EMC²
where information lives®

White Paper

you

Backup-to-Disk Planning Options

Applied Technology

Abstract

This white paper provides a high-level overview of backup-to-disk, and defines what it is and how it compares to tape. Disk technology has become the new standard for backup, replacing tape as the primary backup medium.

August 2006

Part number H2312

Table of Contents

Executive summary

Data backup is a critical but dull IT function that can become quite stimulating if things go wrong. Making sure it goes right is a major challenge.

For decades most businesses have relied upon tape-based backup processes that have changed very little. Now, business demands have altered backup and recovery requirements and operational importance. As a result, disk-based backup has been recognized as a faster and more reliable backup and restore strategy. It has become an essential component of the best data protection and recovery plans.

Backup-to-disk has surpassed tape as the most desired backup solution because it incorporates significant benefits over traditional backup to tape.

Major advantages of backup-to-disk include:

- Faster restore performance
- Faster backup performance
- Enhanced media reliability and data availability
- Improved IT efficiency
- Improved backup and restore predictability

Most importantly, backup-to-disk has emerged as a powerful mainstream solution for rapid recovery of mission-critical data. Customers are adopting use of disk as the target for primary backups, and quite often for longer-term data retention as well.

Along with more stringent business demands, this direction was initially fueled by improved Advanced Technology-Attachment (ATA) disks and more recently by availability of Low Cost Fibre Channel (LCFC) drives. ATA and LCFC disk technologies offer economics comparable to tape, with the performance and reliability benefits of disk. The clear advantage of faster backup and recovery speed has helped disk-based solutions to overtake tape as the most desirable media for operational backup and restore. While tape is still in the mix, the role of tape has shifted primarily to offsite data retention.

Introduction

This white paper provides an implementation and planning overview for backup-to-disk solutions. It defines backup-to-disk, presents recent technology developments and outlines available options. Advantages and considerations are discussed regarding performance, data availability, efficiency, and integration with existing systems.

This white paper:

- Defines backup-to-disk
- Describes the use of disk technology in backup topologies
- Compares and contrasts disk and tape backup
- Highlights planning considerations for disk-based backup
- Profiles options from moderate to very large scale backup-to-disk solutions

Audience

This white paper is intended for IT planners, storage architects, storage administrators, backup administrators, EMC partners and customers interested in a high-level technical overview of backup-to-disk.

Backup-to-disk is mainstream

Typical backup software applications were designed to work with tape devices. If they did write to disk, functionality was often limited. More importantly, the relatively high cost of Fibre Channel drives made disk unaffordable in most backup situations. Tape was the medium of choice.

All that has changed.

Now, backup applications have been optimized for disk and the cost of large-capacity disks has dropped to near the cost of tape. Early adopters demonstrated how using backup-to-disk offered several advantages. In particular, performance gains have enabled faster backup execution and much faster data restoration than with tape. And with business constraints compressing or eliminating backup windows, backup-to-disk has meant that businesses can better meet production objectives. Because disk-based backups deliver fast, near-instantaneous restore times they provide the always-on access to information needed to recover business operations in minimal time.

Disk technology: Larger capacity and lower cost

The benefits from newer disk technology have enabled much of the movement toward backup-to-disk. This means the benefits of online disk approach the economics approaching tape. Disk capacities are much larger and cost-per-gigabyte is lower. This combination enables customers to keep more data online for longer periods of time where previous alternatives were not affordable or justifiable.

Now, with ATA and LCFC drives, IT professionals can simplify management and leverage functionality for both production and backup environments. For example, it is now feasible to mix high-performance Fibre Channel drives and very large capacity ATA or LCFC drives within a backup environment under common management. This can even be accomplished within a single array for combinations that provide considerable deployment flexibility and meet a wide variety of requirements.

Improving Business Productivity with Backup-to-Disk

Faster data backup and restores: Fit more aggressive backup windows and restores data in minutes, not hours or days

Reduced business risk: With built-in hardware redundancy, RAID protection, and high availability assurance of fast and accurate data restoration

Increased application uptime: Fit to smaller backup windows with less impact on application and system availability.

Improved IT efficiency and productivity: Saves hours of staff time previously devoted to handling tape

Using these advances, customers can consolidate applications and take advantage of all of the advanced business continuity and system availability features provided in sophisticated storage arrays. Beyond backup-to-disk, customers also gain flexibility and broader choice for other applications using tiered storage for information lifecycle management strategies.

