Abstract

The dramatic drop in the price of hard disk storage combined with its performance characteristics has given rise to a number of data protection solutions that incorporate disk as a key component of their architectures. These architectures fall into five categories:

- **Continuous data protection:** this approach captures and stores on disk every change to the data volume being protected and requires an alternative backup strategy for medium to long-term data protection due to the volume of transactions being managed.

- **Disk to disk to tape (DDT):** sometimes called “virtual tape,” this approach takes advantage of the speed of disk by providing a short-term buffering or staging area for backup data prior to the data being written to tape.

- **Back-end compression:** this approach focuses on compressing the data to be protected once it is at the backup server, thus reducing the overall tape storage requirements.

- **Filer-based data protection:** this approach relies on the filer as a primary storage control point for all data being protected and then leverages snapshot and management techniques on the filer to provide short-term data protection. This approach still relies on some level of tape archive for long-term filer data protection.

- **Disk-based data protection:** this approach focuses on managing data compression at the source, and therefore it significantly reduces overall storage requirements and minimizes the network bandwidth utilization for full backups. Due to the efficiencies gained by this approach, disk can be utilized as the media for both short-term and long-term protection of data.

*Each approach has strengths, weaknesses and a different architectural fit within the enterprise data protection schema.* Which approach is right for your organization will depend on your corporation’s specific financial, operational and business goals.
Overview

Over the last two decades, computers have evolved from a simple productivity tool into the core infrastructure of the emerging digital economy. They have become indispensable for re-engineering processes and ushering in effective e-business practices. Today, workers require information access in order to do their jobs effectively. As a result, the need to protect information and the ability to restore any lost or corrupted data have become essential elements of corporate planning and long-term success.

In addition, organizations are increasingly setting goals for business continuity that approach “100% uptime,” where data is constantly online and accessible. A key component of near 100% uptime is the protection of data despite incidents such as software malfunction, system failure, or the accidental corruption of data. Managing data protection for the enterprise has become a strategic and challenging issue for companies, accounting for significant percentage of overall spending for IT organizations.

Despite the scale of investment in data protection, enterprises lack the tools to fully protect and manage their information assets. In an effort to improve business efficiency and reliability; data protection processes must be redesigned, systems overhauled and applications upgraded. Yet backup and restore solutions have remained largely unchanged for over a decade. Traditional backup and restore solutions account for nearly 30% of the total cost of data ownership, a staggering amount when you consider that these solutions:

- Fail to backup data more than 20% of the time.
- Fail to restore data 27% of the time from apparently successful backups.
- Do not protect at least 30% of a corporation’s critical data. This data often resides in remote offices or with mobile workers and is costly to bring into a centralized data center. This distributed data, however, is critical information and the amount of this data is growing rapidly.

While these numbers may have been historically tolerable, they are no longer good enough to satisfy emerging enterprise requirements. It is time for businesses to expect more from their backup and restore solutions in order to improve:

- **Integrity** – Businesses can no longer tolerate lower integrity for backups than for primary systems. The data for restores must be there when customers need it—with regular checks for integrity.
- **Speed** – It is no longer “speeds and feeds” that matter but process efficiency. Backups must be quick, non-disruptive and bandwidth efficient.
- **Scale** – Operational efficiency demands a consistent, integrated solution that can grow to meet new organizational requirements in data centers, in remote facilities and for distributed clients.
The root cause of the data protection challenge is data duplication. The natural result of traditional data protection and archival processes is to extract, send and store multiple copies of the same data over and over again. For most customers, weekly full backups can store the same data ten to a fifty times over the course of a year. Daily incremental backups and copies for reliability can double or triple this amount. Every one of these data “duplicates” is created using valuable server, network, storage and human resources, resulting in the escalation of the total cost of ownership. A typical organization with 1TB of primary data will need over 10TB of storage to maintain the necessary backups using a traditional backup strategy after just sixty days of operational backups.

Most enterprises use primary disk for storing current data and interacting with the applications and back up to tape on a periodic schedule. Typically this schedule is daily for incremental backups and weekly for complete backups. The problems with this scheme are well documented. Tape performance limitations make squeezing into backup windows and hitting recovery time objectives a near impossibility for most IT administrators. Backup reliability and the data integrity characteristics of tape are horrendous. The costs of managing a tape backup infrastructure are skyrocketing as the amount of data on secondary media accumulates. Even when a restoration from tape succeeds, the operation is extremely time-consuming.

