WHITE PAPER: DATA PROTECTION

Your Backup Is Not an Archive

Optimizing your backup and recovery environment

INNEO That's IT. Händlerinformation

INNEO Solutions GmbH · inneo@inneo.com · www.inneo.com

Deutschland:
Rindelbacher Straße 42
73479 Ellwangen
Telefon: +49 (0) 7961 890-0
Fax: +49 (0) 7961 890-177

Schweiz:
Ruchstuckstrasse 21
CH-8306 Brüttisellen
Telefon: +41 (0) 44 805 1010
Fax: +41 (0) 44 805 1011

White Paper: Data Protection

Your Backup Is Not an Archive

Optimizing your backup and recovery environment

Contents

Introduction

Data growth, budget pressure, rising service-level requirements, and new technologies like virtualization strain the capabilities of traditional backup and recovery processes. But recent advances in data protection and management unlock an optimized strategy that:

- Dramatically reduces backup and recovery times
- Works smoothly across the virtual/physical boundary
- Introduces policy-based management to data retention
- Minimizes disruption from legal and regulatory Discovery

This paper outlines the elements of a modern data protection and management strategy using the Symantec™ NetBackup™ enterprise backup and recovery solution and Symantec Enterprise Vault™ archiving solution.

Backups and Archives

Definitions first: backup and archiving are both data management disciplines that use similar resources and present similar technical challenges. But they differ in fundamental ways, starting with the reasons for doing them at all.

Backups are for recovery

You know all about backups, but let's review the basics for comparison with archives. Backups protect the active, useful state of an organization's information and processes. They allow recovery of a document, server, or even the entire organization—in full operational context—to a point in time before the information and processes were interrupted, corrupted, or lost.

Archives are for discovery

Archives have an entirely different purpose. They preserve *inactive* information as required by law and company policy—a dull task made necessary because even inactive documents may not be completely useless or irrelevant. For example, a week-old "out of office" email may someday help document an employee's attendance record, an old product brochure help resolve a warranty question, or an expense report help prove a tax deduction. Expiry dates—when documents are judged to have truly outlived their usefulness—schedule document destruction according to a combination of laws, regulations, company policies, and common-sense rules of thumb.

Both backups and archives are precautions taken "just in case"—but the cases are different, and different admins may be responsible for each. Backups help recover information and processes in current use in case they are interrupted, corrupted, or lost. Archives help discover details of information and processes *not* in current use, in case they become useful again because of some unanticipated legal or regulatory event.

Table 1 gives a side-by-side comparison of backup and archiving processes:

Backup and archiving at a glance

Issue	Backup	Archiving
What is it?	Protection for system and data "live state"	Records of inactive document "steady state"
Why use it?	Recovery" restore business operations after data loss, interrruption, or disaster	Discovery: produce evidence to meet legal, regulatory, and policy obligations
Who wants it?	Business stakeholders—CEO, shareholders, your boss	Public stakeholders—courts, regulators
What's in it?	Images in full operational context	Individual objects, especially email
How many are there?	Many: original left in place plus multiple point-in-time copies	One: single global instance – originals replaced by links, or removed altogether from primary storage

Table 1: Backup and archiving look alike on the surface, but differ in purpose as well as operation.

The hidden value of archiving

Archiving lacks the dramatic “bet the company” scenarios that drive backup and recovery, so it’s often overlooked in data-management strategies of even sophisticated companies. That’s unfortunate, because archiving unlocks enormous efficiencies in storage capacity, core network bandwidth, backup/recovery time, user service levels, and more—especially when it’s used together with a modern backup solution. How? The power of archiving is that it recognizes and acts on the *value* of the information it handles, for example:

- Determining whether a document is in use or idle
- Moving idle documents from fast-but-expensive storage to economical archive storage
- Classifying and indexing archived documents for rapid discovery and retrieval
- Recognizing applications to preserve the experience when users access archived data through Outlook, Lotus Notes, SharePoint, etc. Information management challenges

