

SharePoint Backup and Recovery: Ensuring Complete Protection

A Detailed Review

Abstract

The growing use of Microsoft SharePoint in the market is driving the need for comprehensive solutions for managing the protection and recovery of the data and servers that make up a SharePoint farm. EMC[®] NetWorker[®] provides complete data protection capabilities for SharePoint, including next-generation technologies like deduplication, to ensure disaster and operational recovery with a range of solution options to help meet even the most stringent backup and recovery service level agreements.

March 2010

Copyright © 2008, 2010 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.

For the most up-to-date listing of EMC product names, see EMC Corporation Trademarks on EMC.com

All other trademarks used herein are the property of their respective owners.

Part Number h4602.1

Table of Contents

Executive summary	4
Introduction	4
Audience	4
The proliferation of SharePoint	4
The complexity of SharePoint protection	5
Disaster recovery	5
Operational recovery.....	6
SharePoint Recycle Bin.....	6
Extracting content to deliver item-level recovery.....	6
EMC NetWorker: Complete backup and recovery for SharePoint.....	6
VSS integration for coordinated disaster recovery	7
Snapshot-based backup.....	8
Granular recovery for logical SharePoint content.....	9
Simplify content restore with comprehensive search.....	10
Integrated deduplication.....	11
EMC: Driving complete SharePoint solutions	11
Conclusion	12

Executive summary

Companies and organizations are on the hunt for new ways to gain effectiveness in managing corporate information as well as improve productivity and effectiveness in accomplishing business-oriented tasks. Increasingly, customers turn to Microsoft, the established leader in business productivity applications, to improve information management and collaboration with SharePoint. This solution, now an integral part of the Microsoft Office Suite alongside mainstay applications like Outlook, Word, Excel, and PowerPoint, helps users become more effective in day-to-day operations from content management to team projects.

At the same time, as SharePoint provides an improved means of managing information and group collaboration, it also adds complexity to the task of backup and recovery. Data once housed in simple file shares – or on paper – now lives in a federated infrastructure that requires more sophisticated data protection techniques. Backup solution providers are responding to customer needs with tools and best practices that ensure consistency and mask the complexity of restoring SharePoint data. In its mission to help organizations of all sizes manage growing stores of information more effectively EMC extends its solution set and expertise specifically to SharePoint to help innovate ways to store, protect, optimize, and leverage information housed in this increasingly mission-critical application.

Introduction

This white paper begins by explaining how SharePoint has proliferated into the enterprise, and then discusses the complexity of both disaster and operational recovery for SharePoint. The paper then describes how EMC® NetWorker® provides complete backup and recovery for SharePoint.

Audience

The intended audience for this paper is backup, storage, application, and server administrators who require a greater understanding of the challenges of backup for Microsoft Office SharePoint Server and how EMC NetWorker delivers comprehensive SharePoint protection. The paper is also useful for EMC Partners that deliver solutions for SharePoint backup and recovery.

The proliferation of SharePoint

In 2009, Microsoft commented that its SharePoint business is “on fire.”¹ A large number of customers are investing in SharePoint software and infrastructure to help better leverage information both from an internal and external standpoint. SharePoint provides several capabilities that help serve up information to further the interests of the organization:

- **Collaboration** – Gives workers the ability to work together more efficiently from sharing information to project planning to scheduling meetings
- **Content management** – Helps users manage documents and files of all types
- **Business intelligence** – Tools for filtering, analyzing, and publishing information for improved business effectiveness
- **Business forms** – Browser-based forms simplify data gathering tasks such as polling and document review and approval
- **Portals** – An easy-to-use framework for building, customizing, and publishing internal and external portals simplifies presentation of information and corporate Web-fronts
- **Search** – Content indexing simplifies and speeds location of needed information with easy text-based search

Tools that help with these common business tasks have been available for some time but selection, acquisition, and rollout of such software suites have been significant endeavors. SharePoint, however, is

¹ <http://www.microsoft.com/presspass/features/2009/oct09/10-16BallmerSharepoint.aspx>

easy to access, easy to use, and as a result, adoption is swift – so much so that many companies often find it popping up throughout the organization. While SharePoint helps users accomplish business tasks, at the same time it creates challenges for the IT organization:

- If there is no plan for managing the application and keeping pace with its use, the capacity of storage required, etc., the potential for unplanned downtime is great. Downtime means that users will not be doing the things that are important to the business.
- As SharePoint becomes more central to storing copies and versions of documents, the proliferation of duplicate data adds to costs and management complexity.
- Data deleted purposely or accidentally may represent a governance or compliance risk for the company.

