

## **Next-Generation Backup-to-Disk**

### *Building the Foundation for Effective Operational Recovery Management*

**Abstract:** This paper examines the issue of backup-to-disk (B2D) and the central role it plays in operational recovery management. It identifies five immediate benefits of B2D, from faster and more reliable backup and restore to increased efficiency when backing up both SAN and NAS. Similarly, it describes five ways IT can take advantage of B2D to improve operational recovery. Finally, it reviews the key criteria an enterprise B2D product should support—concurrent read/write, Network Data Management Protocol (NDMP), direct file access, and single-step recovery.

Date 1/31/2005

Copyright © 2005 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.

EMC<sup>2</sup>, EMC, Legato, and NetWorker are registered 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.

65010105V1

# Table of Contents

<b>The benefits of backup-to-disk .....</b>	<b>4</b>
<b>Putting B2D to work.....</b>	<b>5</b>
<b>Key features to look for in B2D solutions.....</b>	<b>6</b>
Concurrent read/write .....	6
Network Data Management Protocol.....	6
Direct File Access (DFA).....	7
Single-step recovery .....	7
<b>Achieving the next level of operational recovery management.....</b>	<b>8</b>

IT is feeling the heat. More data, more types of data, and different storage formats, combined with shrinking backup windows, require sophisticated operational recovery management. Meanwhile, demands for ever-faster recovery further compound the challenge.

Storage recovery management is a broad strategy that encompasses archiving and compliance, disaster recovery, and operational recovery. Operational recovery management, which addresses daily backup and recovery needs, presents unique challenges in terms of speed, performance, and service in today's demanding enterprise environment. Traditional tape backup and recovery alone cannot meet the current requirements of operational recovery management. Instead, IT must move to the next level of backup and recovery.

For this reason IT managers are turning to redundancy, RAID, built-in high-availability features, and enterprise backup-to-disk (B2D) that encompass both SAN and NAS. In short, they are turning to disk-based backup and recovery to deliver effective operational recovery service levels. According to a report from the Hopkinton, Mass.-based Taneja Group titled "EMC NetWorker 7.2: Continued Leadership in Disk-Based Data Protection" (January 2005), "A majority of the end users surveyed (53%) indicated that they would be evaluating or deploying disk backup technology within 12 to 18 months."

Fortunately, the emergence of low-cost disks, such as Advanced Technology-Attached (ATA), makes it feasible to achieve this higher level of backup and recovery through B2D. By deploying enhanced backup and recovery tools such as EMC® Legato® NetWorker® 7.x, under a comprehensive operational recovery management strategy that includes both SAN and NAS storage, IT managers can finally provide advanced operational recovery management.

## The benefits of backup-to-disk

The key to operational recovery management is B2D. B2D entails using a backup application to back up data to a disk as opposed to tape. B2D goes beyond mirroring, replication, or a simple copy of existing data on disk. Rather, it plays a central role in the organization's operational recovery management effort.

According to the Taneja Group, "Tape alone is no longer a sufficient backup medium for business-critical applications. Consequently, backup software tools must now have the flexibility to fully support backups of data to a secondary disk medium in addition to regular tape."

The steadily falling price of ATA now makes it economically feasible for IT to back up data to disk instead of conventional tape. IT can store the backed up data sets on arrays of low-cost, reliable ATA disk, taking advantage of the performance characteristics of disk.

Through B2D, organizations will experience a number of benefits:

- Predictable operational backup and recovery due to the greater reliability and consistency of disk storage
- Faster data recovery as a result of the random access capabilities of disk and direct file access
- Faster data backup because of the greater performance of disk
- Increased flexibility due to concurrent read/write capabilities available in advanced B2D products like EMC Legato NetWorker 7.x
- Increased administrative efficiency through the use of a single backup tool for both SAN and NAS backup and recovery

These advantages allow the IT organization to consistently meet more demanding recovery-time objectives (RTOs) and recovery-point objectives (RPOs). Due to the reliability and consistency of disk backup, IT can

deliver higher-quality protection and recovery. B2D eliminates the errors and problems that, despite advances in tape technology, continue to hinder tape backup.

In addition, B2D allows the IT organization to increase the ROI of tape storage itself. Tape can now be used in the roles for which it is ideally suited, such as the primary offsite disaster recovery medium. This takes advantage of the portability of the tape medium. IT can also buffer tape operations with B2D and thereby reduce the impact of tape backup on business processes, which often have to wait while tape backups are completed.