Your backup is not an archive

It's a reasonable question: "Why not use my backup files as an archive - all the information is in there, isn't it?" It is. But try to use backups as an archive to support legal or regulatory discovery, and the differences will become quickly, painfully obvious. Discovery requrements grow every year, and backups simply cannot meet them. Here are a few reasons why:

- *There are too many copies.* Multiple point-in-time backups help you meet recovery point objectives. But in Discovery, multiple copies are discrepancies that need to be reconciled or explained. An archive provides a single “official” indexed record.
- *They’re not organized.* Recovery puts everything back where it came from, so indexing by content isn’t important. In Discovery, you're looking for needles in a haystack: items that meet search criteria, no matter where they are, so indexing is essential.
- *They keep too much of the wrong stuff.* In a backup-based “archive”, expiry usually means date-based destruction of backup tapes. But laws and regulations set different retention dates for different types of information. Archiving supports policy-based retention to meet those requirements.
- *E-Discovery can tie up your staff.* Discovery brings in teams of paralegals, inside and outside counsel, and service providers. If the information they need is on backup tapes, their requests will be routed through your staff. If it’s in a modern archiving solution, they can use their own tools to get it.
- *“Legal hold” can wipe out your storage.* Legal hold stops modification or destruction of documentary evidence. A point-in-time backup of everything is a common approach that wastes a lot of storage. Archives simply flag the documents “do not destroy” for the duration of the hold.

It comes down to using the right tool for the job. In the United States, for example, Federal Rules for Civil Procedure govern the use of electronic evidence, and set requirements for holds. Archiving solutions are designed to meet them; backup solutions never were.

Information management challenges

Now, let’s review the data-management challenges organizations face today. They aren’t hard to find: long-term changes in the volume, use, and value of information are changing the way organizations manage their information, with direct effects on their backup and archiving practices. Here’s a brief overview of some of the trends and their implications:

- *Data growth* continues, leading to longer backups and more storage. And even as storage prices drop, storage management grows more complex.
- *Budget pressures* postpone improvements, spread talent thin, and add risk of coverage gaps and errors.
- *Rising requirements*—internal Service Level Agreements , external regulations, and user requirements rise constantly, and keep the pressure on IT.
- *New technologies*, especially server virtualization, compete with backup processes for access to high-performance storage and network resources.

Backup and archiving processes must address these issues using the most cost-effective technology, without introducing unacceptable risks or compromising business effectiveness.

Special challenges of unstructured data

The technical challenges of archiving are driven by all of the above, plus the volume and variety of unstructured information that organizations must now retain to meet their corporate governance standards, industry regulations, and legal obligations:

- Email – the primary focus of most legal discovery – volume continues to grow, with voicemail, pictures, video and more "unified messaging" on the way
- Most employees at medium-to-large businesses already use Instant Messaging (IM) technologies, with more on the way
- Messaging and collaboration tools such as Microsoft® SharePoint®, the constant “churn” of document formats, and end users' attempts to retain control of files and messages all complicate the environment

The results are large data repositories spread out across the organization, nearly impossible to consolidate and search. The most common attempts to manage them—email quotas and aggressive expiry policies—trigger risky informal workarounds by employees trying to “protect” their email against corporate policy. The resulting collections of Microsoft Exchange PST and IBM Lotus Notes NSF “archives” take up terabytes of storage on desktop hard drives and shared storage, spilling out to portable media, Web mail services, and home PCs. They are difficult to search, notoriously prone to corruption, a burden on backup time and storage, and wasteful of employee time. Worse, they are *all* likely targets for legal or regulatory discovery. Legal discovery from unmanaged records like these risks "smoking gun" revelations that can prove devastating in court—or in the media. Relaxed file-size constraints under Exchange 2010 relieves some of the strain, but doesn’t fundamentally solve the problem. Informal collections add significant cost and risk, leading many organizations to review their backup processes and add managed, disciplined, policy-driven archiving to their information management arsenal.

