Using Hitachi Content Platform v3.1 with Hitachi Data Discovery for Microsoft® SharePoint®

Reducing the Impact of Data Growth in Microsoft SharePoint with the Hitachi Content Platform and Hitachi Data Discovery for Microsoft SharePoint

By Hitachi Data Systems

July 2010
Table of Contents

Executive Summary 3
Overview 4
Technical Description 5
  Hitachi Data Discovery for Microsoft SharePoint 5
  When to Deploy Hitachi Data Discovery for Microsoft SharePoint 8
  Integration with Microsoft SharePoint’s SQL Database 9
  Version Control and Single File Recovery 10
Benefits of the Hitachi Data Discovery for Microsoft SharePoint Solution 11
  Seamless Integration 11
  Low Cost 12
  Lower Microsoft SharePoint Licensing Costs 12
  Reduction in SQL Database Size 12
  Improvement Governance and Reduced Risk 13
  Improved Search and E-discovery 14
Summary 14
Appendix A 16
  Hitachi Data Discovery for Microsoft SharePoint: Feature Overview 16
  Hitachi Content Platform: Feature Overview 17
Appendix B 18
  Techniques to Improve Performance of SQL Database within Microsoft SharePoint 18
Appendix C 19
  Recommendation from Microsoft on Sizing of Storage Environment for Microsoft SharePoint 19
Executive Summary

The Hitachi Content Platform is a robust content platform that enables effective, long term, fixed content data preservation for organizations of all sizes. Designed either to provide an archive appliance or to seamlessly integrate into an existing enterprise storage infrastructure, this high performance, highly available and highly scalable content platform satisfies an organization’s information governance requirements: It ensures the secure long term preservation, protection and fast search and retrieval of valuable business records. Its scalability, open interfaces and ability to provide a secure, immutable repository for unstructured data means that the Hitachi Content Platform can also form the cornerstone of an organization’s data management strategy. With its REST interface (Representational State Transfer, based on HTTP) and its multitenancy capabilities, which provide logical separation of data pools, it can also form the central component of a cloud infrastructure used to deliver cloud services to both internal and external customers.

Microsoft SharePoint is a content management and collaboration solution that is deployed by many enterprises and organizations. In many environments, the data volume is growing rapidly due to the significant growth of unstructured data, which is the most common data type stored in Microsoft SharePoint. Hitachi Data Discovery for Microsoft SharePoint is a Hitachi Data Systems solution that provides organizations a better way to manage their growing Microsoft SharePoint environments. Hitachi Data Discovery for Microsoft SharePoint enables a seamless, policy driven migration of data from Microsoft SharePoint into Hitachi Content Platform v3.1, leaving behind “ stubs” that enable users to continue to access their content through the familiar Microsoft SharePoint interfaces. This "stubbing" provides a streamlined Microsoft SharePoint environment that offers improved performance, faster backups and easy recovery. It enables the environment to support far more content with far less infrastructure growth and without disrupting how users work with their content in Microsoft SharePoint.

The objective of this document is to review the Hitachi Data Discovery for Microsoft SharePoint solution and to highlight the key benefits for organizations that deploy this solution with the Hitachi Content Platform within a Microsoft SharePoint environment.
Overview

The Hitachi Data Discovery for Microsoft SharePoint (HDD-MS) solution is a connector developed by Hitachi Data Systems to automatically migrate Microsoft SharePoint data, on a policy driven basis, from the Microsoft SharePoint repository into either a Hitachi NAS Platform, powered by BlueArc, or a Hitachi Content Platform v3.1 content repository. This document focuses solely on data migrations from HDD-MS to Hitachi Content Platform v3.1. (Note: When Hitachi Content Platform is referred to in this document, version 3.1 is being referenced.)

The movement of data from Microsoft SharePoint can be based on files selected by an individual user or by policies set by schedule or object metadata and applied to a web application, site collection or specific document libraries. If ingestion needs to occur on a given schedule, this can be configured to occur instantly, at a specific time and date, or at a regular interval. Individual users also have the ability to migrate their own data into the Hitachi Content Platform using the standard Microsoft SharePoint interface. Any policies based on objects will supersede any policies set on the document library.

Objects that are migrated to Hitachi Content Platform have stubs left behind to allow seamless access to the data by Microsoft SharePoint users. HDD-MS can also pass to Hitachi Content Platform retention, litigation hold and shredding policies, which are applied to the migrated objects. It is even possible to restore individual files from the Hitachi Content Platform back into Microsoft SharePoint, using HDD-MS, directly from the Microsoft SharePoint interface.

