datadomain

White Paper

Microsoft SharePoint Data Protection Best Practices with Data Domain

Abstract

Microsoft Office SharePoint Server 2007 and SharePoint Services 3.0 provide IT organizations with a new breed of application to better help businesses with content management, collaboration, search, business intelligence and work flow processes. Like any business-critical application, providing data protection for production configurations must be carefully planned and coordinated. Data Domain deduplication storage offers a powerful data protection solution by reducing redundant data across backups and providing a cost-effective replication option for simpler disaster recovery.

This paper looks into some of the options and strategies for developing a data protection plan for your Microsoft SharePoint environment.

Microsoft SharePoint Data Protection Best Practices with Data Domain

Table of Contents

1 Introduction

In many companies today, Microsoft Office SharePoint Server (MOSS) 2007 is being deployed to provide a portal for centralized information sharing. SharePoint enables business collaboration, by combining file-based content with unstructured content, such as blog entries and discussion threads, and storing it all in databases. Portal services and Office applications can then leverage the capabilities of the SQL relational database engine to access this information in powerful new ways. In many cases, organizations are using SharePoint to move away from Exchange as the primary repository.

With this integration of documents, database, collaboration and control, SharePoint administrators must take an application-centric view of data protection. Consideration for the database, application and user access – either web or from Office applications – needs to be worked into any data protection plan. In larger organizations, the complexity of the SharePoint application deployment can rise quickly and as more corporate data is organized into SharePoint, the more business critical this application becomes.

This paper will focus on data protection strategies for Microsoft SharePoint systems which best leverage the capabilities of Data Domain deduplication storage. The intended audience is application architects and storage administrators involved in the planning and deployment of data protection solutions for SharePoint environments using disk-based or virtual tape technology. As such, a familiarity with the SharePoint architecture and basic backup and recovery practices is assumed.

There are several new features in SharePoint that affect availability and recovery mechanisms - these will be addressed as they relate to or affect the typical backup and recovery processes that most administrators face on a daily basis. SharePoint primary storage layout will not be discussed in detail, except as it affects the practices described for the backup/recovery discussions.

The information in this paper has been collected from several sources including Microsoft, 3rd party backup solutions and other Data Domain technical papers. For the most current information, please refer to your vendor's specific solution or the links provided at the end of this document.

2 Data Domain Technology Overview

A Data Domain deduplication storage system is an appliance that is typically used as a target for backup and archive data. The characteristics that make it ideal for this use include:

- Support for most conventional backup applications through multiple protocols for Network Attached Storage (NAS) interfaces (both CIFS and NFS) over Ethernet, a VTL interface option over Fibre Channel, and product-specific interfaces such as Symantec NetBackup 6.5 OpenStorage (OST)
- High-speed, inline deduplication using small, variable-sized segments to identify and eliminate redundant data
- Data Domain Data Invulnerability Architecture (DIA): Integrated data protection technologies such as RAID-6, post-backup data verification, and periodic validation checks of existing data sets
- Network-efficient replication over Ethernet to secondary Data Domain systems to automate disaster recovery (DR).

Data Domain systems are available in a range of sizes and performance levels to match almost any SharePoint backup configuration. Usable capacity starts at just over 300 GB and scales to over 30 TB, before deduplication. Throughput rates can range from 50 MB/sec to over 300 MB/sec per system depending on model and configuration. Interface connectivity can be over standard Gigabit,10G Ethernet or over 4 Gb FC. With data deduplication rates that can range from 10x to 30x, these systems are well suited to maintain multiple weeks of SharePoint backups in an online disk-based solution.

Actual usage scenarios may vary so it is important to work with your local Data Domain team to understand details specific to your environment.

3 Microsoft SharePoint Architecture Overview

The Microsoft Office SharePoint Server 2007 solution is a three tier application environment with a web tier, an application tier, and a database tier.

The web tier is made up of stateless web server nodes that route requests to the appropriate application or SharePoint database server. The Web Front End (WFE) servers can also provide load balancing for larger environments.

The application tier is responsible for SharePoint administration servers, end-user web sites, and shared service providers (SSP). These functions can run on one or more physical servers providing scalability and redundancy.

The database tier provides the SQL Server layer that SharePoint Server 2007 and Windows SharePoint Services 3.0 are built upon. In addition to any content databases, there is a site configuration database. SharePoint maintains all site content in SQL Server databases. This includes documents and other content placed into the SharePoint environment including items like search/index information and shared services content.