Despite these significant issues and limitations, tape still remains the de facto backup medium. However, IT administrators are beginning to realize that the challenges associated with the traditional data protection solutions and the increasingly stringent requirements of today’s business climate have pushed the tape-centric backup architecture to its breaking point. In a November 2004 end-user study of next generation backup technology and trends, Goldman Sachs Global Investment Research learned that many administrators can no longer tolerate the limitations of tape. In fact, two-thirds of those surveyed (66%) indicated that they will start to deploy disk-based backup in the coming year.

Due to the dilemma with existing tape-based backup solutions and the continued decline in disk pricing, a number of data protection solutions that incorporate disk as a key component of their solution architectures are now available. The majority of these architectures, including continuous data protection, disk-to-disk-to-Tape, back-end compression, and filer-based data protection, provide incremental improvements to the backup problem yet still rely on tape media as the ultimate destination for protected data. As a result, organizations implementing these solutions must still deal with the significant issues and limitations of tape-based technology.

The only solution that has been designed from the ground up to take full advantage of the intrinsic benefits of disk technology to solve the backup problem is disk-based data protection. This solution provides a new way to protect data that can better accommodate the driving financial, operational and business imperatives of integrity, speed and cost-effective scale.
Continuous Data Protection

The advent of transaction journaling in database systems and the application of journaling concepts to modern file systems provided the genesis for continuous data protection solutions.

**Description**
In a continuous data protection solution, every write operation to a specified volume is captured and duplicated to a transaction log. This log may be used for audit purposes or may be replayed against a baseline to recreate an image of a volume as it existed at any point in time.

**Continuous Data Protection Strengths**
The primary strengths of this approach center on the continuous nature of transaction logging; this architecture does not utilize a backup window, as every write operation is logged in real-time. Transaction logging also delivers the benefit of being able to recover lost data back to any point in time, but due to the volume of data stored during the transaction logging process, typically only several days of data is kept online at any given time. Therefore continuous data protection solutions are primarily used for short-term rollback of selected high-value data.

**Continuous Data Protection Weaknesses**
Due to its limitations, continuous data protection is generally used for only specific applications and for a short window of protection. While every transaction may be logged under this architecture, it is impractical for operators to select which of the millions of writes captured was the “last good write.” Applications must sort through the transactions to determine consistent restore points, which can require significant indexing and post-processing of data. This architecture can also degrade performance, because it introduces latency in every write operation, since each write operation must be written to both the primary data location as well as the transaction log. This approach also consumes large amounts of disk storage on active systems which must exist on the primary data path of the server being protected. Finally, this approach is still reliant upon an alternative backup strategy for medium to long-term storage due to the amount of data generated in order to enable transaction level restores.

**Architectural Fit**
This solution is ideal for applications where users need the ability to restore data back to very specific points in time. Continuous data protection solutions are usually deployed in conjunction with another approach for backup and restoration.
Disk-to-Disk-to-Tape (DDT)

Tape-based architectures have been the mainstay of the data protection challenge thanks primarily to the initial cost difference in tape versus disk media. However, over the past decade, disk buffers have been introduced in order to provide faster backup and restores, without dramatically changing the tape backup architecture.

**Description**
DDT solutions or virtual tape libraries make a disk array appear to existing backup software like a tape library, without requiring the installation of new applications or a major redesign of existing backup processes. These virtual tape systems store a tape image of a full or incremental backup to disk before the data is ultimately written to tape.

**DDT Strengths**
The primary strength of this architecture is centered on the “status quo.” Specifically, no changes are required to the software environment and there is minimal impact to operational activities. The increased performance of disk-based tape emulation can shorten backup windows, and the greater reliability of disk storage significantly increases the probability of successful backups.

DDT also serves as an “impedance match” for network backups. Often networks cannot supply data at the fast, constant rate required by tape drives. When this happens, the tape drive will go into stop/start mode, which has a dramatic negative impact on the effective throughput of the tape device. The performance of disk-based tape emulation tracks linearly with network throughput.