The HDD-MS and Hitachi Content Platform solution is ideal in environments that have one or more Microsoft SharePoint sites and that are seeing appreciable growth in those environments.

HDD-MS is designed to be transparent to the Microsoft SharePoint user, and one of the reasons it is able to interface so well with Microsoft SharePoint is that it integrates directly with the Microsoft SharePoint SQL database. This gives it far more control over the way it handles the movement of Microsoft SharePoint data and metadata. In fact, there are a lot of benefits of the joint HDD-MS and Hitachi Content Platform solution, including:

- Seamless integration of HDD-MS with the Microsoft SharePoint interface makes it transparent to the Microsoft SharePoint user.
- The HDD-MS solution is low cost because it does not require any additional hardware (making it ideal to bundle with Hitachi Content Platform 300 as an inexpensive Microsoft SharePoint solution).
- Licensing costs with Microsoft SharePoint are reduced as less capacity is stored directly within the Microsoft SharePoint databases and fewer SQL servers are needed (each requiring a license).
- The Microsoft SharePoint SQL databases will be reduced in size, leading to deferred production environment storage upgrade spending, and improved backup and recovery times.
- The performance of the SQL database (and hence Microsoft SharePoint itself) improves with the reduction of the size of the SQL database, so the reduction in SQL database size detailed above will result in an improvement in the solution’s performance. (See Appendix B.)
- As data within Microsoft SharePoint will be migrated to a resilient platform with "write once, read many" (WORM) immutability, object authentication, encryption of data at rest, search and access
security, it is stored in a manner more likely to support information governance and regulatory requirements, thereby reducing the risk associated with the data that is migrated.

- The availability of the Hitachi Content Platform v3.1 Search API to HDD-MS means that data in Hitachi Content Platform can be fully searched, both on a full text basis and on its metadata, therefore providing a powerful search capability. This is a feature of HDD-MS with the search option.

HDD-MS comes in two models:

- **Basic** version is designed for a single server SharePoint environment. Features include: file tiering and archiving functionality, single file recovery, retention, bulk restoration, recovery of files and stubs, shredding and litigation hold capability.

- **Farm** version is designed for a SharePoint farm. In addition to offering the features listed above, it only needs to be installed once on the central administration server.

**Hitachi Content Platform Search option**, available for both Basic and Farm, provides access to the powerful search tools of the Hitachi Content Platform from the Microsoft SharePoint interface. This version also includes the ability to recover files from search results. The combination of these features makes Microsoft SharePoint faster, more scalable and more useful than ever.

### Technical Description

**Hitachi Data Discovery for Microsoft SharePoint**

Microsoft SharePoint environments allow for the quick and relatively easy creation of document repositories, which can be accessed from anywhere. However, these installations tend to rapidly proliferate in most environments into which they are deployed. This is due largely to the rapid growth in unstructured data, which most Microsoft SharePoint environments were deployed to manage.

Microsoft SharePoint stores both its data and its metadata in an SQL database. As this database gets larger, the cost of managing it increases, its performance begins to fall off, and backup and restore times become untenable.

The Hitachi Data Discovery for Microsoft SharePoint solution or HDD-MS integrates directly with the SQL database as well as the Microsoft SharePoint GUI to provide seamless look and feel as well as additional functionality to the end users of the Microsoft SharePoint platform (additional functionality will be described later in this document).

Data is migrated from Microsoft SharePoint to Hitachi Content Platform v3.1 with HDD-MS by the use of policies and defined schedules that determine which data is archived and when. Data movement to Hitachi Content Platform by HDD-MS is performed over the HTTP protocol (it uses CIFS and NFS for deployments with Hitachi NAS Platform). Today, HDD-MS has not been certified with the authenticated namespaces of Hitachi Content Platform v3.1, so today (at time of writing) ingestion of Microsoft SharePoint data will be into the default namespace. There are plans for HDD-MS to support Hitachi Content Platform authenticated namespace in a future release.

