



WHITE PAPER

Disk Backup and EMC: Addressing Today's Business Problems

By Heidi Biggar

May, 2008

Table of Contents

Table of Contents..... i
The Changing Face of Data Protection..... 1
 Disk Backup Continuum..... 1
 Integrated Data Protection Ecosystem 3
EMC's Disk Backup Portfolio 4
 Flexible Disk Platform Support..... 5
ESG's View 7

All trademark names are property of their respective companies. Information contained in this publication has been obtained by sources The Enterprise Strategy Group (ESG) considers to be reliable but is not warranted by ESG. This publication may contain opinions of ESG, which are subject to change from time to time. This publication is copyrighted by The Enterprise Strategy Group, Inc. Any reproduction or redistribution of this publication, in whole or in part, whether in hard-copy format, electronically, or otherwise to persons not authorized to receive it, without the express consent of the Enterprise Strategy Group, Inc., is in violation of U.S. copyright law and will be subject to an action for civil damages and, if applicable, criminal prosecution. Should you have any questions, please contact ESG Client Relations at (508) 482-0188. This ESG White Paper was developed with the assistance and funding of EMC.

The Changing Face of Data Protection

There's no question about it: The face of data protection has forever changed—and that new face is disk. Today, more and more organizations prefer disk as the nightly backup target. For an increasing number of these organizations, it is the only solution. Corporate and regulatory requirements, eDiscovery/litigation events and voluminous data pools all demand quick, easy recovery—and you get this only from backing up to disk.

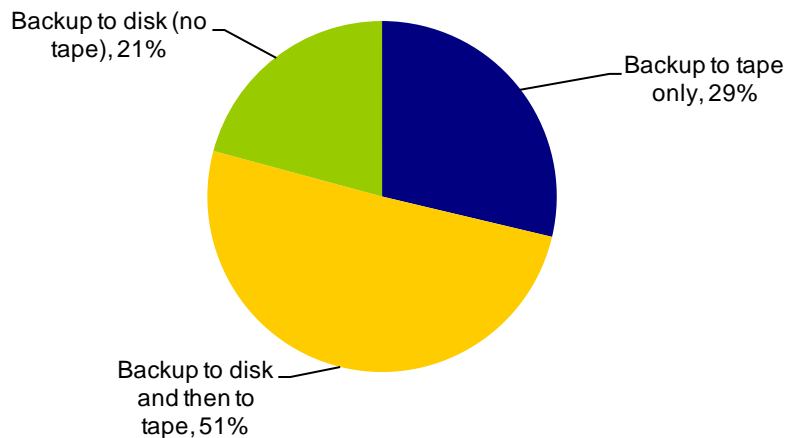
In fact, in a recent ESG Research survey, 21% of respondents said they backed up to disk exclusively (i.e., they used no tape at all in their environments) and 51% said they used a combination of disk and tape for backup (see Figure 1). This means that 72% of the users who participated in our research are using disk-based backup—strongly validating its value. This statistic should come as little surprise to most IT professionals, given the business and economic climate we live in and the demanding SLAs IT departments are continuously dealt.

While improving backup performance is still a key consideration in an organization's decision to implement disk backup today, users tell us that improving recovery performance and reliability is just as—if not more—important. For a large percentage of organizations, being able to recover data quickly isn't just a "nice-to-do;" it's a "have-to-do" with dire consequences. In regulatory situations, for example, delayed recovery can result in fines ranging in the hundreds of thousands of dollars, or more. In OLTP situations, it can mean lost opportunity and revenue. In either case, it can mean permanent or long-lasting damage to a company or brand reputation.

There are other compelling reasons to back up to disk aside from improved backup and recovery performance. Disk, by nature, is a lot more user-friendly than tape from a media management perspective (there are no tape cartridges to keep track of or keep from the falling into the wrong hands), it tends to be a lot more flexible for remote data protection and it comes in various configurations (e.g., from just "dumb" disk to VTL), which helps users ensure that their data protection practices match specific data requirements.

