EMC²
where information lives

DISK-BASED DATA PROTECTION:

The New Data Backup and Recovery Imperative

DATA BACKUP HAS BEEN A CHALLENGE FOR DECADES: slow, costly, labor-intensive, and unreliable. The traditional medium of choice, tape, has its limitations—it is mechanical as well as serial. Although vendors have made great strides in improving the hardware and software underlying tape backup, the inherent nature of this technology limits its effectiveness, especially in today's environment.

Enterprises, however, are enhancing their tape backup strategies with new disk-based options. By tapping the capabilities of disk, such as concurrent read/write and random access, enterprises can complement their tape backup strategies to achieve faster and more reliable backup and recovery while continuing to use tape for what it does best—long-term archiving, for example.

A number of factors are driving organizations to rethink their tape backup strategies. The explosion of data alone—industry research firms estimate annual data growth of 50% or more—is leading enterprises to augment backup solutions.

Regulatory mandates in many industries now require that organizations protect critical data and ensure its accessibility. In other cases, organizations are under great pressure to guarantee that they can not only meet more and more aggressive backup windows, but also recover from failures, disasters, and other emergencies quickly and reliably, often within minutes rather than days or even hours. Competitive pressure and the imperatives of the 24x7 global marketplace also drive the need for robust business continuance strategies in which data backup and recovery play a key role.

Heightened interest in corporate and IT governance, and the corresponding pressure to improve operational efficiency, are driving organizations to review every business and IT process with an eye toward speeding it up and reducing its cost. Data backup and recovery is a prime target for this kind of effort. Some analysts estimate that backup accounts for over 50% of IT's time. Similarly, management is setting demanding service-level requirements for a wide range of processes and measuring performance through service-level agreements.

All these factors converging on the IT world have led the storage industry and storage strategists to develop new approaches to data backup and recovery. New technologies, such as low-cost ATA disk and enhanced backup software, are giving IT management new options for meeting heightened backup and recovery requirements.

NEW OPTIONS FOR BACKUP AND RECOVERY

The new backup and recovery options revolve around the emergence of low-cost disks. "Disk capacity is growing really big and the cost of disk capacity is dramatically falling," says Bob Lyon,

co-founder of LEGATO, leading provider of backup and recovery software NetWorker, acquired by EMC in 2003. Taking advantage of greater capacity and lower pricing, storage vendors and strategists have been designing new backup solutions, such as backup-to-disk and tape library emulation, also known as virtual tape (for example, the CLARiiON Disk Library).

These new options promise a number of benefits over conventional tape backup:

- Faster data recovery, which gets the organization up and running sooner
- Quicker backup, which addresses the challenge of diminishing backup windows
- Increased confidence in the backup due to redundancy, RAID, and other high-availability features
- Greater reliability due to the better reliability of RAID protected disks compared to tape

In short, the new disk-based backup options finally solve the problems that have been plaguing conventional tape-based backup for decades.

THE BUSINESS CASE FOR DISK-BASED BACKUP

The payback from disk-based backup and recovery comes from the increased speed and efficiency, greater reliability, and reductions in labor it confers over tape-based approaches. Specifically, disk-based backup delivers payback in the following ways:

- **Faster and more reliable recovery and restoration—** Organizations can recover their data and be up and running more quickly, which reduces the period of non-productive time.
- **Shorter backup times—** Organizations can back up their data faster, which shortens the length of time their business applications are offline or impacted while data is being backed up. In many cases, the need for overnight backup windows is greatly reduced or altogether eliminated.
- **Improved reliability and confidence in the restored data—** Managers can be certain that the data they backed up will be recovered in an immediately useful format due to the various reliability and high-availability features of RAID disk. As a result, organizations will avoid the labor-intensive, time-consuming task of having to repair and even recreate corrupted data before it can be used. The risk of a single cartridge failure compromising a restore is removed.
- **Streamlined management and administration—** Managers can more easily administer data backup and recovery centrally. At the same time, the need for cumbersome and often costly manual tape handling is greatly reduced.
- **Investment protection—** Organizations can leverage existing disk capacity and existing SAN, NAS, and LAN infrastructures. Disk-based backup can leverage the same industry-standard disk already employed throughout the enterprise.

