



**STORAGE AND
FILE AREA
NETWORK**

Data Protection: Understanding the Benefits of Various Data Backup and Recovery Techniques

When implementing a backup and recovery solution, IT organizations should carefully define their recovery objectives and align their technology strategy with their business requirements.

With the growing value of data as a strategic corporate asset, today's IT organizations face the challenge of implementing reliable backup and recovery solutions in the most efficient, cost-effective manner. To meet this challenge, they need to carefully define their business requirements and recovery objectives before deciding on the right backup and recovery technologies to deploy.

This paper describes some of the key deciding factors and potential techniques these organizations should consider as they build the right backup and recovery infrastructure for their particular needs. For organizations that need assistance in finding the best solution, Brocade® Services provides a wide range of offerings for designing, enhancing, and managing backup and recovery infrastructures.

THE NEED FOR BETTER DATA PROTECTION

As today's businesses increasingly recognize corporate data as a strategic asset that must be protected against a wide range of risks and threats, data protection has become a high-priority objective for IT organizations. In turn, these organizations now face the challenges of identifying, deploying, and efficiently managing their data backup and recovery infrastructures.

One of the greatest challenges is justifying the costs associated with data protection solutions. This is because data protection is often viewed not as a revenue-generating function but rather a "necessary evil"—an overhead cost that most organizations would prefer to keep to a minimum. However, recent well-publicized events (natural and manmade disasters, lost backup tapes, and stolen corporate laptop computers) have increased the visibility and elevated the importance of data protection.

Although there is no standard definition for “data protection” and the functions it encompasses, most organizations agree that any comprehensive data protection strategy should include the following areas:

- **Data security:** Preventing unauthorized access to data, which might involve the use of technology such as encryption as well as application-based security technologies.
- **Data availability:** Ensuring that data is highly available to business applications. This generally entails deploying high-availability storage solutions that utilize technologies such as Redundant Array of Inexpensive Disks (RAID), multiple/redundant storage controllers, remote mirroring, multiple I/O paths, and clustered storage systems. These solutions focus on eliminating single points of failure in the data path and/or the storage target.
- **Data backup and recovery:** Ensuring that a point-in-time copy of data can be restored in order for business operations to resume.

It is important to note that these functional areas are not independent of each other. For example, a backup of a database is created to meet data backup and recovery requirements. However, it might be important for that backup copy to be encrypted from a data security standpoint, especially if the backup resides on portable storage media (such as a tape cartridge) that is taken out of the security of the data center and susceptible to unauthorized access.

BACKUP AND RECOVERY TECHNOLOGIES

Organizations can choose among numerous backup and recovery technologies that are available today—ranging from traditional methodologies to leading-edge solutions. These technologies generally fall into one of the following solution categories:

- **Core technology:** The traditional backup and recovery technology that most organizations use. This typically consists of one or more types of backup servers and backup clients. Backup clients send data over the LAN to the backup media server, which writes the backup objects to some type of storage. The master server maintains the catalog of backup objects, manages the backup schedules, and manages the media. Most midsize to large organizations have a core solution such as IBM Tivoli Storage Manager (TSM), Symantec NetBackup (NBU), EMC Legato NetWorker, CommVault Galaxy, and HP Data Protector.
- **Storage-based technology:** The technology that utilizes functions within the storage system to create additional copies of the data. These functions typically include full-copy data snapshots, pointer-based data snapshots, and synchronous and asynchronous replicated copies (typically over long distances). This is generally a homogeneous technology requiring similar storage systems (although virtualization is beginning to change this requirement). Most major storage vendors support native backup and recovery technologies within their storage systems.
- **Specialized technology:** A catch-all term for leading-edge technologies that have not yet gained wide acceptance within the industry and/or are targeted at specific applications or environments.

Because IT organizations need to meet their backup and recovery requirements in a cost-effective manner, they often deploy multiple technologies for various purposes. For example, an organization that utilizes TSM as its core backup and recovery technology might decide that TSM cannot cost-effectively meet the recovery objectives for its business-critical applications. In that case, the organization might decide to use a storage-based solution to provide backup and recovery for those particular applications.