FIGURE 1. ORGANIZATION'S CURRENT DATA BACKUP PROCESS

Which of the following best describes your organization's current data backup process? (Percent of respondents, N = 228)



Source: Enterprise Strategy Group, 2007

Disk Backup Continuum

ESG recognizes that not all data is created equal and therefore, should not be treated equally. Some data types simply have more intrinsic value to an organization—from a business, strategy or IP standpoint—than others and therefore, require a higher level of protection.

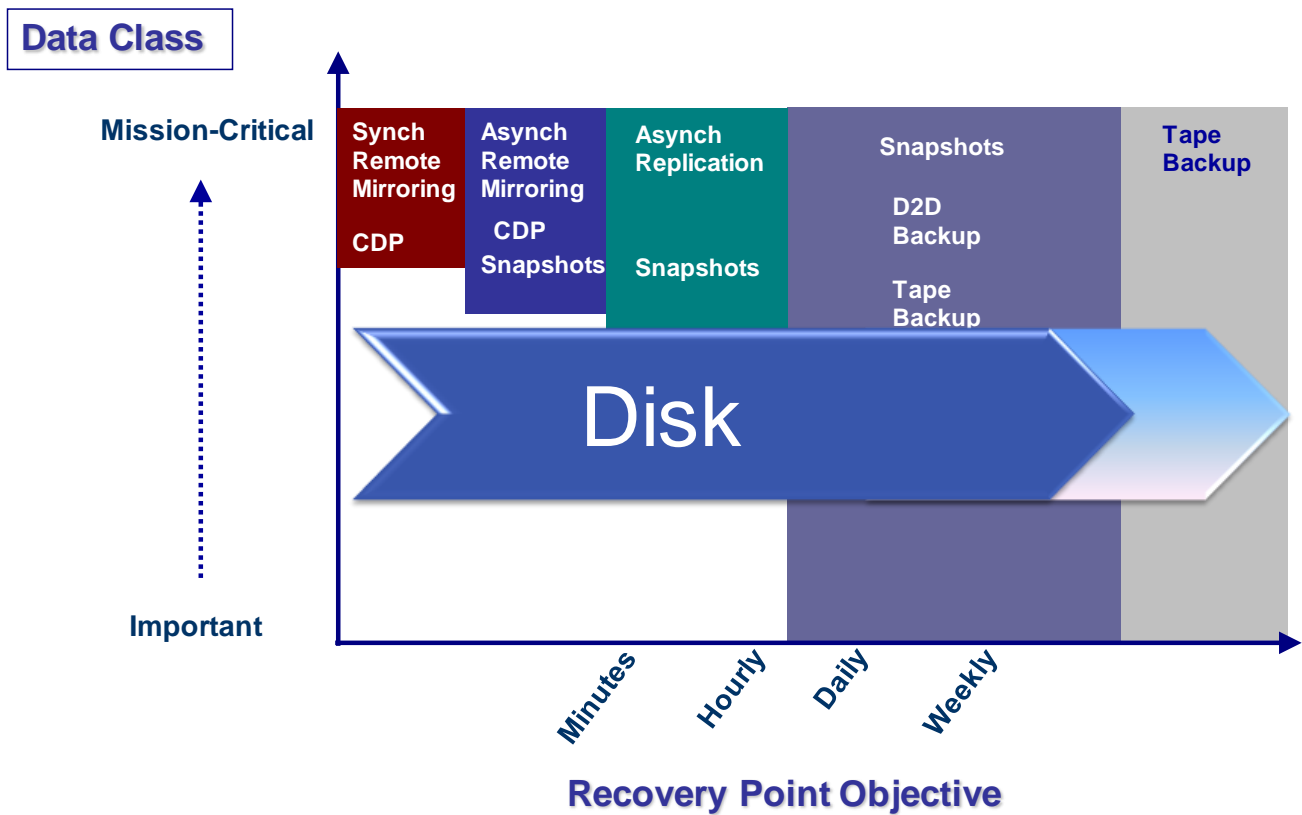
**Disk Backup and EMC:
Addressing Today's Business Problems**

For years, organizations really had only two choices when it came to protecting their data. They could back up to tape or, if a higher level of protection was needed, they could leverage some type of array-based replication or snapshot technology. By default, organizations chose tape for the bulk of their data, reserving replication and snapshot for their most-mission-critical data. Thanks to the advent of SATA disk, new twists in replication/snapshot (improved implementation and recovery granularity) and data de-duplication, users have a lot more flexibility when it comes to how they protect their data.