The savings from disk-based backup can be significant. It is estimated that a company using tape will endure 30 days of unavailability per year to execute backups [(5 hours per full backup x 52 weeks) + (1.5 hours per incremental backup x 6 days x 52 weeks)]. A backup performance improvement of 30% from disk would improve availability by nine days. That's nine more days a year that revenue-generating applications are available and employees are productive.

Because most companies have become heavily dependent on computer systems to conduct normal business operations, each hour of "downtime" can have a significant negative effect on the bottom line of the organization. It is estimated that the average cost of downtime as a function of idle labor for a large enterprise is more than $1 million an hour (Meta Group). For companies such as financial or telecommunication firms, the possible impact of delays in recovery of just one mission-critical server often justifies the investment in a disk-based backup solution. Companies have a lot at stake not just from a productivity and revenue standpoint, but also in terms of intangible costs such as customer and partner confidence, company reputation, shareholder value, and employee morale.

The payback from disk-based backup is so compelling that more than half (53%) of the managers interviewed in a recent study by the Taneja Group reported intentions to add disk-based backup in 2004. "The historic disparity between tape and disk prices would have made such a strategy cost-prohibitive several years ago. But with the emergence of ATA-based arrays and other low-cost disk solutions, the economics are now far more palatable," writes Arun Taneja, Taneja Group senior analyst, in a profile of EMC Legato NetWorker. Although disk-based backup

will carry a small cost premium, at least for now, it will be more than offset as organizations "achieve significant improvements in backup performance and reliability, restore time objectives, and data integrity," he adds. Total cost of ownership is the key—not acquisition cost.

DISK-BASED BACKUP: TWO ENTERPRISE OPTIONS

Two primary enterprise options have emerged for disk-based backup: tape library emulation, and backup-to-disk. Although the two approaches are very similar and use the same underlying disk and backup software, each has different characteristics that will make it more appealing to an organization based on its specific requirements and situation.

An important distinction to consider is that disk-based backup is focused on operational backup and restore, not archiving. Tape still delivers significant advantages in terms of the removeability and mobility of the tape medium for purposes of offsite storage and its proven ability for long-term archiving. Instead, disk-based backup options are intended to augment tape backup, allowing for faster, more reliable, and more efficient backup and recovery of the content that today resides on-site. Disk-based backups can then be cloned or staged to tape for shipment to an offsite location for disaster recovery without impacting daily operations.

TAPE LIBRARY EMULATION

Tape library emulation is a physical disk solution that emulates a tape library. This disk capacity appears to the organization's backup software and applications as a physical tape library. From a backup point of view, the emulated tape library is seen as a standard backup target. As a result, no changes are required to the backup software, the applications, or the data backup process. The only noticeable change is that

Disk-Based Backup Offerings

	BACKUP-TO-DISK	TAPE LIBRARY EMULATION
Environment	Direct Attached LAN-based with NAS SAN-based arrays	Direct Attached LAN-based SAN-based
Backup and production data can be on same disk array	Yes – Multi-use.	No – A dedicated appliance, in most cases.
Backup Application Changes	May require licensing and configuration changes.	Minimal or no changes.
Simultaneous Read/Write on same file or virtual device	Yes	No
Disk Administration Requirements	Storage management, virus scanning, and ongoing tuning.	Storage is automatically configured and tuned.
Offsite Tape Creation	Via backup application staging, clone.	Via backup application clone, virtual eject, or appliance functionality.
On-board Compression	Typically not included. Check vendor specs.	Typically included. Check vendor specs.
Data Replication	Supported.	Typically not supported.

data backup and recovery happens significantly faster.