DATA RECOVERY OBJECTIVES

Before they can begin developing an effective backup and recovery strategy, organizations should explicitly define their data recovery objectives for all of their business data in terms of:

1. **Recovery Point Objective (RPO):** The RPO is driven by two key parameters:
 - a. *The number of transactions that can be lost.* This can be actual business application transactions (such as a transaction-oriented database) or time units (such as hours). This defines acceptable data loss.
 - b. *The point in time from which it must be possible to recover data.* That is, how far back must the data be recoverable from? This parameter generally drives the retention settings used for the protected data.
2. **Recovery Time Objective (RTO):** The RTO defines how quickly the business function must be restored. Note that the RTO generally refers to the resumption of business function, not necessarily the amount of time available to perform data recovery. The window for data recovery is sometimes called the data recovery window.

Until they define these requirements, organizations will not be able to know whether their core solution (such as NetBackup or TSM) can actually meet their business requirements. They should consider these recovery objectives in regard to data recovery events, which vary widely and can result from operator error (such as a deleted file), hardware failure (such as a faulty disk drive), data corruption (such as a corrupted database), or a catastrophic event (such as a site disaster). Data recovery events are typically one of two types:

- **Operational recovery events:** Occur during normal business operations and are the events that backup administrators, system administrators, and database administrators deal with on a relatively frequent basis.
- **Disaster recovery events:** Involve the loss of a complete data center or a large portion of the data in a data center.

When planning for such events, it is important for organizations to differentiate between their *backup* requirements and their *archiving* requirements. The purpose of a backup is to satisfy the data recovery requirements of a specific data recovery event. Archives, on the other hand, satisfy business-specific records management requirements.

For example, at the end of a project an organization might decide to make a copy of all the files associated with that project and retain them as an archive for a specific number of years. Because this copy of the data is being created to meet specific requirements, it should:

- Be a separate copy of the data (not utilize the “backup” copy of the data)
- Have its own recovery objectives

CORE BACKUP AND RECOVERY TECHNOLOGY

For years, the “core” backup and recovery technology has been client/server software in open systems environments. This technology, coupled with disk and tape storage, is what most organizations continue to rely on (although most organizations also utilize storage-based and specialized technologies for specific recovery requirements).

Because core backup and recovery technologies such as NetBackup, TSM, and NetWorker can be relatively complex, they often require significant expertise and skill to manage. After all, backup and recovery touches every host platform, operating system, the LAN, the WAN, the SAN, disk storage, tape storage, applications, file systems, and databases—and it impacts all IT groups (storage, operations, systems, database, and applications). Further complicating matters, backup and recovery must operate in the “background” without impacting normal business operations or users.

Due to the complex nature of backup and recovery design, most organizations approach it from both a “macro” and “micro” level:

- **Macro design:** Addresses high-level design aspects of the overall backup and recovery environment, including factors such as the physical location of media servers with respect to the data to be backed up; whether duplicate copies of the backup images will be created; and the kinds of backup storage used. Organizations should use these macro design aspects and their associated workloads to size the required backup and recovery infrastructure.
- **Micro design:** Addresses specific backup and recovery considerations for individual applications, servers, and data sets. For example, organizations might decide to perform full daily backups for e-mail servers over the LAN with the backup images retained on disk.

Macro Design Decisions

During the macro design of the core backup and recovery infrastructure, software selection is a key consideration. Having to switch from one product to another or to another vendor’s product can be both expensive and disruptive. Moreover, investments in personnel training and site-specific processes and procedures should be a key part of the cost/benefit analysis during the selection process.

After selecting their software products, organizations need to make other key architecture decisions, including:

- Whether the backup servers and backup clients will be in the same physical location or separated from each other.
- Whether a secondary copy of the primary backup images will be created and, if so, where and how it will be created. Electronic vaulting is the process of creating secondary copies across a WAN to a remote physical location. Electronic vaulting eliminates the need to physically eject cartridges from a tape library and ship them to a remote location (thereby reducing potential failures and risk).
- What kind of storage technology will be used. Magnetic tape has been the traditional storage technology for core backup and recovery, but software vendors are increasingly supporting disk as a high-performance yet cost-effective backup target. Disk storage vendors have also developed products for existing core backup and recovery environments in the form of Virtual Tape Libraries (VTLs). These products look like standard SCSI tape libraries to the backup and recovery software, and they are relatively easy to integrate and manage.

Micro Design Decisions

After they have developed an overall architectural design and deployed a core backup and recovery infrastructure, organizations need to plan how to use that technology for each category of backup client within the environment.