These three tiers can run on a single server for smaller environments or across multiple and redundant servers for larger enterprise configurations. For multi-server configurations, the notion of a SharePoint "Farm" adds another level of management complexity.

The general structure of SharePoint is depicted in Figure 1 below.

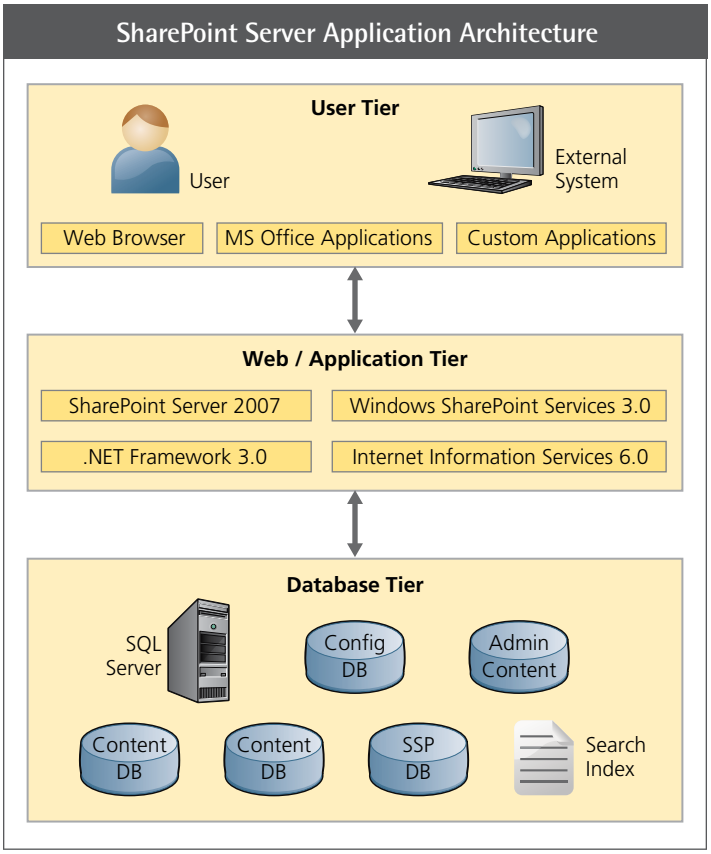

Figure 1 - SharePoint Server Application Architecture

4 Retention Level Objectives – The Hidden Service Level Agreement

Most people associated with data protection, backup, recovery and DR tasks are now familiar with the service level agreements (SLAs) used with the solutions they implement. As IT infrastructures get larger and more complicated, there is another emerging SLA that needs to be considered.

For lack of a better term, this new SLA can be referred to as the Retention Level Objective (RLO). In its simplest form, this SLA helps define the relationship between a particular Recovery Point Objective (RPO), Recovery Time Objective (RTO) and the administration aspects required to deliver them. Before delving deeper, the current SLA definitions and how they might relate to each other should be examined.

4.1 RPO and RTO – The Standard SLAs

According to SNIA: (http://www.snia.org/education/dictionary/r/)

Recovery Point Objective (RPO)

The maximum acceptable time period prior to a failure or disaster during which changes to data may be lost as a consequence of recovery.

Data changes preceding the failure or disaster by at least this time period are preserved by recovery. Zero is a valid value and is equivalent to a "zero data loss" requirement.

Recovery Time Objective (RTO)

The maximum acceptable time period required to bring one or more applications and associated data back from an outage to a correct operational state.

These definitions are fine for individual data protection reference points within a system, but how do you define the multiple conditions that may exist in a typical (operational) data protection implementation for a product suite as complicated as SharePoint?

Within SharePoint, there are multiple ways to provide different levels of RPO/RTO SLAs, each of which has an implication on what data is protected, where it is maintained and for how long. It is the latter criteria which drives many implementation decisions.

4.2 Defining RLO – One Possible Approach

When looking at a complex collection of components and protection strategies that could exist in a SharePoint deployment, it is likely that your business will define multiple types and levels of RPO and RTO requirements. The associated Retention Level Objectives would need to capture the following information at minimum:

- **Recovery Point** – which copy of the data is being used to meet this objective? For example; deleted object from the Recycle Bin, previous version from the object repository, alternate copy from a storage snapshot, restored copy from Full or Incremental backup, etc.
- **Retention Period** – how long is this copy of the data maintained in the system? For example; deleted items kept until certain storage levels obtained, up to "N" number of online versions maintained in SharePoint, snapshots on disk for one week, backups maintained for 90 days, etc.
- **Retention Location** – where is this copy of data maintained? For example; deleted items and versions are in active SharePoint structures, snapshots are maintained by storage platform, backups kept on disk on site for RTO < 24 hrs, backups kept off-site for RTO > 24 hrs., etc.