The primary advantages of emulated tape over physical tape are drastically improved backup and restore performance, significantly enhanced reliability, and improved manageability. Disk-based data can be accessed directly; the serial characteristics of physical tape make this impossible. Tape emulation can be substituted for a physical tape library in plug-and-play fashion. In addition, emulated devices do not suffer from mechanical delays such as robotic tape picking, loading, and tape fast-forward and rewind. Searching for data on a tape becomes an immediate operation through the advantages of disk's random access capabilities.

Another advantage of tape emulation is the ability to create more tape drives than were previously available in the physical environment. By providing additional tape drives (and thus tape targets) to an existing backup environment, the backup administrator can re-deploy backup jobs that may have been waiting for tape drive resources to become available.

The backup administrator may also choose to re-deploy backup jobs that previously shared tape drives, a complex operation that is no longer necessary with emulated tape libraries. Eliminating tape sharing, which is no longer necessary to improve overall backup performance, will optimize restore performance. Running more backup jobs concurrently also allows administrators to run additional backups in parallel and complete more backups in less time.

In the end, eliminating the physical tape drive as a major constraint increases backup bandwidth and reduces backup times. At the same time it is increasing backup and recovery performance, tape library emulation improves reliability and predictability by eliminating the potential for mechanical failure, which is a constant danger in physical tape backup.

BACKUP-TO-DISK

Backup-to-disk in which disk is presented as a file-system-based volume is similar to tape library emulation in many respects. It uses increasingly low-cost disk capacity to back up data more quickly and reliably and, most importantly, restore the data quickly and reliably. In this case, the backup application treats disk as disk. Although it performs a job similar to backup-to-tape, it allows the organization to take full advantage of the inherent capabilities of disk, especially that of random access, through which data can be retrieved in non-sequential order, and concurrent read/write, in which the disk can read and write data on different parts of the disk at the same time.

As a result, treating disk as disk provides several key advantages:

- Better performance than tape for backups, clones, and restores.
- The ability to restore data from a completed save-set while the backup to the same group on the same device is still in process, as with EMC Legato NetWorker DiskBackup™ Option (DBO), is not possible with tape. For example, although a full backup may take eight hours to complete, an issue may arise at a point, say, six hours into the backup, that requires a restore of something that was backed up earlier in the process. When backing up to disk as disk, you can restore the backed-up data immediately without having to interrupt the ongoing backup process.
- Cloning, the process of making a mirror copy of stored data, likewise can begin on any completed save-set even when the full backup process is still running (another unique feature of NetWorker DBO). Traditionally, cloning could only begin after completing all save-sets in a group. But by taking advantage of concurrent read/write capability, backups and clones can finish in nearly the same time it normally takes to complete only a backup. Once cloned, if a restore from tape is ever necessary, data can be recovered directly to the affected clients in a single step.

Backup-to-disk generally requires some changes to the backup process. The backup software must be reconfigured to see the disk as a backup target and treat the disk as disk so the organization can take full advantage of concurrent read/write and other disk capabilities. The changes, however, are modest, and the latest backup software, such as NetWorker, supports a wide range of options, including disk as disk and disk as emulated tape.

CHOOSING A DISK-BASED BACKUP OPTION

To determine which approach is right for a given organization, managers need to look at the environment, operational concerns, and budget considerations. The advantages of disk-based backup are so compelling that managers can't make a wrong choice in adopting a disk-based approach to

backup. However, though either solution can likely fit into any environment and produce desirable results, organizations may find they favor one over the other based on specific benefits. The following guidelines will help steer managers to the best option for their organizations.

Basically, backup-to-disk works best where organizations:

- Are comfortable and experienced in disk management
- Want to take advantage of the flexibility and advantages of being able to partially restore or clone before a backup process is fully completed
- Have existing disk within an environment that can be allocated to backup (or have an array to which ATA drives can be added)

The tape library emulation approach works best where organizations:

- Are comfortable and experienced with the tape paradigm
- Want to gain disk speed advantages but choose not to alter their standard tape management practices
- Want the advantages of disk backup extended to their NAS NDMP backup infrastructure
- Have backup teams that cannot make a traditional disk decision but do have a budget for backup media

As noted above, disk-based backup is not intended to completely eliminate tape. Tape may still be required for offsite storage and for long-term archival purposes.