Ultimately, the products or technologies an organization implements depends on the data protection SLAs it establishes. Invariably, this boils down to a discussion about recovery point objectives (RPOs) and recovery time objectives (RTOs). RPOs refer to the amount of data loss an organization can tolerate; RTOs, to the length of downtime an organization can tolerate. Both are measured in a unit of time (e.g., minutes, hours, days, weeks, etc.). The more mission-critical the data, the lower the RTO or RPO typically.

In many cases, disk-based technologies can—and should—be used in concert to meet protection requirements. Figure 2 shows today's current data protection continuum—and the various technologies organizations can leverage, depending on the value of the data and established RPOs. CDP or synchronous remote mirroring may be most suitable for your organization's most mission-critical data, disk backup for data that is a little less mission-critical and can tolerate a little more data loss and tape backup for the least mission-critical information.

FIGURE 2. TODAY'S DISK-CENTRIC DATA PROTECTION CONTINUUM



Implementing a disk-centric data protection ecosystem, or continuum, has several key benefits:

- **It enables a more granular recovery process.** This allows users to match recovery objectives (e.g., both RPO and RTO) to the importance of the data it generates and, importantly, adjust the level of data protection over time as the data's value changes. All changes in the frequency of access, the age of the data, etc., should be considered to ensure data is protected appropriately.
- **It makes for a more efficient data protection environment.** Having a continuum of data protection solutions ensures efficient data protection from recovery, management and operational standpoints. It allows users to meet backup and recovery windows, reduce or eliminate media management headaches and overhead (common with tape environments) and improve backup and recovery reliability.
- **It reduces business and data risk.** A continuum of data protection products will better equip organizations to respond to DR, regulatory and eDiscovery situations. Organizations are able to get to data when they need it—and in appropriate time frames. Doing so minimizes business risk.
- **It reduces CAPEX and OPEX costs.** Too much of good thing can bad. Take CDP, for example. While an organization's most-critical data may require continuous data protection, it doesn't make sense to apply CDP to all data types. Doing so would be costly and inefficient, consuming unnecessary disk space and keeping organizations from using disk as their primary defense for all data types. The data protection ecosystem allows users to apply the right level of data protection to the right data at the right time. It is a fluid architecture that allows users to maximize protection and minimize both capital and operational (CAPEX and OPEX) costs.

End-User Perspectives: Backup and Recovery Challenges

Backup Windows: "Backup times are always a challenge. We're running high-end research systems that need to be available from 7:00 AM until 4:00 AM the following morning, which leaves us a very short backup window."

Media Management: "One problem we have is that we can't recycle backup tapes fast enough. This has really affected our ability to back up and restore desktops. Our backup applications backup all daily changes across thousands of desktops, so tapes get filled up pretty quickly. It sometimes takes us three days to restore a desktop because the user's data is scattered across 40 different backup tapes!"

Consumption of Human Resources: "With tape-based systems, the human overhead of managing tapes is much more pronounced than with disk-based products, unless you have a really high-end tape library."

Recovery time: "Growth in ERP data will really strain our backup windows in the near future. As it is, we probably don't have enough protection against data loss. If it takes us six hours to back up 500GB, I figure we'll need twice that to recover. Twelve hours is just not an acceptable RTO for a company of our size."

Integrated Data Protection Ecosystem

Clearly, implementing a data protection ecosystem is a good thing, but implementing an ecosystem of integrated applications can be even better.