Complex environments like SharePoint can keep and track multiple copies of objects within the application and backup applications must regularly extract this data. Therefore, the potential amount of storage required to properly protect the solution can quickly grow. However, by leveraging a Data Domain deduplication storage system, the impact of keeping more information on hand and on site to improve recovery SLAs becomes an easier solution to maintain.

5 SharePoint 2007 Data Protection Best Practices

When developing a data protection plan, it is often useful to start with an understanding of what needs to be protected and the tools available to perform the various data protection (backup) tasks. From here you can develop a backup and recovery plan that supports the organization's needs and includes the right people, processes and infrastructure components.

5.1 What needs protecting?

There are many parts or objects in the MOSS 2007 environment that can be protected and recovered. They are:

- Farm (entire collection of objects for a given site)
- Configuration and Central Administration content databases
- Content Databases (SQL Server)
- Third Party Databases
- Search SSPs (Shared Service Providers)
- Site Collections
- Web Applications
- Web Sites
- List / Document Libraries
- Document Library Folders
- Document Library Files
- List items
- Versions
- IIS Configuration Settings
- Customization Files
- System / Server Settings
- The 12 Hive (site specific settings)

There are several methods and tools for protecting the objects in SharePoint. They are:

- Recycle Bins (built into SharePoint)
- Versioning (integral feature for collaboration and development)
- Web Delete Event (trigger to support internal copy operations)
- stsadm.exe Backup and Recovery Operations – (built-in command line tool, only supports UNC path)
- Office SharePoint Server Central Administration
- SQL Server Backup / Restore
  (available with full version of SQL server)
- Systems Center Data Protection Manager 2007 (DPM)
- Windows Server Backup Tool
- Supported Third Party Backup Tools
- Documented Change Procedures
- Reinstall from Distribution Package

5.2 Data Protection Planning

With an understanding of the objects and the methods available, you can now determine the plan (or plans) for the backup solution. There are a couple of additional key considerations that can now be addressed.

The first decision to make is how each object will be handled. The choices are to perform a Full or Incremental backup. Incremental backup methods (differential or cumulative) should be considered based on business and operational needs.

When using Data Domain deduplication storage, we recommend Full backups whenever possible. While this will usually require more time for the actual backup operation, it provides a clearer and often easier to administer recovery point (less restore and playback). Full backups do not incur significant overhead from the Data Domain systems due to its deduplication capabilities.

The next decision is to determine the specific SharePoint components protected in any one operation, whether it is the entire Farm configuration or just a selected set of specific components, depends primarily on the structure and complexity of the SharePoint environment and the targeted SLAs. For larger environments, it may be necessary to protect individual parts across several different backup operations and servers.

Full backups do not incur significant overhead from Data Domain systems due to its deduplication capabilities.

In any case, it is recommended that all SharePoint backup save sets be directed to a common Data Domain system. This will help leverage the commonality and deduplication results achieved. If possible, other Microsoft data sets not currently included in the SharePoint configuration (e.g., file shares, Public Folders, Exchange, etc.) could also be targeted to the same Data Domain system. The deduplication effects of the Data Domain system can thus be further leveraged across multiple related Microsoft application spaces that may not be in the current SharePoint environment.

The Data Domain systems can be presented to the local servers through a UNC (CIFS) path or in some cases as a Fibre Channel Virtual Tape Library (VTL). For environments that may include UNIX backup servers, NFS is also supported. The underlying data will be deduplicated regardless of the protocol or transport used to get the data to the Data Domain system.

Data Domain Replicator software enables network-efficient WAN replication, making it possible to have a copy of the backup data quickly and easily at another remote location. This capability can help better protect SharePoint deployments at remote offices by making copies of the backup data to the primary data center over low bandwidth connections. It can also be leveraged in organizations that have multiple data centers and want a higher degree of DR options between each data center.

One of the final considerations in the planning process involves the scheduling of the backup operations. For the built in Microsoft tools, batch files can be used to run the backup tasks. There are some automated approaches that use qualified third party backup tools. Data Domain deduplication storage can be used as the target for either solution and is supported and easily integrated with all of the leading backup applications.

As a last measure, keep a record of all service packs and hotfixes that are installed on the SharePoint server when you perform backups. SharePoint restore jobs may not complete successfully if the databases were backed up with different patch levels. Some of the recommended recovery procedures for certain components of SharePoint are best handled with a re-install of the software.