Backup and archiving solutions

New solutions address the challenges of backup and recovery—including the complications raised by unstructured data—in three important ways:

- *Deduplication* helps get the most from server, storage, and network infrastructure
- *Virtual Machine* solutions deliver granular recovery and off-host protection to meet the special requirements of virtual machines and the applications and data they contain
- *Archiving* helps cut backup volume, time, and cost, and reduces legal and regulatory exposure

Let’s consider them one at a time:

Deduplication

Backing up duplicate data wastes time, storage capacity, and network bandwidth. Block-level deduplication, on the other hand, lets physical and virtual machine backups run 70% to 90% faster, and can cut their storage footprint up to 95%. When deduplication is done at the client instead of the media server, those improvements mean not only a smaller storage footprint, but shorter backup windows and reduced consumption of network bandwidth during backups. Of course, deduplication at the client consumes processor resources, so deduplication at the backup media server is more

appropriate for performance-sensitive applications during peak usage hours. But in general, the best practice is to deduplicate as close to the source as possible, for resource savings in all downstream processes.

Backing up virtual machines

Virtualization is transforming the world's data centers by consolidating multiple virtual machines onto physical hosts, improving hardware and facilities utilization, and power and cooling efficiency. But virtualization involves tradeoffs that affect backup and recovery, for example:

- *More backups*—the ease of deploying virtual machines can create "virtual server sprawl", increasing the time, bandwidth, and storage needed to back them all up. And many IT departments have adopted the practice of backing up *both* the virtual machines and their physical hosts (VMware VMDK or Microsoft VHD environments)—for another doubling of backup resources.
- *Resource competition*—virtualization maximizes utilization of CPU and I/O resources on physical hardware, but backups need those same resources. As a result, a backup on one virtual machine slows down every virtual machine instance on that host.

Resolving the unique challenges of virtual machine backups demands a balanced approach. "Virtual-only" backup point solutions raise complexity with another technology to plan, manage and support. But storage infrastructure or APIs optimized for virtual environments can be used instead for benefits like these:

- *Faster VM backups* using VMware's vStorage API for Data Protection and media server deduplication to cut backup times by two-thirds
- *Low-impact backups* in VMware and Microsoft Hyper-V environments, using block-level incremental and off-host backups
- *Greater storage reduction* through intelligent deduplication across both physical and virtual machines
- *Granular recovery* for VMware and Microsoft Hyper-V, offering single pass backup file-level or complete VM recovery with half the time and storage

Archiving

We have already mentioned archiving as a special case of deduplication for unstructured information. But a well-designed archiving solution improves backup and recovery processes in other ways, for example by:

- *Shrinking data stores* for faster backup and recovery at lower storage cost
- *Supporting self-service recovery* to build user satisfaction and increase service levels while reducing demands on IT staff
- *Supporting search* by corporate Legal and outside counsel to locate relevant documents without IT assistance
- *"In place" legal hold* that protects documents or document classes from tampering without building a duplicate archive

This last two points shouldn't be underemphasized: the "drop everything" urgency of legal discovery can disrupt IT operations and pull technical specialists away from their primary responsibilities into unfamiliar and uncomfortable roles.

But the situation is improving: Organizations such as the Electronic Discovery Reference Model (EDRM) Project have created frameworks for electronic legal discovery that help balance legal requirements with the interests and capabilities of inside and outside counsel, IT, internal staff, specialist legal discovery firms, and solution providers.