Companies adopting SharePoint must carefully consider and establish backup procedures that will ensure data is protected – and that will meet required service levels to ensure the application and data are available to continue business operations.

The complexity of SharePoint protection

SharePoint installations come in many shapes and sizes from small single-server installations to multi-system farm configurations. Once an organization makes the decision to deploy SharePoint as a mandated corporate tool, most do so with a full SharePoint farm consisting of multiple systems with the intention of serving a significant amount of data to its user base with load-balancing access to ensure performance and availability. Information within SharePoint is presented with logical view sites, folders, lists, etc., that allow users to organize and present information in ways that make the most sense for their project or business. With this shift from housing data on simple file shares to leveraging a content management system comes greater IT complexity. Information and collaboration are improved – but at the same time management grows more complicated and strained, including backup and recovery. Without the right tools and a well-thought-out strategy, data protection and recovery are a challenge.

Disaster recovery

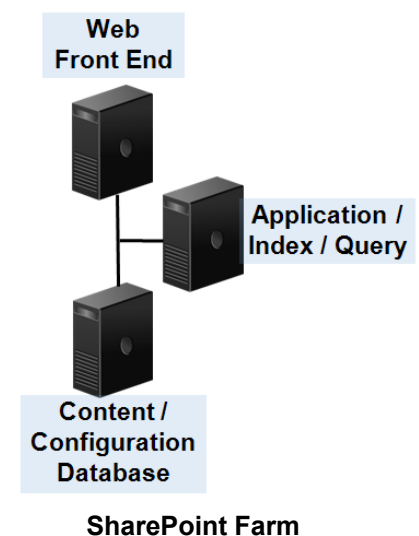
A disaster recovery plan for SharePoint should take into account requirements for restoring individual components that make up a SharePoint farm – or the entire farm itself. For any issue that compromises access to a database, a drive, a system, or an entire physical site/location – what would it take to return to service?

A SharePoint farm is made up of a three-tier topology:

- Front-end server
- Application server
- Database server

These three tiers may be spread across multiple servers. Protecting each of these servers in a SharePoint farm and all of the critical components running on each including the configuration database, search databases, search indexes, content databases, system states, etc. – and protecting each in context with the other – is a critical aspect to complete recovery of a SharePoint farm. Because recovery consists of much more than restoration of simple files, it takes careful planning and the right set of tools to ensure data is captured consistently so when needed, data can be recovered appropriately and in a way that guarantees a restartable SharePoint system.

To aid in the process, Microsoft has provided the Volume Shadow Copy Service (VSS) feature available with Windows Server. VSS is a backup infrastructure provided within the Windows operating systems and applications that creates consistent point-in-time copies of data. Microsoft Office SharePoint Server 2007 supports backup and recovery of its databases and indexes within this infrastructure. The SharePoint VSS Writer is responsible providing a full inventory list of all of the servers, databases, and files within the SharePoint farm that need to be protected. VSS also



puts the servers in the SharePoint farm in a consistent state at the beginning of a backup operation and helps to maintain consistency throughout the process for backup and recovery. VSS simplifies SharePoint backup and restore operations but also enables users to move beyond simple streaming backup operations to leverage snapshot management workflows that help remove the impact of backup from production hosts and improve recovery service levels.

Operational recovery

While DR is a very key capability in the purview of the IT administrator, to the SharePoint user, recovery requirements are distinctly different. Knowledge workers who rely on SharePoint in their day-to-day operations will have little concern of recoverability of databases or indexes or any of the components that make up the SharePoint environment. SharePoint users require restores at a more basic level. They simply want a return of content they have uploaded into SharePoint – a PowerPoint that is suddenly missing or corrupt, a team calendar that was accidentally deleted, etc. – so they can resume their jobs.

Uploading information into SharePoint delivers great benefits – search to help find files, versioning to help track changes, etc., but if we accidentally delete a file within SharePoint, what choices are available for recovery? For recovery at the logical view the best approach is a granular restore of content. A complete solution will deliver recovery for the full range of content in SharePoint. This means more than documents. Recovery levels should also include Web applications, site collections, sites, lists, libraries, calendars, wikis, blogs, etc.

