over-ethernet-techbook.pdf
[55] Matthew McKenzie. "Your Storage Questions Answered." Information Week, SMB. http://www.informationweek.com/whitepaper/Hardware/Peripherals/your-storage-questions-answered-wp1278003067980?articleID=151500066
[56] "Network File System (protocol)." Wikipedia. http://en.wikipedia.org/wiki/Network_File_System_%28protocol%29
[57] "Introduction to NAS - Network Attached Storage." http://compnetworking.about.com/od/itinformationtechnology/l/aa070101a.htm