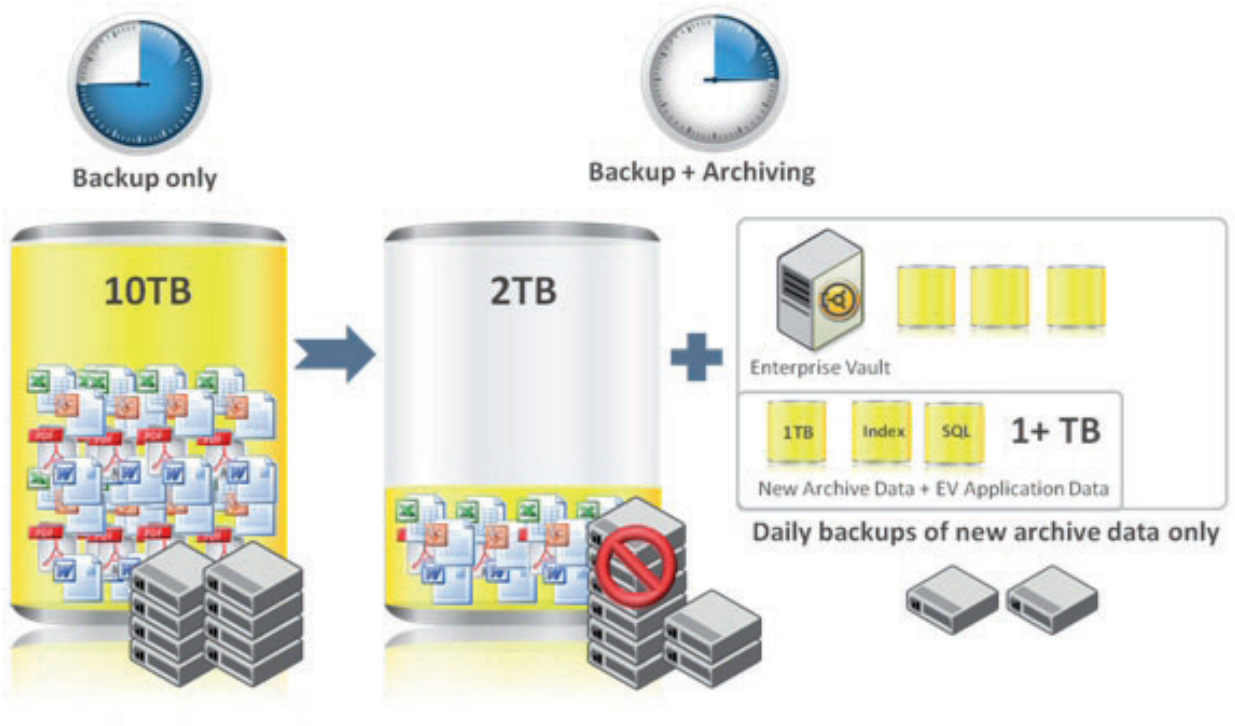

Figure 1: By removing historic data from production storage, Archiving saves space on expensive storage assets, and cuts backup, restore, and disaster recovery times.

The archiving contribution to backup

Client-side deduplication and specialized support for virtual environments are important innovations that solve growing problems. Archiving is an established solution—to a *different* problem—that is finding new relevance as backup grow more strained.

Without archiving, much of the time, bandwidth, and storage spent on backups is simply wasted. Multiple-instance geographic and point-in-time backups of non-current information cut into backup and recovery windows, waste vast amounts of storage, increase costs, disruption, and delay during legal discovery, and introduce unnecessary risks when multiple incompatible records must be reconciled or explained in court.

Archiving is the correct tool for protecting, organizing, and searching this unstructured information: It provides an organized, searchable store of single-instance records, with capabilities to meet the specialized, absolute demands of legal discovery. Deduplication, integrated storage management, and self-service search offload work from backup systems and messaging and collaboration applications, reduce storage costs, and satisfy user expectations for accessibility of emails and files.

Archives prove their worth during the first legal hold or discovery requirement. But even without any such requirement, they pay back quickly by simplifying and reducing the storage and management burden on backup processes.

More than the sum of their parts

Backups and archives perform separate functions—but the capabilities of each help the other work better and more efficiently.