Figure 1 shows the flow of data in a HDD-MS environment, which can include Hitachi Content Platform as well as the Hitachi NAS Platform family.
Figure 1. Data Flow for Office Data into Hitachi Content Platform v3.1 and Hitachi NAS Platform

Data migration into the Hitachi Content Platform can be performed by individual users on their data, but migration can also be performed on a document library. This data migration ingests data into Hitachi Content Platform with various policies, such as object retention and shredding, which will then be enforced by the Hitachi Content Platform. It is also possible to set litigation hold or release a litigation hold on objects ingested into Hitachi Content Platform via HDD-MS. Ingestion of data from a Microsoft SharePoint document library can be scheduled to run across an entire web application or as granular as the document library level to happen immediately, on a specific date, or at a regular specified interval (e.g. weekly or monthly). This can include object retention policies. Figure 2 shows a Microsoft SharePoint document library and the ingestion scheduler.
Figure 2. Scheduling the Ingestion of Microsoft SharePoint Document Library Data in Hitachi Content Platform v3.1 Using Hitachi Data Discovery for Microsoft SharePoint

Files that are ingested into Hitachi Content Platform will leave behind a stub file in the Microsoft SharePoint Office environment so that users will still be able to access their files as if they were in their previous location (see Figure 3). In this way, users have the same, familiar interface with extra functionality, such as the ability to apply shredding to their data and specify retention periods.
When to Deploy Hitachi Data Discovery for Microsoft SharePoint

As mentioned earlier, the very nature of the Microsoft SharePoint environments and the systematic way in which they are deployed in many organizations means that the Microsoft SharePoint environments soon become very large and difficult to manage. Moving data back out of the Microsoft SharePoint environment would be one solution, but because the Microsoft SharePoint SQL database is managing both the data and its metadata, and uses this metadata to manage the files in its repository, this is not possible without a tight integration with the Microsoft SharePoint SQL database. HDD-MS has a very close coupling with the SQL database that underlies the Microsoft SharePoint structure, and is therefore able to allow users and even scheduled automated policies to tier Microsoft SharePoint data out of the Microsoft SharePoint environment and into the Hitachi Content Platform.

Therefore, the environments most likely to benefit from these solutions are those that:

- Started with a relatively small Microsoft SharePoint deployment but are now struggling with its proliferation through their environment
- Are finding it difficult to back up their Microsoft SharePoint Environment within their given backup window
- Are struggling to meet certain internal service level agreements (SLAs) due to the poor performance of a large Microsoft SharePoint environment
- Are reaching scalability limits in Microsoft SharePoint’s underlying SQL server environment
- Are preparing to migrate from Microsoft SharePoint 2007 to Microsoft SharePoint 2010

The low cost of the HDD-MS solution means it can be deployed for any size environment and will pay for itself very quickly.

A bundled solution of HDD-MS and a Hitachi Content Platform 300 in a Microsoft SharePoint environment can quickly alleviate many of the most pressing issues, but at a very low price point compared to alternative solutions. When a HDD-MS solution incorporates the Hitachi Content Platform 500 as its content repository, you have a scalable solution that will tackle these issues in even the largest Microsoft SharePoint environments.

**Integration with Microsoft SharePoint’s SQL Database**

As mentioned earlier, one of the main strengths of HDD-MS as a solution for migrating Microsoft SharePoint data into Hitachi Content Platform is its tight integration into the SQL database that forms the content repository of Microsoft SharePoint environments. This integration means that it is able to separate the data from its metadata, and then move the data and a copy of the metadata into the Hitachi Content Platform. As Microsoft SharePoint retains control of its metadata, and with the migrated data replaced by a small stub in SQL Server, Microsoft SharePoint maintains its original look and feel, and the functionality it provides is unaltered with the exception of new features that pertain specifically to the archiving of data. These features include the ability to:

- Schedule when archiving takes place based on policy.
- Specify retention periods.
- Enable shredding on objects being ingested into Hitachi Content Platform (which Hitachi Content Platform would then enforce at the hardware level).
- Synchronize changes in metadata.
- Migrate data to Hitachi Content Platform from SQL Server. This is important because of the scalability issues around the SQL database.

Microsoft recommends that the SQL database used with Microsoft SharePoint environments should be at least 200% the size of the data contained within Microsoft SharePoint (see Appendix C). So, if the Microsoft SharePoint environment contained one million documents, with an average size of 100KB each, the resultant 100GB of data would require a 200GB of storage space for the SQL database, based on Microsoft’s own recommendations. If you then allow a factor of between 1.5 and 3 times this amount to cater for anticipated growth, it is easy to see how this can quickly get out of control.

Figure 4 shows how this separation of Microsoft SharePoint data and metadata occurs, with the metadata only going into the SQL database, and the actual data being offloaded into the Hitachi Content Platform.
Documents in the Hitachi Content Platform can still be searched via the HDD-MS search option, a feature designed for Hitachi Content Platform. Also, because information was migrated to Hitachi Content Platform with the required compliance and regulatory settings, risk is reduced for the data.