## Putting B2D to work

B2D has emerged as a powerful addition to any operational recovery management strategy. Highly flexible, B2D gives IT numerous options:

- Staged backup to reduce impact on production systems—by backing up to disk before copying the backup to tape, organizations speed backup while continuing to use tape for offsite storage
- Comprehensive enterprise backup—backing up both SAN and NAS data through a single, streamlined B2D backup tool
- Backup consolidation—backing up multiple diverse systems with a single B2D backup and recovery product
- Selective recovery—recovery of selective data very quickly rather than having to wait through a complete restore process
- Fast recovery—using concurrent read/write capabilities to recover some files even while others are still being backed up

B2D does not usually eliminate the use of tape backup. To the contrary, most organizations continue to use removable tape media to meet a variety of data retention and disaster recovery requirements because of its portability and low cost. B2D, however, can be deployed in a way that is nondisruptive to the existing systems environment and flexible enough to fit within a company's backup window and disaster recovery guidelines.

In fact, combining B2D with tape adds immediate value, particularly in the area of restore. Traditionally, customers had to choose between fast backups and fast restores. In order to maximize backup throughput and avoid both tape delays and the wear and tear of tape drives, backup administrators tune the environment to keep the tape drives streaming during backup. This entails configuring the backup application to allow multiple streams (multiplexing) of data from multiple sources to be interleaved into a single data stream sent to a device. With multiplexing, the drives are kept spinning, thus offering the best performance with the least amount of impact on the tape drives. A price is paid, however, when it is necessary to recover. To restore a specific piece of data the backup application must demultiplex or unwind the data stream that is intermixed with multiple sources. As a result, the more the environment is tuned to provide fast backups, the more recovery performance suffers.

With B2D the organization can avoid having to make this choice by backing up first to disk. B2D continues to enable multiple streams to be sent to a device concurrently, but on disk it no longer interleaves these streams. Each is written contiguously. This maintains backup performance while eliminating the need to demultiplex data for a restore. In effect, administrators can tune for both backup and recovery performance.

Once backed up to disk, the data can be subsequently moved from disk to tape through cloning or staging. Both cloning and staging take advantage of the concurrent read/write capabilities of advanced B2D backup software like NetWorker 7.x to speed the process. Restoring from tapes that have been created by staging or cloning after data has first been sent to disk is consistently faster than restoring from data sent directly to tape because demultiplexing is avoided, as noted above.

## Key features to look for in B2D solutions

Not all backup software can handle B2D. This is especially true when IT wants to treat the disk as disk, not tape. Furthermore, even those that do claim to treat disk as disk often may not offer the full set of features required to get the most out of a B2D strategy. According to the Taneja Group, “Backup-to-disk capabilities will be one of the key criteria, if not the top criterion, against which backup products will be evaluated.”

Currently, four key features set apart top-performing B2D backup products:

1. Concurrent read/write
2. Network Data Management Protocol (NDMP)
3. Direct File Access (DFA)
4. Single-step recovery

The following sections look at these features more closely.

### ***Concurrent read/write***

Advanced backup products like EMC Legato NetWorker 7.x can handle read and write operations simultaneously. Such concurrent read/write gives the organization several advantages.

First, it allows for fast restores even while the backup is still in process. In effect, data sets that have been backed up earlier in the process can be accessed even before the remaining data sets have been backed up. This is not possible with standard disk or tape backup.

Why would a company want to do this? Surprisingly, the need is more common than most realize. During a backup, a user may mistakenly delete or damage an important file. Rather than wait for the entire backup to be completed, the IT group can restore that file, if it has already been backed up, before the backup is complete. This can save the user hours of time otherwise spent waiting, a clear productivity win for the organization.

Second, concurrent read/write speeds cloning. Organizations that use cloning to copy backed up data from disk to tape can begin the cloning process on any completed save set, even while the backup-to-disk is still running. Traditionally cloning could begin only after all the save sets in a group had been backed up. As a result, a backup and cloning process can finish in nearly the same time it takes to complete only the backup portion.