SharePoint Recycle Bin

A facility for user self-service recovery of deleted files is available within Microsoft Office SharePoint Server 2007. The SharePoint Recycle Bin provides a landing place for purposely deleted items and maintains them for a specified period of time to allow an easy restore if needed. After expiration from a user Recycle Bin, deleted content can remain available in a second-tier administrator Recycle Bin as well. For many recovery use cases this is a great tool to help with quick recovery with little impact to IT. Recycle Bin does not, however, meet all use cases for granular recovery that are demanded by most SharePoint users. Limitations include:

- Deleted SharePoint Web Applications, Site Collections and Sites do not land in a Recycle Bin and so cannot be recovered by the tool.
- Content that is corrupted or missing due to system error is not recoverable from the Recycle Bin.
- Since Recycle Bin content counts against user quotas for storage, retention is rarely more than 30-60 days before content is permanently deleted.

Typically another solution is required to ensure item-level operational recovery beyond the capabilities of Recycle Bin.

Extracting content to deliver item-level recovery

As SharePoint has evolved, so have the interfaces into it for managing its content. Microsoft provides an interface into SharePoint that enables management and migration of SharePoint logical content. The SharePoint Content Migration API lets any third-party application export SharePoint content from an entire website to an item in a list or library and stores these objects into XML-formatted files. It enables saving of metadata along with files, and delivers capture of full or incremental content. Through this object model backup applications can deliver a full granular solution for SharePoint recovery.

EMC NetWorker: Complete backup and recovery for SharePoint

Given the complex nature of SharePoint and its trajectory toward a mission-critical application, companies investing in the platform should look for a backup and recovery solution that provides a complete range of protection capabilities. EMC NetWorker delivers comprehensive capabilities for SharePoint and empowers IT organizations to ensure its protection and recovery for disaster recovery and operational recovery.

VSS integration for coordinated disaster recovery

For disaster recovery, NetWorker integrates with Microsoft Volume Shadow Copy Services and with the SharePoint VSS Writer in particular to ensure coordination and enforce order in capturing farm and database-level backups. Other market solutions attempt to capture SharePoint farms through SQL backups loosely coupled with system-level protection. The danger, however, is that there is no interaction and coordination with the SharePoint farm itself to ensure that the right context and a consistent view of the farm are presented before data is delivered for backup.

Before attempting any DR-level protection, NetWorker calls into the farm through the VSS framework to obtain a complete list of the entities that make up the SharePoint farm.

By protecting all of the components in the SharePoint farm as well as the system components required to appropriately restore a server, NetWorker can ensure complete recoverability in the event of any hardware failure, hardware migration or replacement, or natural or manmade disaster. Administrators can perform recovery of an entire SharePoint farm or just a database recovery as required. NetWorker protects the following entities that make up a SharePoint farm:

The **NetWorker Module for Microsoft Applications (NMM)** is a single unified solution that leverages the Microsoft Volume Shadow Copy Service (VSS) for snapshot-based protection and recovery of SharePoint, Exchange, SQL, Hyper-V and Active Directory.

Table 1. NetWorker protection of SharePoint entities

SharePoint entity	Protected data
Web front end	<ul style="list-style-type: none"> – SYSTEM COMPONENTS:\ – All system volumes (for example, C:\, D:\) – APPLICATIONS:\Microsoft Office SharePoint Services\
Search server	<ul style="list-style-type: none"> – SYSTEM COMPONENTS:\ – All system volumes (for example, C:\, D:\) – APPLICATIONS:\Microsoft Office SharePoint Services\
SQL Server	<ul style="list-style-type: none"> – SYSTEM COMPONENTS:\ – All system volumes (for example, C:\, D:\) – All SQL databases (including the configuration and content databases)