Integrated platforms give users the flexibility to cover more—or even all—data protection bases from a single source. This contrasts with "point product" approaches in which multiple technologies (typically from multiple vendors) are used in combination to meet data protection requirements. The problem with a point product approach is that there is generally very little integration between applications. This can make management difficult (because multiple GUIs are needed to manage multiple data protection applications), increase CAPEX and OPEX costs (again because of the multiple applications required) and, importantly, make the data protection process a lot less fluid. Again, the idea behind the data protection ecosystem is being able to apply the right data protection technologies to the right data at the right time and at the right price point throughout the data life-cycle. This requires a level of communication between participating data protection applications. The idea behind

integrated platforms is to enable users to initiate and manage multiple data protection processes (e.g., volume snapshots, continuous data capture, replication, and even data de-duplication) from a single console.

Of course, the level of benefit an organization will see from an integrated vs. a point product approach will vary, depending on the types and granularity of the data protection applications offered (e.g., replication, CDP, snapshot, VTL, data de-duplication, etc.) and how tightly the products are integrated. ESG contends that vendors that integrate these applications on the front-end and provide easily accessible and searchable secondary storage pool options on the back-end will have a clear advantage in the market (see Sidebar: Keeping It Simple).

ESG Research¹ finds that this type of platform resonates with users. In fact, 62% of respondents of a recent ESG survey said they were interested in a more unified approach to data archiving. While this survey was specific to the digital archiving space, it is indicative of a general trend in the industry toward a more streamlined approach to data management due to the various technical, operational and financial challenges that result as IT environments become more diverse and complex. Such as:

- Evaluating, purchasing, implementing and managing an increasing number of discrete point products.
- The difficulties involved in efficient sharing or re-use of corporate information assets due to competing standards or the lack of a common data repository.
- Training end-users on multiple solutions is consuming precious IT and business resources.
- The high price tags associated with the purchase and operation of different solutions.

EMC's Disk Backup Portfolio

EMC offers a comprehensive portfolio of products that span the data protection continuum (see Figure 3), and with the introduction of data de-duplication for its Disk Library family, EMC further extends de-duplication support across its backup portfolio. This breadth of disk and data de-duplication capabilities allows users to pick-and-choose solutions to match the data protection requirements of different data types. For mission-critical data, that may mean taking frequent snapshots of data via EMC NetWorker or doing continuous capture via EMC RecoverPoint. For data that is a little less mission-critical, backing up to an EMC Disk Library (via EMC NetWorker) with or without data de-duplication may be appropriate. For remote offices and data in VMware environments, EMC Avamar is appropriate. With Avamar, backup data is de-duplicated (globally) at the source (i.e., on the client) before it is moved across the network and written to disk.

From a management perspective, EMC has already begun to integrate these technologies. Integration makes the data protection continuum more fluid. The more fluid the architecture, the better able organizations will be to meet changing data protection requirements and the more efficient the data protection environment from a capital and operational standpoint. An integrated environment means fewer points of management, fewer agents to install/manage and less hardware to buy/manage.