Backup-to-disk approaches

There are essentially two ways that backup-to-disk can be accomplished: backing up to a disk file or using emulation for backup to virtual tape.

Backing up to a disk file

Writing backup data directly to a disk file in a file system is an approach that has matured over the past few years. Nearly every commercially available backup application now supports the ability to "back up" data to disk in this fashion. Using this approach, data is often written in tape format to disk. Depending on the

application, various parameters within the backup application may have to be set to perform a backup-to-disk operation, and to direct the backup to go to the file system on disk.

High-speed Fibre Channel, LCFC or ATA disk technology can be used with this approach, with each having its unique benefits. Depending on the specific backup application to be used, higher speed Fibre Channel drive technology may be the best fit for the initial backup operation where maximum performance is required – and where the data is expected to be kept on the Fibre Channel disk for only a limited amount of time (hours or a few days).

LCFC or ATA technology fits well where good disk performance is still required, and the retention period of the data will be days, weeks, or perhaps months in some scenarios.

This combination also permits the use of a multi-tiered storage approach where backup data that is sent to high-speed Fibre Channel drives for a brief period can then be moved to LCFC or ATA technology for longer term retention. In most backup environments, LCFC or ATA drives are recommended as the technology choice and will satisfy the bulk of backup-to-disk requirements.

Emulation or virtual tape

The second approach, referred to as emulated or virtual tape, writes backup data to a disk that is presented as a tape device to the backup application. The backup application "sees" tape devices and writes the backup data to these devices just as if they are physical tape devices when they are actually disk. In this scenario, the disk is front-ended with a tape emulation engine that presents the disk array storage as any number of tape devices. Advantages of disk technology are maintained, while preserving the operational consistency of traditional tape backup methodologies without the disk management involved associated with the first approach above.

This white paper focuses primarily on the first approach, where data is written to a file system on disk.

Advantages of backup-to-disk

Tape has long been the backup medium of choice because of cost-per-MB advantages over disk. Now, disk economics have narrowed that gap. For example, EMC's storage implementations can economically combine high speed and large capacity drive technologies for production as well as backup scenarios in EMC® CLARiiON® or Symmetrix® storage arrays.

The advantages of using disk over tape with backup solutions can be grouped into four major categories:

Restore performance

- Faster recovery time is often realized for disk drives over tape devices. The difference can be seconds or minutes versus hours with tape.
- Disks support random and sequential access. Tapes support sequential access only. This enables faster access of data files on disk, improving overall performance.
- When data is on multiple tape media, time must be spent on loading and unloading the media into drives, positioning the tape to the appropriate point to start actual data read operations, rewinding tape, and other tasks. Time to first byte can be milliseconds for disk, versus seconds or minutes for tape.

Backup performance

- Large capacity drive technology can be faster than many of the new-technology tape drives such as SDLT and LTO.
- Fibre Channel drive technology can be faster than many tape devices – particularly when used as a primary storage pool.

- Tape technologies often respond to a minimal data stream by excessive positioning or "shoe-shining." This physical behavior can significantly reduce a tape drive's performance. Disk does not experience this behavior because disks are random access devices.
- Robotic/mechanical movement of tape devices (libraries, drive load/unload operations, and so on) adds time to an overall backup operation. Disk technology does not suffer from these delays.

Media reliability and data availability

- Disk RAID protection enhances data availability and prevents data loss in the event of a disk drive failure, whereas tape-specific media errors are common.
- Tape handling is reduced or eliminated. Maintaining tapes from a tape library can be problematic and involves issues such as manual management, offsite shipment, and even potential loss.

Overall IT efficiency

- Disk does not require tape handling/positioning.
- Because of increased reliability, there may be less need to perform frequent full backups. When fewer backups need to be performed, network and CPU cycles are minimized.
- Because of the fast access and reliability, incremental backups become truly practical. There is no risk of missing or unreadable tapes that make all subsequent backups useless.
- Tape undergoes a technology shift every few years, so a conversion process from old to new media must be undertaken at that interval. Disk technology does not go through these transitions.
- New larger capacity disk drives reduce space requirements compared with equivalent-capacity tape libraries.
- More data can be kept online, for a longer period of time, due to the improved cost-per-MB of disk technology.