"Given the current pressures of shrinking backup windows, the need for backup reliability, and aggressive objectives for restore times, most backup administrators have come to the realization that traditional backup technologies and methods no longer cut it. Customers are in serious pain and are looking for disk-based backup solutions to solve many of their backup woes," says Taneja.

EMC is in a unique position to help organizations determine where and when to choose backup-to-disk and tape emulation solutions. With hardware choices ranging from tape library emulation—the CLARiiON Disk Library—to a range of ATA-based arrays, including CLARiiON CX and Celerra NS, and software choices including NetWorker and NetWorker DiskBackup Option, EMC can support virtually any disk-based backup deployment. With NetWorker, customers can choose a tape library emulation environment, a backup-to-disk environment, or a mixture of the two.

Taneja concludes, "Overall, with NetWorker 7.0 and 7.1 and EMC behind it, we expect that users will reward Legato for its timely technological innovation."

To learn more about the latest in high-performance disk-based backup and restore, click now on www.legato.com/diskbackup.

Frequently Asked Questions

How do I know my organization is ready to consider a disk-based backup solution?
Disk-based backup is helpful if your organization depends heavily on tape for backup and recovery; if you are looking to purchase or upgrade tape libraries; if you can't meet service levels for backup/restore; or if you are encountering backup performance, reliability, or management issues.

What changes are required to adopt disk-based backup?
Adopting a disk-based backup requires very few or no changes, depending on your approach. A tape library emulation approach is simply plug-and-play; no changes to your backup process are required. A backup-to-disk approach typically requires minor configuration and licensing changes.

Do I have to choose between a backup-to-disk and a tape library emulation approach?
With some products you will have to choose, but with EMC Legato NetWorker 7.x you can pursue both backup-to-disk and tape library emulation strategies.

How can disk-based backup help us meet SLA requirements?
Disk-based backup allows for faster, more reliable recovery. Depending on your approach, it can enable concurrent recovery even while backup may still be under way, which improves responsiveness. Finally, it allows direct access to target files, eliminating the time typically consumed by overhead of mounting, un-mounting, and positioning tapes. All these factors help meet the most demanding SLA commitments.

Can I use low-cost ATA disk for backup?
Yes, these disk technologies are ideal for disk-based backup. EMC Legato NetWorker not only supports ATA disk but also SCSI and FC, should you have excess capacity in these disk formats.

ABOUT EMC

EMC Corporation (NYSE: EMC) is the world leader in information storage systems, software, networks, and services, providing automated networked storage solutions to help organizations get the maximum value from their information, at the lowest total cost, across every point in the information lifecycle. Information about EMC's products and services can be found at www.EMC.com.

ABOUT LEGATO SOFTWARE FROM EMC

Legato software from EMC sets the standard for open, enterprise-class software solutions and services for information management—helping organizations simplify management and control of information and enable business continuity and compliance at the lowest TCO. Legato software's information protection, automated availability, and messaging management solutions are delivered through a worldwide network of strategic partnerships and alliances, as well as a direct sales force.

For more information about Legato storage management products, visit www.emc.com/legato or call 888.853.4286 (outside the U.S.: +1.650.210.7000).

Copyright (c) 2004 EMC Corporation. All rights reserved.
EMC believes the information in this publication is accurate as of its publication date. The information is subject to change without notice.

THE INFORMATION IN THIS PUBLICATION IS PROVIDED "AS IS." EMC CORPORATION MAKES NO REPRESENTATIONS OR WARRANTIES OF ANY KIND WITH RESPECT TO THE INFORMATION IN THIS PUBLICATION, AND SPECIFICALLY DISCLAIMS IMPLIED WARRANTIES OF MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE.

Use, copying, and distribution of any EMC software described in this publication requires an applicable software license.

EMC², EMC, and [Documentum, Legato, or EMC product names] are trademarks of EMC Corporation. All other trademarks used herein are the property of their respective owners. All other brand names are trademarks or registered trademarks of their respective owners.