**DDT Weaknesses**
Over the past decade, tape-based backup and archiving systems have become highly attuned to the constraints of tape systems. These constraints have led to many of the limitations we experience with backup today, including the requirement for repeated large data transfers, the difficulty of maintaining multiple indexes to track data location, and the relatively inefficient use of tape capacity. In fact, an effective backup strategy typically requires a tape library that is five to ten times as large as the amount of data being protected, in order to hold repeated full copies of the data in addition to incremental copies; thus an enterprise with 1 TB of data to protect can require a 10 TB tape solution to provide effective data protection. DDT solutions eliminate none of these constraints, since DDT solutions still rely on tape systems for medium to long-term data storage.
Other weaknesses with DDT solutions include the introduction of additional complexity and cost in the overall backup architecture with the deployment of another set of hardware and the increased overhead of managing the DDT solution. The usage of DDT solutions may also require the purchase of additional backup software licenses since typical backup software packages are licensed on a per drive and/or tape library basis.

**Architectural Fit**

DDT solutions are an extension of the current disk to tape solution implemented over the last two decades. DDT solutions provide an incremental improvement to a few of the problems that organizations face with tape-based backup solutions, including backups that take too long to complete, the low reliability of data restores, and excessive time and effort spent recovering data from tapes.
Back-End Compression

Back-end compression solutions are an evolutionary improvement of the disk-to-disk-to-tape solutions aimed at reducing the amount of disk needed for staging and the amount of tape needed for long-term archives.

Description
Back-end compression solutions seek to reduce the total data storage requirements by employing various data compression techniques. While reducing hardware costs, these solutions inherit both the advantages and disadvantages of disk-to-disk-to-tape solutions. Back-end compression solutions can be implemented as tape emulation devices or as NFS mounted file systems.

Back-End Compression Strengths
In addition to the strengths of disk-to-disk-to-tape solutions, this approach reduces the hardware cost through more efficient storage of the backup data and provides for longer retention of data on disk compared to disk-to-disk-to-tape solutions of the same capacity.

Back-End Compression Weaknesses
In addition to the weaknesses of disk-to-disk-to-tape solutions, this approach is only effective within the device the compression software is loaded on. Also, given the amount of CPU resources required to perform compression and decompression of data, there is a potential performance impact when introducing back-end compression into data protection architectures.

Architectural Fit
Back-end compression solutions are typically deployed in place of or in addition to disk-to-disk-to-tape solutions to reduce the overall hardware that must be purchased to support the DDT deployment.
Filer-Based Data Protection

Dedicated NAS filer appliances have proven to be a very stable and easy to administer platform for many storage applications. Given the appliance nature of these devices and the close integration and control of the underlying file system, there has been a natural evolution to supply logical volume management services such as block level snapshots and replication.

Description
Modern filer appliances often provide the ability to create block-level snapshots, point-in-time captures of the entire file server that can be used when restores are necessary. These snapshots are generally limited in number and stored on the primary storage device. In addition, some filer appliances enable block-level replication across a dedicated network or even WAN in order to copy these snapshots to another device or location.

Filer-Based Data Protection Strengths
The primary strength of this architecture is similar to disk-based solutions in that block level change detection is more efficient than incremental backups. Snapshot administration is integrated with filer administration, and this approach delivers highly reliable short-term data protection and online restore for data contained in snapshots.

Filer-Based Data Protection Weaknesses
In some cases, snapshots can overrun their storage allotment and reduce the overall storage available on the file server. In addition, snapshots work only for the file server in question; they do not work for heterogeneous systems with their own internal or external storage, which require their own data protection solution. Furthermore, most organizations still archive snapshots to tapes due to the cost of snapshot storage and the limitations on the number of snapshots supported by a given file server.

Architectural Fit
Filer-based data protection solutions are usually deployed in conjunction with another approach for backup and restoration. Filer-based data protection solutions provide many of the same benefits as DDT solutions, including reduction in backup windows and rapid restores for recent data. Many organizations utilize filer-based data protection to protect filer data in the short-term but still use traditional tape-based backup solutions to protect data for the long-term.
Disk-Based Data Protection

Tape backup has historically been the default choice for protecting enterprise data repositories. Most backup and recovery software packages, therefore, are built around tape media and automated tape libraries. But with its dramatic price declines, disk storage has become increasingly attractive as an addition or replacement for tape media in modern data protection solutions. Disk storage has a number of important advantages over tape media, including random access to stored data, higher reliability and faster access times. As a result, data protection solutions are quickly evolving to adopt hard disk storage as an integral part of backup and recovery for the enterprise.