**Version Control and Single File Recovery**

Because of the granular nature of the data movement into Hitachi Content Platform by HDD-MS, it is possible to retrieve single files from the Hitachi Content Platform back into Microsoft SharePoint. The key feature with this capability is that end users can retrieve their files by themselves, without the assistance of the IT department. Upon retrieval of the file, it can be worked upon and versioned, and then re-ingested into Hitachi Content Platform.

Figure 5 shows the process to recover a single file from the Hitachi Content Platform repository back into Microsoft SharePoint.
Figure 5. Single File Recovery from a Hitachi Content Platform Repository using Hitachi Data Discovery for Microsoft SharePoint

Benefits of the Hitachi Data Discovery for Microsoft SharePoint Solution

The HDD-MS solution in combination with the Hitachi Content Platform provides a cost-effective, easy to manage and efficient method of controlling the growth of Microsoft SharePoint environments. In addition, the seamless integration with Microsoft SharePoint and also with its underlying SQL database, means that the look and feel of Microsoft SharePoint is preserved, with just the addition of a couple of extra options to give end users the ability to archive and retrieve their own data. The following features are some of the main benefits of deploying the joint HDD-MS and Hitachi Content Platform solution in a Microsoft SharePoint environment.

Seamless Integration

With some solutions that can be deployed to help manage Microsoft SharePoint environments, a whole new application is introduced into the environment, with its own interface and its own management requirements. This will typically involve extensive education on the part of all intended users of the new system, and a period of adapting to the new look and feel of the new platform, as well as its functionality. In many cases, the software to help manage the Microsoft SharePoint
environment is module of a much larger software suite (e.g. Symantec Enterprise Vault for Microsoft SharePoint).

HDD-MS is a specialist solution targeted at helping to alleviate problems organizations may be having with their Microsoft SharePoint environments. In terms of the interface presented to the users, due to the level of integration between HDD-MS and Microsoft SharePoint, this is effectively the same Microsoft SharePoint interface to which end users would already been accustomed. As such, the same look and feel of the interface is maintained, training requirements are minimized and users gain the added advantage of extra capabilities to manage their data.

In addition to integration with the Microsoft SharePoint interface, HDD-MS also integrates seamlessly with the Microsoft SharePoint SQL database, which allows it to direct data from the Microsoft SharePoint repository into Hitachi Content Platform, and only send the object metadata into the SQL database. There are many benefits of doing this, which are described in some of the following sections.

**Low Cost**

One of the compelling attributes of the HDD-MS solution is the low cost of the solution compared to other solutions on the market that attempt to deliver a similar function. Many of the competitive products are modules of much larger enterprise software suites, and carry a price tag to match. As the HDD-MS solution focuses solely on Microsoft SharePoint, it is a very streamlined product and therefore very cost-effective. Due to the tight integration with Microsoft SharePoint, management of the solution has minimal cost associated with it and the reasonable cost for the HDD-MS software. The solution is cost-effective enough to be deployed in organizations of all sizes, from the smallest company (or department with its own requirement) up to the largest enterprises. For smaller deployments, a bundled solution of HDD-MS combined with the Hitachi Content Platform 300 can give a low cost solution to resolving issues around Microsoft SharePoint growth, and hence give a better return on investment (ROI) justification for deployment of the solution.

**Lower Microsoft SharePoint Licensing Costs**

Microsoft SharePoint sites are licensed per SharePoint farm or can be available for a single server implementation of SharePoint. Because the HDD-MS solution offloads actual data that would other reside in the SQL database of Microsoft SharePoint, the net effect of deploying it is that the SQL databases and SQL servers are greatly reduced in both size and quantity. This greatly reduces the licensing cost associated with the SQL servers. Similarly fewer Microsoft Windows operating system licenses are required, as well as the physical hardware (servers and disks) to maintain the environment. As such, the total cost of ownership (TCO) of the solution is dramatically reduced.

**Reduction in SQL Database Size**

As mentioned above, one of the main benefits in HDD-MS migrating data from the Microsoft SharePoint environment into the Hitachi Content Platform is that the Microsoft SharePoint SQL database is reduced in size. In addition to the benefits this brings in relation to licensing for the Microsoft SharePoint environment, this also has very beneficial effects when it comes to backing up the Microsoft SharePoint environment. As the database is smaller, backup times are correspondingly reduced, and this alleviates any problems relating to meeting backup SLAs or completing backups.
within predefined backup windows. This reduction in backup size also has the complementary benefit that database restores from the backup would be more efficient (see Appendix B).