5.3 Using Backup Applications

There are several existing third party backup applications that include special agents and options for addressing the protection of all or parts of SharePoint as well as the built in Microsoft backup and recovery functions discussed above. The goal is to choose a solution that satisfies the objectives of the business and can be supported by the IT staff. For larger environments, a third party backup application may provide more functionality and could also be used for other, non-SharePoint, data protection needs.

Regardless of the backup application selected, there are some basic recommendations that will likely improve the results obtained by the use of Data Domain systems. This section addresses these recommendations.

Perform full system backups of the SharePoint servers, including system state, system volume, and any other volumes that contain SharePoint-related files and folders. When you make full system backups, you protect the entire SharePoint environment. There are several locations and sources of data and configuration information. Consult the latest SharePoint data protection documents from Microsoft to make sure all of the necessary items are addressed.

Where possible, make SharePoint backup selections from the appropriate Farm identifier under Microsoft SharePoint Server Farms, rather than from the individual SharePoint servers. Microsoft SharePoint server farms represent the entire farm topology in your environment, which include all of the SharePoint resources that reside on a single computer.

There are several common options in backup applications that should be avoided when using Data Domain systems for backup.

The main reason for not using these options is to ensure that the data sent to the Data Domain system can be effectively processed for the removal of redundant data. The options to avoid are: compression, encryption and multiplexing. All of these choices will alter the data streams from their original content format and thus reduce the advantages of Data Domain systems for effectively storing several copies of the backup data.

Do not perform backup image validation during the backup process. The main reason for verifying backups is to check the tape media which is notoriously unreliable. This time consuming step is essentially eliminated with disk based systems that have the capability to continually check the integrity of the data on that system. For more information, refer to the Data Domain Data Invulnerability Architecture white paper found in the references section.

5.4 Content Recovery

There are some new features in SharePoint that aid in simple content recovery for end users. While these capabilities do not protect against content corruption, they do provide a certain level of protection against item deletion.

Two-Stage Recycle Bin – SharePoint now provides a convenient mechanism for dealing with accidentally deleted files, documents, list items, lists and document libraries without resorting to performing a restore from the content databases. This is accomplished through a multi-stage Recycle Bin capability. Deleted items are first stored in a user level Recycle Bin. When the user empties the first-level Recycle Bin, the items are moved into a second-level Recycle Bin that is then managed by the administrator.

Web delete events – SharePoint can be configured to detect and capture the content from the Web Delete event. When a Web Delete event is detected, the feature archives the site to a file share before the site is removed from the configuration database and the content database. Deleted sites are saved as .bak files in the specified backup directory. You then can use the restore stsadm operation to restore a site. Note that this capability is supported by a third party (non-Microsoft) tool. The Data Domain system used for the backup tasks can also be used for this storage area if needed.

Versioning – SharePoint also provides a document versioning capability that allows users to collaborate better by checking documents into and out of SharePoint content databases. With this capability, it is possible for users to retrieve previous versions of documents without performing a restore operation. By enabling versioning control and providing sufficient "previous versions," SharePoint can provide a certain level of protection against inadvertent changes by using a check-out/check-in model of object control.

With versioning enabled, the content databases will grow in size relative to the change rate and version levels retained. The increase in content can be mitigated during backup by exploiting data deduplication solutions and using a Data Domain system as the backup target.

5.5 Site Recovery

Content databases – these can be backed up and protected with traditional SQL server techniques. Content restores follow the same level of capability as SQL Server databases. You can recover content databases back into the existing site as needed. You can use the native backup capabilities of SQL Server or the database options of your current third party backup tools to manage this aspect of the protection strategy.

Data Domain systems provide an optimal solution for backing up databases – particularly when full backups are preferred. For larger enterprises, the local database administrator (DBA) should be involved with this aspect of SharePoint data protection strategies.

5.6 Disaster Recovery

When planning strategies for full disaster recovery, it is first important to determine if an alternate site needs to be considered. If the DR plan is to fail over to another location, it is critical to have the proper server, storage and backup content available at both sites. Some levels of continuous replication or clustering may be called for as well as a complete data recovery plan.

When looking at a complex collection of components and protection strategies that could exist in a SharePoint deployment, it is likely that your business will define multiple types and levels of RPO and RTO requirements.

When using Data Domain deduplication storage for backup operations at the primary site, you have the option of also deploying Data Domain Replicator software to efficiently replicate SharePoint backup data to an alternate site using low bandwidth connections. With this approach, data that is written during the backup operation is replicated to the DR site. The bandwidth required to support this is significantly less than the original backup operation due to the data deduplication applied before the content leaves the primary site.