**Description**
Unlike other architectures, disk-based data protection takes full advantage of disk storage as the final repository for backup and archival data. By re-designing the backup architecture around the inherent advantages of disk storage, disk-based data protection solutions can eliminate the need for tape storage altogether. Using content-addressable storage and other high-efficiency techniques, these solutions can store full copies of backup data in a fraction of the amount of space required for traditional full and incremental tape backups. For instance, with its intelligent client agents, Avamar Axion identifies and filters redundant data stored in files within a single system, across systems, and over time at each client so that each unique sub-file data block is only backed up once within the enterprise, reducing the overall data for a full daily backup that must be transmitted over the network and stored on the data protection server by over 99%.

Disk-based data protection solutions also offer fault tolerance capabilities through the use of conventional disk RAID or more advanced architectures, such as Axion Redundant Array of Independent Nodes (RAIN), providing far higher availability than other solutions. Disaster recovery can be provided by mirroring data across geographically distributed sites using solutions like Axion Replicator, which send very small amounts of data across LANs or WANs on a scheduled basis. With its efficiency advantages, disk-based data protection solutions can provide online access to all backup and archival data and replication, within the cost-envelope of traditional tape solutions.

**Disk-Based Data Protection Strengths**
The entire purpose of backup and archive is to enable the timely restoration of data when needed. Disk-based data protection provides online access to backup data
that can be immediately restored to a desired location. Disk-based backup solutions can also eliminate the challenges associated with long backup windows for traditional tape backups. By only sending the changed blocks of data on a daily basis, disk-based data protection solutions move quickly through the backup process and greatly reduce the utilization of network bandwidth for daily backups.

Another advantage of disk-based data protection solutions is the far higher level of data integrity when compared to traditional backup solutions. Because all backup data is available on disk, these solutions can perform regular consistency checks and validation, ensuring that data can be restored successfully when it is needed.

Lastly, by eliminating the necessity of tapes, these solutions significantly reduce the manual effort and cost of managing data protection.

**Disk-Based Data Protection Weaknesses**

Although disk storage has been used reliably in enterprises for decades as primary storage, enterprises must be willing to make fundamental changes in their data protection architecture in order to adopt a disk-based data protection solution. Often, these solutions are designed to replace existing backup solutions, and enterprises may be reluctant to change their data protection architecture. Also, because disk-based data protection solutions are relatively new, they may not provide the range of platform support available compared with legacy solutions.

**Architectural Fit**

Disk-based data protection solutions were designed from the ground up to take full advantage of hard disk storage as the final repository for backup and archival data, instead of retrofitting disk storage to fit into the existing tape-based backup paradigm. As a result, these solutions have a tremendous advantage in efficiency over existing backup solutions. Disk-based data protection solutions can provide backup, recovery, archive and replication to an offsite location for disaster recovery, all for the same cost as a traditional tape backup solution with offsite tape vaulting.
Summary

For several decades, IT administrators have coped with the operational challenges of managing tape-based backups. Today, cost-effective data protection solutions incorporate disk as a key component of their architecture, providing significant business and operational benefits.

These solutions can be categorized into five categories: continuous data protection, disk to disk to tape (DDT), back-end compression, filer-based data protection and disk-based data protection. Each of these solutions has its strengths and weaknesses. The first four solution architectures are additive in nature to the legacy backup solution, still relying on tape infrastructure to archive data over the medium to long-term. These solutions add complexity and cost for the benefits they provide. The only solution that replaces the existing tape infrastructure, significantly reduces the total cost of ownership of data protection and resolves the shortcomings of tape-based backup solutions is disk-based data protection.

Avamar has rebuilt the backup and restore process from the ground up to create the first disk-based backup and restore solution that delivers all the benefits of disk-based technology without the limitations of tape. Avamar offers these capabilities through the Axion line of products that offer complete solutions for storing, protecting, managing and accessing enterprise data. Today, Axion is setting new standards for data integrity, storage efficiency and cost-effective scalability. We invite you to see for yourself how Avamar Axion provides the benefits of disk-based data protection while reducing overall total cost of ownership. To begin an evaluation, please contact us at http://www.avamar.com/contactus2.html.