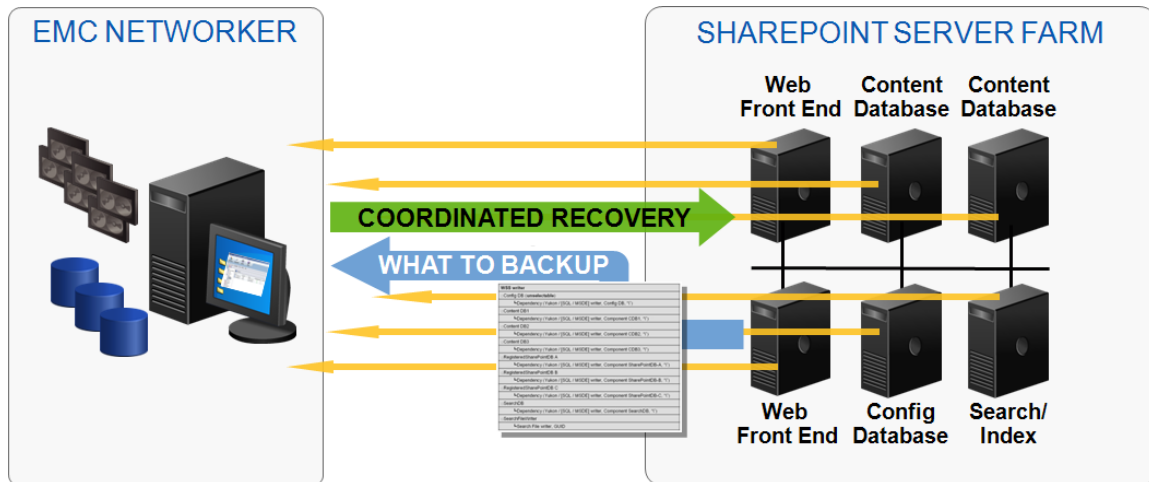


Figure 1. Protection with VSS coordination

Snapshot-based backup

Given the round-the-clock nature of most businesses, any SharePoint recovery capability should provide a fast resumption of business. Having to restore backup data from a tape that is 12 hours old – or more – no longer meets the service levels required by most customers. Through integration with VSS, NetWorker enables instant low-impact snaps for the servers in a SharePoint farm. This enables frequent, consistent recovery points that allow for quick recovery with minimal data loss but enables off-host, off-hours backup to secondary media as required. NetWorker delivers easy scheduling and coordination of SharePoint snapshot creation through the NetWorker Management Console with a set of snapshot policies listed in Table 2.

Table 2. NMC snapshot policies

Policy	Description
Snapshot frequency	The number of farm snapshots to take per day
Snapshot retention	The number of farm snapshots to retain at a time
Snapshot expiration	When to expire snapshots
Snapshot backup	Which snapshot to “roll over” to secondary such as tape, disk or virtual tape

NetWorker takes advantage of both software- and hardware-based VSS snapshots to capture and protect SharePoint data. Customers using intelligent storage arrays that support inbox copies via a VSS hardware provider can take advantage of flexible options for moving data from snapshots to backup media such as disk, tape, and virtual tape. NetWorker supports:

- LAN-based backup from the SharePoint servers in the farm
- LAN-based backup from a proxy client that shares a storage area network (SAN) with the production SharePoint servers
- LAN-free backup from a proxy Storage Node that directs backup data to a SAN-attached target device.

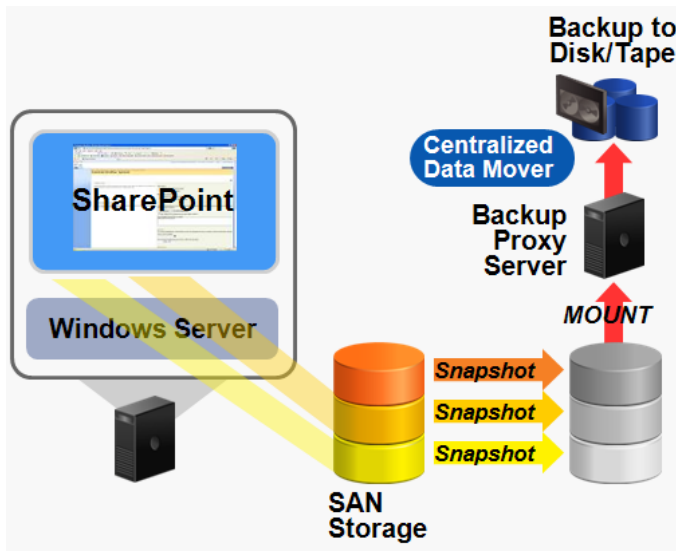


Figure 2. Off-host backup from snapshots eliminates the impact of backup

Proxy-based backup works by leveraging hardware-based “transportable” snapshots that can be mounted to non-production hosts. Through off-host and LAN-free backup, NetWorker delivers several significant benefits for SharePoint users, including the following:

- Eliminates network traffic
- Removes the impact of backup from production farms to ensure users get the best response from the SharePoint farm – even during backup

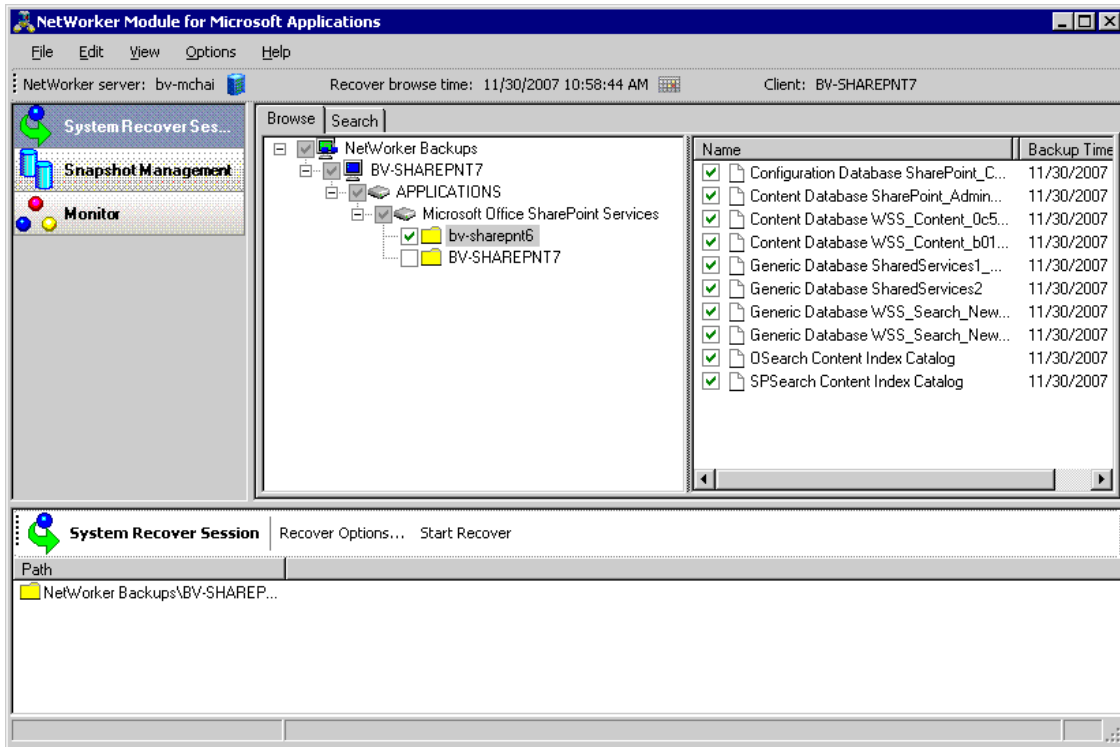


Figure 3. SharePoint farm recovery

Granular recovery for logical SharePoint content

Side-by-side with DR protection, NetWorker also provides complete backup and recovery for logical content within the SharePoint farm. NetWorker integrates with the SharePoint Content Migration API to capture data at the logical level to ensure that on request, a backup administrator can quickly restore items to an original or alternate SharePoint location.

NetWorker provides the flexibility for an organization to target the specific SharePoint sites and locations for which granular recovery is desired or required. You can select all of your SharePoint data for granular protection – or you can selectively narrow its scope to target specific sites and data in a farm. This gives administrators a choice in the type and amount of data it desires for reducing, allowing for managing just the key data for which quick recovery is desired.

Given the growing types of information being managed in SharePoint, the flexibility to protect and recover the gamut of content including versions and metadata is a key requirement. NetWorker protects and recovers a full range of objects housed in SharePoint beyond simple documents. The list of protected and recoverable items includes:

Portal sites	Top-level sites	Portal webs	Subsites
Document libraries and documents	Picture libraries	Galleries	Events
Links	Images	Contacts	Lists and list items
Surveys	Wikis	Discussions	Tasks