For example, they might want to develop unique strategies for e-mail servers, NAS filers, large databases, and so on. Some of the many micro design decisions include:

- Whether the data will be transferred through a backup media sever via the LAN or directly between the backup client and storage media (LAN-free).
- Whether an additional piece of software (called an application agent or API agent) will be used in conjunction with the standard backup and recovery client software to more closely interface with the application (such as a relational database or e-mail).
- What kind of backup job will be used. Some options are full backup, full plus incremental backup, and “incremental forever” (used by TSM).
- When and how often the backup jobs will run. The impact on the production server is a critical factor, including the recovery objectives.
- How long the backup images will be retained. These retention values should be driven by data recovery objectives.
- What other job-specific requirements must be addressed. This includes scheduling methods, number of simultaneous sessions, monitoring/alerting requirements, restart requirements, and storage target and pre- and/or post-processing requirements.

Infrastructure Sizing

Infrastructure sizing and capacity planning are also key components of a successful core backup and recovery environment, including proactive capacity planning to support continued growth. Two aspects of capacity are critical:

- **Bandwidth:** The amount of data backed up in a given amount of time. This is generally the number of gigabytes or terabytes backed up per day. The bandwidth is driven by the amount of data being protected and how that data is protected.
- **Storage:** The total capacity of all the backup images that must be maintained by the backup software. The storage component stems from the amount of data being protected as well as the retention parameters of the data.

A properly designed backup and recovery infrastructure must be based on a sensible operational model. The operational model defines the number of hours available during a 24-hour day to complete specific backup and recovery operations. This is similar to the manner in which batch processing tasks are defined in a mainframe environment. Each operational task has an “operational window” (for example, the number of hours) during which it must complete. Operational compliance means that all tasks are completed within the defined window.

Workload Management

As described previously, the backup and recovery workload consists of both a bandwidth and storage component. The total cost of ownership of the backup and recovery infrastructure is proportional to the workload: the greater the workload, the greater the amount of infrastructure and personnel resources required to meet those needs. Therefore, organizations should define the most realistic requirements and deploy the most effective backup and recovery policies to meet those requirements.

Often, the existing backup and recovery policies are based on decisions made years earlier when certain technologies were not even available. To avoid inefficient resource utilization and higher costs, organizations should make sure they update these policies for their most recent business requirements and technologies.

ADDITIONAL STORAGE-BASED BACKUP AND RECOVERY TECHNOLOGY

Although this paper focuses on core backup and recovery technology, it's important to understand other data backup and recovery technologies, including *storage-based technology* and *specialized technology*. Organizations typically utilize these technologies to augment their infrastructures because:

- The existing core technology simply cannot meet the RTOs, RPOs, or both
- The existing core technology cannot meet the RTOs, RPOs, or both in a cost-effective manner
- Certain applications or data sets require specialized functionality

In the case of storage-based technologies such as disk storage systems, storage capacity takes one of two formats:

- **Block format:** In this format, disk capacity is presented to the host as a LUN that appears as a disk drive. The host is then responsible for formatting the LUN and building the appropriate semantics (such as a file system) so that the applications running on the host can use it. In open systems, this form of storage is accessed by the host via SCSI commands, typically in a Fibre Channel SAN. SANs were driven primarily by the need to share large block-based disk storage systems among multiple hosts. These systems are typically classified as tier-1, tier-2, or tier-3 systems based on overall capacity and throughput; the number of "front-end" Fibre Channel connections; the disk technology being used (Fibre Channel, SATA); the size of the cache; the overall Reliability, Availability, and Serviceability (RAS); and the software functionality, including backup and recovery features.
- **File format:** In this format, the storage system presents the disk capacity as a usable file system to the host via the TCP/IP network. This storage is commonly known as Network Attached Storage (NAS) and is accessed via the IP network. Users access data via CIFS for Windows environments and NFS for UNIX environments. These NAS systems have proliferated in Windows environments in part because the native features help organizations overcome the difficult task of providing backup and recovery for large file systems (1 TB in size and larger) that contain millions of files and directories.

Note that some storage-based technologies can support both block- and file-based services, accessing block-based services via Fibre Channel (or iSCSI in some cases) while accessing file-based services via the IP network.

Backup and Recovery Functions

Disk systems have provided backup and recovery functionality for years, utilizing the processing capacity of the storage system to make copies of the primary data for recovery purposes.