Comparing disk- and tape-based configurations

Because tape is a sequential-access medium, it is not possible to perform both a backup and a restore operation using the same tape drive at the same time. So, if a restore must use tape media that is already in use for a backup job, either the restore must wait for completion of the backup, or the backup must be aborted. Disk, on the other hand, is a random-access medium and does not suffer from the simultaneous write/read limitations of tape. With disk, it is possible for backup and restore operations to occur simultaneously when needed.

To maximize the investment in tape backup devices, many enterprise-class backup software applications support backing up multiple streams of data to the same tape device concurrently. This process is called *multiplexing*. With multiplexing it would be possible for a host system with, say, multiple data disks to back up all disks simultaneously to the same output device. This helps keep the tape device streaming – and can improve overall backup performance to tape. However, during restore operations, multiplexed data on tape may be spread out over a large portion of the tape, or even multiple tapes, requiring a longer period of time to fully read the data. The net effect is poor restore performance.

With disk used as the backup target, restoring multiplexed data does not negatively impact restore performance due to the random access nature of disk. In fact, it is no longer necessary to multiplex data with disk in the first place.

When comparing and contrasting the performance of a backup-to-disk implementation to a backup-to-tape implementation, both the throughput performance and overall time to complete a backup or restore operation must be considered. There are large differences between tape and disk in the overall time it takes to perform a backup or restore task. Figure 1 and Figure 2 highlight how the restore times can vary from a tape- and disk-based environment.