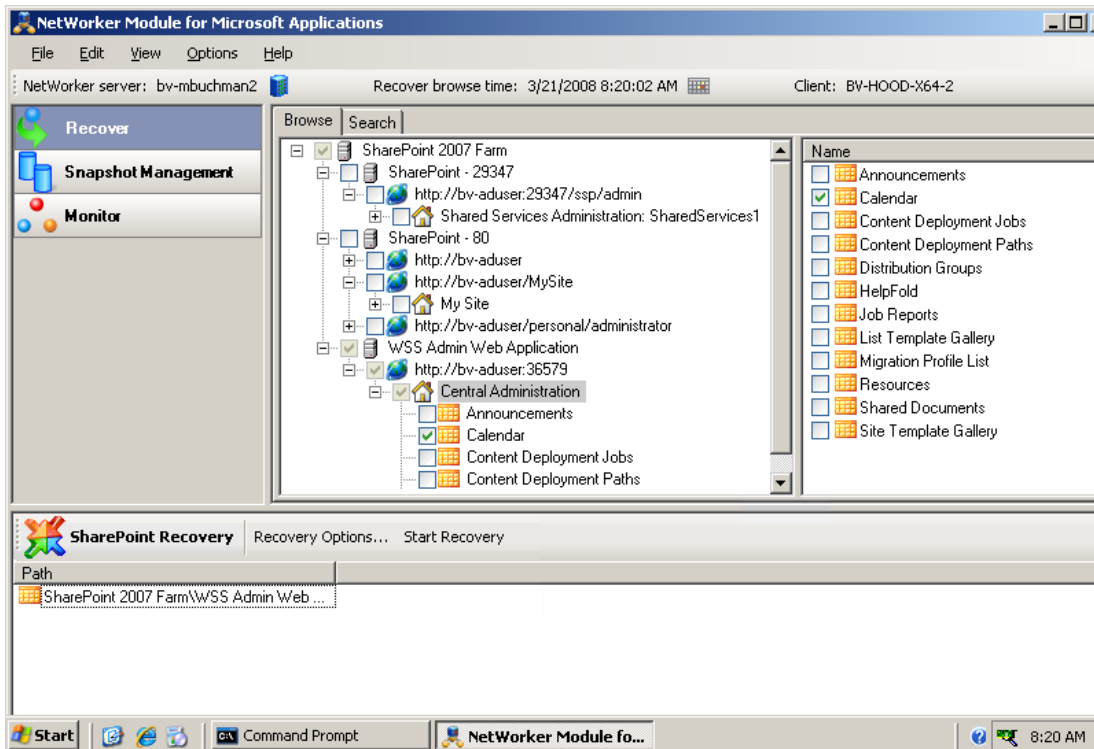


Figure 4. Granular SharePoint recovery

Simplify content restore with comprehensive search

One of the significant value propositions of NetWorker backup for SharePoint is the ability to capture and manage a large amount of SharePoint data and versions captured over time. At the same time, with SharePoint content growing, finding the right data needed for recovery can be a daunting task. NetWorker provides a graphical display of SharePoint sites and contents in an easy-to-use tree view, but for sites that may have a significant number of sites and content under management, NetWorker also simplifies location of data for recovery with a comprehensive search capability. NetWorker's search interface lets you search based on a number of criteria including:

- Name
- Object type
- Creation date
- Backup time
- Modification date
- Author
- Size
- Title

Any items that match the provided criteria are displayed. Restores are quickly executed from the returned results.