These functions typically take one of the following forms:

- **Full-copy snapshot:** A full-copy snapshot makes a complete copy of a LUN (block format) or a volume (file format) onto a separate set of disks within the disk system. In this case, 100 percent of the capacity of the original data is required to make the copy. For example, a snapshot of a 50 GB LUN needs 50 GB of capacity. After the first copy, some systems enable incremental updates at a later time, which creates a newer snapshot but writes only updated or new blocks.
- **Pointer-based snapshot:** A pointer-based snapshot provides a similar backup and recovery function as the full-copy snapshot with one major exception. The snapshot does not actually copy all the blocks of the primary LUN/volume but rather creates a set of pointers to the location of the original data. As the original data changes, a copy of the original data is made before it is updated ("copy on write"), and the snapshot reflects the new location of this data. As a result, a pointer-based snapshot might not require capacity equal to the size of the original LUN/volume.

- **Remote mirroring/replication:** Remote mirroring, also called replication, involves copying data from one storage system to another. In general the two storage systems must be “like” systems (same manufacturer and same product line). There are basically two modes of replication: synchronous and asynchronous. With synchronous replication every write transaction at the local site is replicated to the remote site before a write acknowledgement is returned. This ensures that both sites have a consistent copy of the data. With asynchronous replication write transactions are performed at the local site and then queued up for transmission to the remote site. As a result, the remote system might be “behind” the local system by anywhere from seconds to hours depending on the rate of writes, latency, bandwidth, and other factors. Note that mirroring alone does not provide foolproof data recovery in the event of data corruption.
- **Snapshots and remote mirroring:** Organizations can combine snapshot and remote mirroring functions to provide additional recovery options. A local snapshot can be taken and replicated remotely (generally asynchronously) to provide a remote copy of the data from a specific point in time.

Disk Storage Virtualization

One of the key drawbacks of storage-based solutions is that they tend to be tied to the vendor (and often a specific model) of the storage system. This makes it difficult if not impossible to deploy a flexible multivendor storage strategy.

Virtualization, as it applies to disk storage, enables organizations to move the same kinds of backup and recovery functions out of the disk storage systems. As a result, functions such as snapshots and remote mirroring can utilize heterogeneous disk systems. For example, organizations can copy a production database on a tier-1 disk system from one vendor into a snapshot to a tier-2 disk system from another vendor. This snapshot can then support backup and recovery, application development, and/or quality assurance.

A variety of storage virtualization architectures are available today:

- **Host-based storage virtualization:** Is an extension of the volume management software running on each host in the network with an additional platform that enables management from a single location.
- **Storage system-based virtualization:** Is generally constrained to the storage capacity within one system. Recently, however, some storage vendors have added the ability to manage the storage capacity of other storage systems.
- **Network-based virtualization:** Moves the virtualization function into the SAN and is classified as *in-band* (symmetric), *out-of-band* (asymmetric), or *split-path*. In-band solutions place a virtualization server in the data path between the host and storage. Out-of-band solutions require an agent running on each host. And split-path solutions utilize intelligence within the Fibre Channel SAN to properly direct I/O transactions. The split-path architecture provides the greatest scalability, performance, and manageability.

Regardless of the specific architecture, each of these storage virtualization techniques provides additional backup and recovery options.

Disk-to-Disk Backup, Virtual Tape, and Redundant Data Elimination

As noted previously, an important macro design consideration is the kind of storage media used to store the backup images. Within the past few years, the advent of SATA disk technology and the decreasing cost of disk capacity have made disk a much more cost-effective storage option for backup images.

Most organizations generally deploy this disk capacity for backup in the form of a VTL. The VTL appears to the core backup and recovery software as a SCSI tape library, so it is relatively easy to integrate into the environment. Organizations can also build their own VTL by:

- Procuring the VTL software from the software vendor
- Procuring their own VTL server hardware
- Procuring the required storage from a storage vendor

This “build-your-own” model provides the most architectural flexibility but requires independent support for each component. Organizations can also buy an appliance version of the VTL that contains all of the components (software, server, and storage) and is ready to be plugged into the SAN and configured to the backup and recovery software. The advantage of this version is that there is only one vendor responsible for support and maintenance. The primary disadvantage is that there are a limited number of models from which to choose, and scaling a VTL is subject to the particular vendor’s constraints.

A relatively new storage feature called Redundant Data Elimination (RDE) is emerging for backup and recovery, specifically for storing backup images. RDE provides a “super compression” capability. Instead of traditional compression ratios of 2:1 or 3:1, ratios of 30:1 and higher are possible. This level of compression makes using disk as the final repository for backup images even more cost-effective.