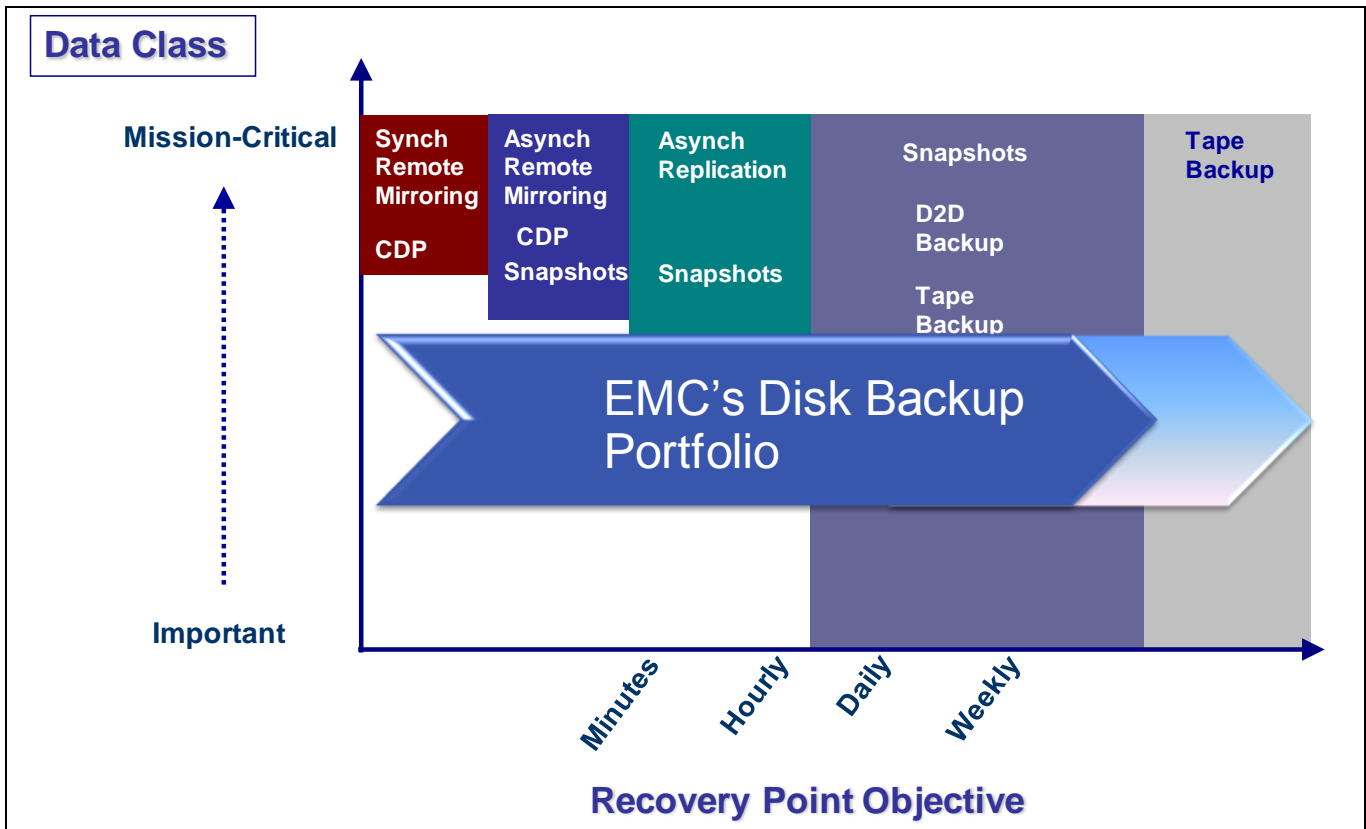
Due to its existing footprint in the data center, NetWorker is the primary integration point of EMC's disk backup portfolio. From NetWorker, organizations can initiate and manage other data protection applications, including file-level backup to disk, volume snapshots (Symmetrix, Clariion, and Celerra), block-level continuous capture (RecoverPoint) and sub-file source de-duplicated (Avamar) backup to disk. Which application organizations use depends on the value, or criticality, of the data and data protection SLAs.

EMC has also integrated the management of Avamar and RecoverPoint with NetWorker. Today, this means that NetWorker customers benefit from Avamar's source and global de-duplication capabilities without having to install, or run, a second client; the de-duplication technology has been integrated directly into NetWorker.

¹ ESG Research Report: *Digital Archiving: End-User Survey & Market Forecast 2006-2010*, March, 2006.

RecoverPoint snapshots are fully indexed within the NetWorker catalogue. EMC customers can manage both de-duplication and CDP—configuration and monitoring—through their NetWorker consoles.