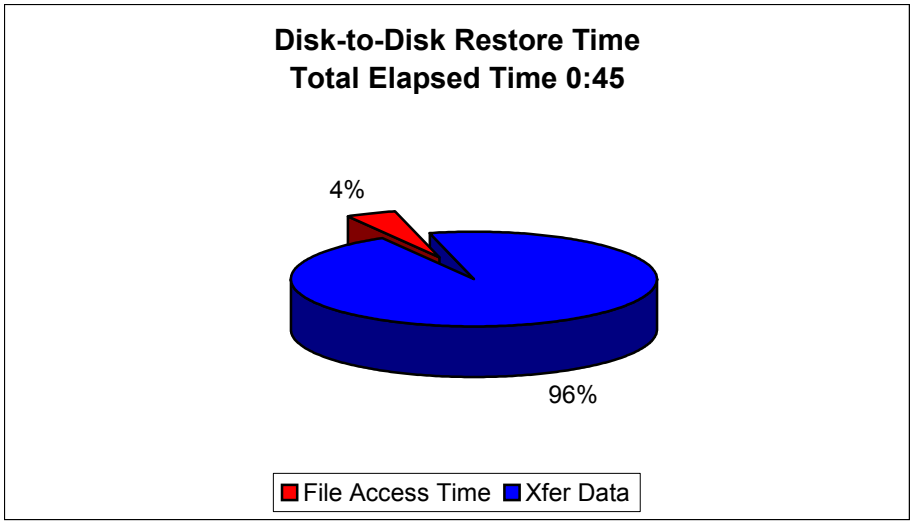

Figure 1. Disk-to-disk restore time

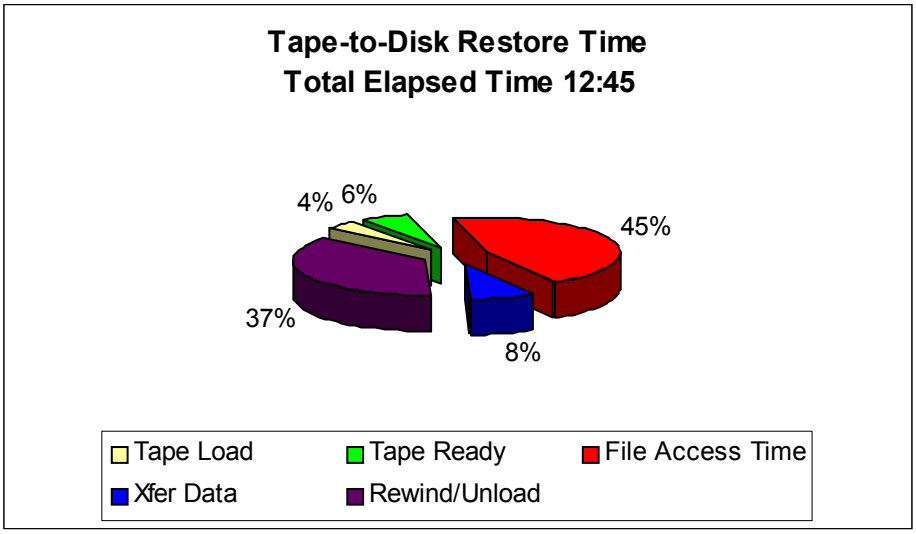

Figure 2. Tape-to-disk restore time

This example shows a scenario where a 1.5 GB data file is restored. As Figure 1 shows, the disk configuration was able to restore the data in about 45 seconds. In the tape recovery scenario in Figure 2, it took more than 12 minutes to complete. In the tape example, a significant amount of time was consumed by tape overhead operations, which included the fact that the requested data was located on several different pieces of media (three different tapes in this case) that all had to be loaded and unloaded. This overhead must be taken into account when comparing and contrasting tape versus disk performance.

Recovery time objective, reliability needs, and footprint requirements all must be considered when determining an optimal backup solution.

Cloning and replication

When enterprise-class storage arrays, such as EMC's CLARiiON and Symmetrix, are used as backup targets they address the many issues found in tape and also offer many features that tape cannot directly provide. These include:

- RAID protection that protects data from the failure of a single-disk drive
- Snapshots, mirrors, and clones that can provide rapid and near-instant backups and restore of business continuance volumes (BCVs)
- Remote data replication using array-based software applications eliminates the need to manually manage tape movement and improves physical data security by eliminating the risk of lost tapes.

Storage array disks, in addition to being a target for backup-to-disk, can also be used in variety of replica-based backup scenarios. This is a very powerful capability that depends upon storage array functionality.

For example, array-based software options are available for point-in-time copies and clones/BCVs of production volumes to be created by the array. These "copies" of the data can then be used as the source data by the backup application for normal backup operations. As a result, backup operations can be executed without impact to performance in production environments.

These operations can be performed independently from the backup application. Depending on the specific backup application however, it may be possible to invoke the array-based snapshot and replica capabilities from within the backup application GUI. For example, EMC's NetWorker™ provides extremely fast and flexible backup and recovery capabilities with some advanced options that include powerful snapshot management, advanced backup-to-disk functionality, and dynamic drive sharing. This creates a tightly integrated and centrally managed backup operation.

Conclusion

Backup-to-disk has emerged as a solution with significant benefits over traditional tape backup approaches. Changes in disk technology make backup-to-disk solutions economically feasible and technically straightforward. Business adoption has made this a mainstream solution, and disk solutions have improved existing backup infrastructures. More importantly, they have made business operations more effective and cost efficient.

Major advantages of backup-to-disk include:

- Faster restore performance
- Faster backup performance
- Enhanced media reliability and data availability
- Improved IT efficiency
- Improved backup and restore predictability

With all leading backup software applications supporting backup-to-disk functionality, the combination with Symmetrix and CLARiiON disk storage systems enable businesses to better meet stringent backup and critical recovery time.

This works because, when taken as a whole, disk backup is more affordable — and because the cost of downtime is totally unaffordable.

References

Specific performance and implementation details are available in expanded white papers tailored to individual backup applications. Please refer to these papers on EMC Powerlink for additional detail:

- *Backup-to-Disk Building an Effective Long-Term Strategy Applied Technology*
- *EMC CLARiiON Backup Storage Solutions: Backup-to-Disk with CA's BrightStor ARCserve Backup*
- *EMC CLARiiON Backup Storage Solutions: Backup-to-Disk with CommVault Galaxy*
- *EMC CLARiiON Backup Storage Solutions: Backup-to-Disk with EMC NetWorker*
- *EMC CLARiiON Backup Storage Solutions: Backup-to-Disk with EMC Retrospect*
- *EMC CLARiiON Backup Storage Solutions: Backup-to-Disk with VERITAS BackupExec*
- *EMC CLARiiON Backup Storage Solutions: Backup-to-Disk with VERITAS NetBackup*
- *EMC CLARiiON Backup Storage Solutions: Backup-to-Disk with IBM Tivoli Storage Manager*