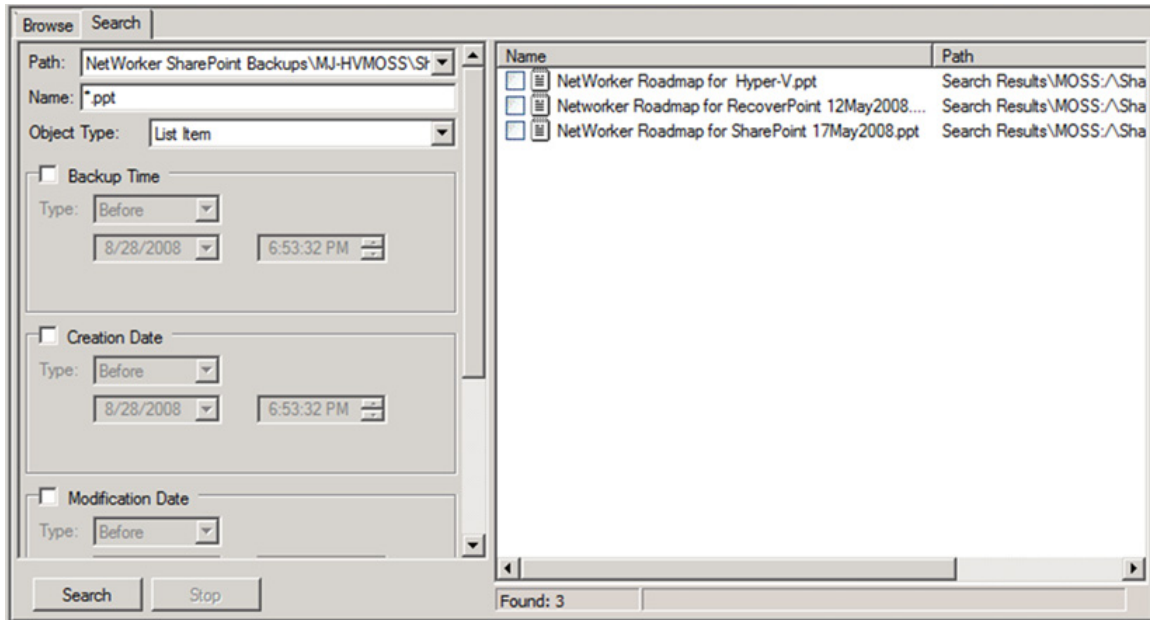


Figure 5. Easy search for SharePoint recovery items

Integrated deduplication

NetWorker provides deduplication support to address backup challenges introduced by the proliferation of redundant data. This is particularly advantageous in SharePoint environments where multiple document copies and versions are likely to be housed. Identifying and eliminating duplicate data in the backup and recovery environment delivers significant advantages:

- Retain backups longer, using 20 to 50 times less disk
- Replicate data offsite faster with 99 percent bandwidth efficiency
- Recover reliably from disk with advanced data integrity
- Optimize server consolidation with faster virtual server backups
- Reduce power, cooling, and space requirements

Through integration with EMC Avamar[®], NetWorker delivers deduplication intelligence at the backup client level, identifies unique data segments across sites and servers, and transfers only new data during the backup process, resulting in fast and efficient backup. One of the easiest ways to take advantage of deduplication in a NetWorker environment is to leverage EMC Data Domain[®] deduplication storage systems to dramatically reduce the amount of disk storage needed to retain and protect enterprise data. With Data Domain, redundant files and data segments are identified and eliminated as they are being stored. Deduplication with Avamar and Data Domain can be utilized under NetWorker centralized management and delivers significant advantage for both VSS-based disaster recovery as well as granular operational recovery of SharePoint.

EMC: Driving complete SharePoint solutions

EMC is committed to delivering products that help you enhance and control your Microsoft Office SharePoint Server deployments. Whether you are looking to design and deploy an enterprise-class, large-scale SharePoint infrastructure to support many thousands of users, or looking to deploy a smaller-scale deployment, EMC offers a robust portfolio of validated solutions, expert services, and industry-leading technologies that can accelerate your time to value and lower your risk of deployment. Benefits include the following:

- Proven experience across diverse SharePoint environments for both physical and virtualized environments

-
- Documented reference architectures and best practices to address a range of user workloads and business requirements
 - Comprehensive, unified, tiered storage portfolio to address current and future workload
 - Broad range of expert services to address planning, design, implementation, and support requirements
 - Strong business alliances and engineering relationships with Microsoft

EMC will convert your strategic vision into an intuitive, easy-to-use, collaboration system that is simple to manage and that will scale with your business.

Conclusion

As SharePoint usage grows with companies of all sizes, solutions for easy, reliable, and fast recovery will become increasingly critical to the success of the platform. EMC NetWorker provides simple, open, and secure backup and recovery with unparalleled performance for SharePoint from system- and database-level disaster recovery to logical content-level granular operational recovery. NetWorker enables IT managers to craft the best SharePoint protection solution to meet their business needs. Through simplified, centralized, and automated backup and recovery in a single solution, NetWorker streamlines operations and meets strict backup and recovery windows. Off-host backup workflows ensure the impact of data protection is removed from your production SharePoint farm. Data deduplication helps reduce the amount of data sent and stored for backup and recovery, lowering impact and costs. EMC is committed to SharePoint and its successful partnership with Microsoft. With EMC your organization can continue to grow and leverage SharePoint with the assurance of tested and validated solutions backed by extensive experience.