FIGURE 3. EMC'S DISK-CENTRIC DATA PROTECTION ECOSYSTEM



Flexible Disk Platform Support

In addition to providing organizations with a range of data protection applications (e.g., snapshot, continuous capture, de-duplicated backup to disk), EMC also offer a variety of SATA-based disk systems to back up to, including straight disk systems and its EMC Disk Library, which has a virtual tape interface. Again, these options give users flexibility to meet the different data protection requirements of different data types—and, equally important, do it within IT budgets. The target for NetWorker-based backup could be low-cost SATA disk or EMC Disk Library, for example.

The EMC Disk Library works with multiple backup applications, including EMC NetWorker, and simulates a variety of tape products. With the Disk Library, backups can be done in a traditional fashion where the backup software runs on a separate media server or, alternatively, users can choose an implementation in which the backup application (EMC NetWorker or Symantec NetBackup) runs from the Disk Library. This type of implementation simplifies the backup process by allowing users to initiate and control key media management functions (e.g., tape copy, tape cloning, etc.) from the backup application. These processes, as well as the backup process itself, are monitored via EMC Backup Advisor.

In adding data de-duplication support to the Disk Library, EMC is now able to offer both source- and target-based de-duplication to its customers. This flexibility allows organizations to choose the de-duplication solution that is best-suited to their environments and the data they are protecting—from a single vendor. For example, organizations that are hampered by network or other resource bottlenecks may opt to implement EMC Avamar for de-duplication because it de-duplicates data at the source before data is sent over the network. Others may choose to do the de-duplication at the target in the Disk Library itself.

Target de-duplication is available with EMC's new Disk Library 3D 1500 and 3D 3000 LAN backup to disk series as well as an option for its DL4000 series VTLs. In either case, data de-duplication reduces the amount of physical disk that is needed for backup, which has significant implications from both an RTO and an RPO perspective. By offering de-duplication at different places in the storage environment, EMC widens the range of RTO and RPO scenarios in which disk can play in the data protection continuum.

Keeping It Simple

There are hundreds, if not thousands, of discrete applications on the market today—all creating and managing data separately. While many of these products support or integrate with other applications, there is rarely any commonality or correlation of the data they generate. This means there is no easy way for organizations to share data among these applications, no easy way to search for specific content across them and no easy way to access the various types of data they produce. Without this commonality, there is no way to turn the enormous amounts of data organizations generate into information—information that can then be leveraged for both IT and business benefit.

While many solutions provide single-pane-of-glass views of other applications (e.g., backup, archive, replication, snapshots, etc.), allow users to kick off applications from within other complementary applications or even set policies for multiple applications, few products do more than that today—but those that do will have a clear competitive advantage.

Consider traditional backup applications. Most were designed to make copies of primary data for DR purposes. Replication, archiving and snapshot capabilities, etc., have been added over time, as dictated by business SLAs and other IT requirements. The problem is that in many cases, these applications still function discretely. They become individual containers with little or no commonality among them. Eliminating these silos has potentially huge benefits to organizations. Creating an integrated platform, or data protection ecosystem, is the first step toward that end.

ESG's View

In today's IT environment, data and applications abound. An integrated data protection ecosystem has the potential to provide organizations with a level of efficiency and reliability they never dreamed of. Instead of managing multiple applications from multiple vendors, organizations manage multiple applications from a single GUI. Doing so has the immediate benefit of simplification and cost control, but has the long term potential to change IT and business dynamics all together.

Today, revenue loss is typically calculated in time lost (in trying to recover data quickly and completely in recovery situations). What about lost opportunities? While organizations stand to lose potentially hundreds of thousands, if not millions, of dollars by not protecting data resources adequately, they leave significant dollars on the table by not fully leveraging the data assets (or information) they already have. An integrated data protection platform can help organizations better leverage this information.

By providing a continuum of data protection solutions—and integrating them—EMC not only helps ensure that organizations have the right data protection tools for the right data at the right time—which has a range of benefits as described in this report—but puts organizations in a better position to maximize the value of the data assets it protects.



20 Asylum Street
Milford, MA 01757
Tel: 508-482-0188
Fax: 508-482-0218

www.enterprisestrategygroup.com