RDE is not a definitive backup-to-disk solution, but it clearly makes disk a more cost-effective option for storing backup images. Although RDE reduces the overall disk storage requirements and can drive down the WAN bandwidth needed to support disaster recovery, it can also slow recovery speed, especially when many streams of recovery are required simultaneously (such as during disaster recovery).

SPECIALIZED TECHNOLOGY

Other promising technologies being deployed for backup and recovery are Continuous Data Protection (CDP) and file services data protection.

Continuous Data Protection

In its purest form CDP tracks all the I/O operations of a given application so that no transactions are lost and the application can be rolled back, virtually instantaneously, to a previous point in time in a granular manner. The “holy grail” of CDP is an infinite number of restore points. As a result, the frequency of backup and restore granularity separate CDP from the core backup and recovery technology.

File Services Data Protection

Unstructured file data generally constitutes the majority of the data within an enterprise. This data might be centrally located within one or more large data centers and/or distributed across many branch offices and other satellite locations. Users typically access this data via a network file sharing protocol such as CIFS (for Windows) or NFS (for UNIX). Protecting this data is one of the major challenges facing IT organizations, because traditional core backup and recovery technologies have not been effective in protecting this type of data.

A relatively new class of File Area Network (FAN) technology has evolved to address the management—including backup and recovery—of distributed file services. These FAN solutions solve several key challenges:

- **Common, centralized address space:** A single view of all available file services is a key enabler for many follow-on functions. A common address space decouples the physical file servers and storage from the user view and enables the flexibility to effectively manage the backend physical storage environment. This is similar to the function storage virtualization provides for block-level access.
- **Data mobility and migration:** Storage and file server administrators must constantly move data as a result of normal operational performance and capacity management as well as technology refreshes. When users are tied to physical servers and storage, planned outages must occur on a regular basis. FAN solutions virtualize these file services and enable transparent, non-disruptive data migration.
- **Backup and recovery:** As with block-level access to storage, file server data must be backed up to meet recovery objectives. Large file servers and distributed file data present a significant backup and recovery challenge, and the standard mechanisms provided by the core backup and recovery technology are often unable to meet these objectives. FAN solutions provide backup and recovery technologies that specifically accommodate distributed and remote file data.
- **WAN bandwidth and latency:** File data distributed across remote geographic locations presents its own set of challenges, particularly when those sites are connected to the main data centers using low-bandwidth, high-latency network connections. These connections make centralized backup and recovery difficult if not impossible—meaning that local backup solutions are required. FAN solutions address these challenges by providing a centralized address space for file servers, local file caching via appliances, synchronization with centralized servers, and WAN optimization for IP-based file service protocols.

LEARN MORE

As the need for reliable, efficient backup and recovery continues to grow, today's IT organizations can enhance their existing infrastructures with strategic technologies, processes, and solutions. To learn more about Brocade backup and recovery offerings and how Brocade Services can help organizations implement the right solutions for their particular requirements, visit www.brocade.com.

Corporate Headquarters

San Jose, CA USA
T: (408) 333-8000
info@brocade.com

European Headquarters

Geneva, Switzerland
T: +41 22 799 56 40
emea-info@brocade.com

Asia Pacific Headquarters

Singapore
T: +65-6538-4700
apac-info@brocade.com

© 2007 Brocade Communications Systems, Inc. All Rights Reserved. 11/07 GA-WP-952-00

Brocade, the Brocade B-weave logo, Fabric OS, File Lifecycle Manager, MyView, SilkWorm, and StorageX are registered trademarks and the Brocade B-wing symbol, DCX, SAN Health, and Tapestry are trademarks of Brocade Communications Systems, Inc., in the United States and/or in other countries. FICON is a registered trademark of IBM Corporation in the U.S. and other countries. All other brands, products, or service names are or may be trademarks or service marks of, and are used to identify, products or services of their respective owners.

Notice: This document is for informational purposes only and does not set forth any warranty, expressed or implied, concerning any equipment, equipment feature, or service offered or to be offered by Brocade. Brocade reserves the right to make changes to this document at any time, without notice, and assumes no responsibility for its use. This informational document describes features that may not be currently available. Contact a Brocade sales office for information on feature and product availability. Export of technical data contained in this document may require an export license from the United States government.



BROCADE