### ***Network Data Management Protocol***

Network-attached storage has become well accepted for enterprise storage. NAS enables organizations to consolidate file system data, reduce file storage administrative overhead, and lower management costs. Because NAS devices typically have specialized operating systems optimized for a single function, protecting these systems via traditional backup methods, which were designed primarily for block-level storage, is problematic. Most NAS vendors, in fact, discourage the running of additional software, such as backup agents that may degrade performance, on these devices.

In response, the industry developed the Network Data Management Protocol to provide centralized control of backup and restore for NAS devices. By defining a protocol that can be easily implemented by NAS vendors and other system and tape device vendors, NDMP allows backup, restore, and other data transactions to be performed without the installation of specialized software. More than 25 companies have adopted NDMP, and the protocol has achieved wide acceptance among backup vendors, NAS vendors, and tape library manufacturers.

NDMP is an open standard for centralized backup of storage devices in the enterprise, allowing a backup application to control native backup-and-restore facilities on NAS. With NDMP, vendors follow common interfaces for common architecture data transfers. Similarly, NDMP uses common commands to pass file metadata (index data) to and from the backup software, regardless of the software package being utilized. Because NDMP integrates with a standard backup program on the NDMP server, the data format on tape is consistent no matter which backup vendor is acting as the NDMP client.

In the simplest NDMP protection configuration, the tape device or library is directly attached to the NAS system to be backed up. Although this reduces traffic on the network, directly attaching the NDMP backup to the NAS system presents drawbacks in organizations where IT wants to simplify, streamline, and consolidate backup. In that case, NDMP can be configured to back up the NAS system through a central backup server.

EMC has adapted NetWorker to take full advantage of its NDMP Client Connection to deliver backup and restore of mission-critical data residing on NDMP-enabled NAS file servers. Prior to the release of NetWorker 7.2, leveraging B2D with NDMP for NAS simply was not possible. NetWorker's disk-based backup option, however, solves the problems that have been plaguing conventional tape-based backup systems for decades. NetWorker's native functionality enables enterprises to automate, centralize, and consolidate the protection of all elements in a heterogeneous environment including NDMP-compliant NAS.

## ***Direct File Access (DFA)***

Direct file access (DFA), a unique part of the NetWorker 7.2 B2D feature set, uses the EMC Advanced File Type Device (AFTD) configuration to enable direct disk access to recover save set data. In practice, the NetWorker recover command uses the direct path to the save set file, automatically bypassing NetWorker's standard media-management process, to read information directly from the AFTD.

The benefits of DFA are increased availability and connectivity and a reduction in the latency (wait time) involved in using the recover program to restore data from NetWorker storage media. In short, it delivers faster and easier recoveries. Since file systems support concurrent access, DFA also enables true concurrent access to AFTDs. This means that save sets can be recovered simultaneously through multi-read/multi-restore access to save sets. If for any reason a NetWorker client cannot directly read and access a save set file that is stored on disk, then the recover program automatically uses the standard access process through NetWorker rather than DFA to recover the data.

## ***Single-step recovery***

Single-step recovery, as its name suggests, gives organizations the ability to complete a recovery from disk or from data that has been cloned or staged from disk to tape media in a single step. To provide single-step recovery the B2D product should be intelligent enough to understand where the backup data resides and enable recovery directly from tape, if appropriate, without having to first bring it back to the original medium by undoing the clone/stage operation. Some solutions in the market cannot do this.

## Achieving the next level of operational recovery management

To move to the next level of operational recovery management, IT must simplify and centralize backup and recovery operations across the wide variety of platforms and storage topologies, both SAN and NAS, within the enterprise. Whether the organization has one large data center or a far-reaching network of branch offices, management needs assurance that all the data is safe, and most importantly, readily accessible.

EMC Legato NetWorker 7.2 makes it possible for IT to confidently provide that assurance. With NetWorker 7.2 IT can centralize operations and provide automation for efficient operational recovery management. With its comprehensive B2D features, such as concurrent read/write, NDMP support, DFA, and single-step recovery capabilities, NetWorker 7.2 makes it possible for IT to meet the demand for higher service levels while making the most efficient use of its backup and recovery budget. By leveraging disk storage, IT can deliver fast and reliable backup and recovery operations. In capitalizing on the B2D capabilities of NetWorker 7.2, IT gains an effective alternative to the tape-only backup solution for managing data—an alternative that allows IT to raise operational recovery management to a higher level.

**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](http://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](http://www.emc.com/legato) or call **888.853.4286** (outside the U.S.: +1.650.210.7000).