Backups provide essential protection for the archive. Although the idea of backing up the archive seems paradoxical, it actually makes perfect sense. The archive, after all, is part of the "active, useful state" of the organization's information and processes—if the archive were corrupted or lost, the business would face unacceptable risks during litigation or regulatory review.

But archiving solutions help backup processes, too. First, as we've seen, they reduce the useless burden of backing up multiple copies non-current information. That means backups run faster and consume less network bandwidth, processor cycles, and system input/output. Faster, less resource-intensive backups can be run more often—and that means better business protection.

Just as important, the policy-based migration tools built into the best archiving solutions can move older or low-priority archived information from disk storage to tape or other media managed by the backup solution. This kind of storage tiering keeps the archive manageable without sacrificing search or discovery, and frees up high-performance storage that can reduce backup windows and accelerate recovery.

Symantec solutions for archiving and backup/recovery

Symantec Enterprise Vault™

Enterprise Vault is the world's leading solution for email and content archiving. The solution archives unstructured information from messaging, file, and collaboration systems, optimizing storage, classification, and retention processes to help companies to stop buying storage and discover data faster. The solution:

- Moves older, less-frequently accessed data from primary storage to cheaper disk or even tape
- Gives users undiminished access to archived files and email whether they are online or off
- Deduplicates and compresses information to store a single lightweight copy of each file or message, regardless of the number of times it was archived or its original location
- Streamlines search, discovery, and legal hold of archived information without wasting storage or staff time

Symantec NetBackup™

Symantec offers the world's leading enterprise backup and recovery solution, Symantec NetBackup™, which offers:

- Market-leading, scalable data protection for physical and virtual server environments
- A single platform to manage, protect, and recover data across storage tiers, locations, and operating systems
- Deep integration with VMware vStorage API for Data Protection (VADP) and Microsoft Hyper-V
- Built-in client and media server deduplication
- Granular recovery across Exchange, SharePoint, Active Directory, VMware and Hyper-V

Symantec Enterprise Vault™ and Symantec NetBackup™ together

The combination of Symantec Enterprise Vault and Symantec NetBackup delivers an end-to-end solution for managing, protecting and archiving critical information across the enterprise, with capabilities like these:

- Accelerated backup and recovery using Enterprise Vault to reduce the footprint of data stores protected by NetBackup
- Efficient, fully compatible protection of the Enterprise Vault archive using the NetBackup for Enterprise Vault agent
- Policy-based migration tools to move archived data from disk managed by Enterprise Vault to tape or other media managed by NetBackup, maintaining full access to archived files

Conclusions

Backup/recovery and archiving solutions meet different business requirements, using technologies that have kept pace with the scale and complexity of enterprise IT. But an efficient, comprehensive approach to managing and protecting information uses both types of solution to maintain and even raise service levels in the face of growing data volume, budget pressure, legal and regulatory requirements, and technological complexity.

About Symantec

Symantec is a global leader in providing security, storage and systems management solutions to help consumers and organizations secure and manage their information-driven world. Our software and services protect against more risks at more points, more completely and efficiently, enabling confidence wherever information is used or stored.

INNEO That's IT. Händlerinformation

INNEO Solutions GmbH · inneo@inneo.com · www.inneo.com

Deutschland:
Rindelbacher Straße 42
73479 Ellwangen
Telefon: +49 (0) 7961 890-0
Fax: +49 (0) 7961 890-177

Schweiz:
Ruchstuckstrasse 21
CH-8306 Brüttisellen
Telefon: +41 (0) 44 805 1010
Fax: +41 (0) 44 805 1011

For specific country offices and contact numbers, please visit our website.

Symantec World Headquarters
350 Ellis St.
Mountain View, CA 94043 USA
+1 (650) 527 8000
1 (800) 721 3934
www.symantec.com

Symantec helps organizations secure and manage their information-driven world with **storage management**, **email archiving**, **backup & recovery** solutions.

8/2010 21075780