Replicating between Data Domain systems can also improve security by minimizing the potential for data on tapes to be compromised or lost. By controlling the network and assets on both ends, the need for tape encryption can be reduced or eliminated.

5.7 Content Archival Systems

With the growth of multi-purpose systems like SharePoint (and Exchange) that can hold increasing amounts of semi-structured data, there is also an option to integrate third party content archiving solutions to help manage the size of these systems over time. Acting as a second tier of storage for less actively used content, these content archiving solutions could also leverage the Data Domain system as a repository for the migrated information.

The combined deduplication effect of archiving and backup will usually be better than archiving alone. This effect comes from the result of data passing through the application as backup data and then again getting stored to the same target through the archiving application. The specifics of these solutions and the results may vary so it is always best to check with your software solutions providers and IT staff to determine the best alternatives.

6 Summary

With the growing functionality and popularity of SharePoint to collect, publish, collaborate and manage corporate information, the need to properly protect this environment grows as well. Protecting your SharePoint environment requires a thorough understanding the options for managing the information maintained inside SharePoint as well as the tools used for performing the data protection and backup tasks.

The following list outlines the basic steps for protecting a Microsoft Office SharePoint Server (MOSS) 2007 environment with Data Domain.

1. Understand the data protection options available for your configuration. Include business units, DBAs and system administrators in the review of the SharePoint architecture for your specific site.
2. Examine available settings within SharePoint for built-in levels of recovery. These include: Recycle Bins, Versioning and Delete event handling.
3. Perform full database and file system backups whenever possible.
4. Do not use compression settings for the backup data or within the backup application. Leverage the data deduplication capabilities of the Data Domain system to yield best results.
5. Do not use encryption settings for the backup data or within the backup application when writing to the Data Domain system.
6. Do not use multiplexing settings for the backups targeted for the Data Domain system. Optimize individual streams and leverage the aggregate throughput of the system.

With the deduplication capabilities of the Data Domain systems, administrators have more flexibility in setting the thresholds for optimal management (and their protection strategies) of SharePoint environments.

Appendix: References

Microsoft links:

http://office.microsoft.com/en-us/sharepointserver/default.aspx

http://www.microsoft.com/technet/windowsserver/sharepoint/V2/reskit/c2861881x.mspx

http://go.microsoft.com/fwlink/?LinkId=102839&clcid=0x409

http://go.microsoft.com/fwlink/?LinkID=108961&clcid=0x409

http://download.microsoft.com/download/1/b/d/1bdfced8-30b2-4dd4-bfd3-1ff44cf76f9d/SharePointDocumentCollaboration_GS_E.ppt

http://technet.microsoft.com/en-us/library/cc706867.aspx

http://technet.microsoft.com/en-us/library/cc287876.aspx

Data Domain links:

Appliance Series	http://www.datadomain.com/products/appliances.html
Data Invulnerability Architecture	http://www.datadomain.com/products/DIA.html
Data Domain Technology	http://www.datadomain.com/products/technology.html
Archiving and Compliance	http://www.datadomain.com/solutions/archiving.html

Other:

http://searchstorage.techtarget.com/magazinePrintFriendly/0,296905,sid5_gci1258014,00.html

http://searchwinit.techtarget.com/generic/0,295582,sid1_gci1319577,00.html#

http://media.techtarget.com/searchExchange/downloads/X_tips_SharePoint_DR_Planning_pdf.pdf%20

http://searchstorage.techtarget.com/magazineFeature/0,296894,sid5_gci1331595,00.html?asrc=SS_CLA_306606&psrc=CLT_5

http://www.quantum-logic.com/sharepoint-revision-control.html

http://eval.symantec.com/mktginfo/enterprise/fact_sheets/ent-entvault7_for_sharepoint_01_2007.en-us.pdf

AvePoint links:

http://www.avepoint.com/assets/sharepoint_whitepapers/Best-Practices-for-SharePoint-Backup-and-Recovery.pdf

http://www.avepoint.com/assets/doc50/DocAve_Backup_and_Recovery_v5.0_Datasheet.pdf

http://www.avepoint.com/assets/sharepoint_whitepapers/MOSS-2007-Backup-Strategies.pdf

http://www.avepoint.com/assets/DocAve-4.5-Data-Protection.pdf

Data Domain | 2421 Mission College Blvd., Santa Clara, CA 95054 | 866-WE-DDUPE, 408-980-4800