**Improvement in SQL Database Performance**

All databases manage internal tables and, when items are added to the database, perform a number of lookup functions as well as database record update functions. The data may also need to be indexed. For a small database, these processes will run fairly quickly, but as the database gets larger it takes longer and longer to perform certain functions. This becomes most apparent when uploading a large quantity of data into the database. The user experience can become seriously impacted, and simply deploying more Microsoft SharePoint sites or SQL servers just increases the overall cost of the solution.

Deploying HDD-MS in the environment, however, ensures that only the metadata of objects migrated to Hitachi Content Platform are stored in the Microsoft SharePoint SQL database. As the size is contained, it runs far more efficiently and the user experience is less impacted as the volume of data grows in the environment.

**Improvement Governance and Reduced Risk**

The Hitachi Content Platform is a WORM repository for the preservation of fixed content unstructured data. In addition to the WORM capability, it also authenticates all data stored within its repository to ensure that it is immutable, that it has not changed or been tampered with in any way. It can also encrypt data, automate the disposition of data upon expiration of its retention period, and even shred confidential information once it has been deleted. As such, it is the perfect platform for the long term preservation and protection of content. Therefore, data placed in the Hitachi Content Platform is far more likely to support information governance and regulatory guidelines than data residing in the Microsoft SharePoint repository. In the event of an audit, data within the Hitachi Content Platform repository will also provide greater evidentiary weight as a result. By migrating data from the Microsoft SharePoint environment into Hitachi Content Platform, HDD-MS therefore is one tool that can support a reduction in litigation risk and increase the level of information governance within the overall solution.

HDD-MS moves the data from Microsoft SharePoint in a manner that maintains the "chain of custody" of the ingested data. This is also important from a data authenticity point of view, and this authenticity can be continuously verified once the data is located in the Hitachi Content Platform.

In addition to this, the fact that the migration of the data to Hitachi Content Platform can be automated by HDD-MS, and that retention, shredding and retention hold or release can be set from within the HDD-MS interface, further reduces the risk of violating information governance requirements. Figure 6 shows how the HDD-MS interface can be used to place an object on retention hold (litigation hold) in Hitachi Content Platform or release the hold if it has already been placed on the object.
Improved Search and E-discovery

Microsoft SharePoint does have its own search interface, but HDD-MS provides a separate search option, which connects to the Hitachi Content Platform search and allows users to search within the Hitachi Content Platform for their data. This powerful search capability supports complex Boolean queries, over 377 file formats and 77 different languages. User have the added functionality to export search results to either XML or CSV files for further analysis, as well as being able to place objects in a result list on litigation hold, which will prevent them from being deleted. This is an important function for legal counsel who are conducting an investigation or responding to a litigation request. With HDD-MS search option, it is also possible for the Microsoft SharePoint user to search non Microsoft SharePoint data stored within the Hitachi Content Platform, allowing for federated searches from the SharePoint interface across all their data from multiple sources.

Summary

The Hitachi Data Discovery for Microsoft SharePoint, when combined with the Hitachi Content Platform, provides cost-effective, scalable, seamless integrated method of migrating data out of a Microsoft SharePoint environment into Hitachi Content Platform. This migration is based on defined policies acting on an entire document library, or individual files migrated by end users, who may wish to store their data with various information governance polices, like retention, shredding or retention hold.

HDD-MS performs its migrations to Hitachi Content Platform by ensuring the data is sent to Hitachi Content Platform rather than to the Microsoft SharePoint SQL database. This drastically reduces the size of the Microsoft SharePoint SQL database. The efficiencies that this reduction brings to backup and restore, licensing costs and Microsoft SharePoint site performance help to make the joint solution a compelling solution for organizations dealing with the ever expanding Microsoft SharePoint environment.
In addition, the information governance benefits of migrating data to the Hitachi Content Platform means the overall litigation risk can be dramatically reduced within the environment when coupled with appropriate practice and procedure.
### Appendix A

**Hitachi Data Discovery for Microsoft SharePoint: Feature Overview**

<table>
<thead>
<tr>
<th>Feature</th>
<th>Capabilities</th>
<th>Value</th>
</tr>
</thead>
</table>
| **Single File Recovery**         | • Ingested files can be retrieved from Hitachi Content Platform back into Microsoft SharePoint.  
                                    | • Retrieved files can be versioned and ingested back into Hitachi Content Platform. | • Users can recover files themselves, offloading IT from recovering backups. |
| **User Roles**                   | • There are three predefined user roles: Microsoft SharePoint Administrator, Site Administrator and End User. | • Restrict user management based on roles. |
| **Retention, Litigation Hold, Shredding** | • Manage retention, litigation hold or release, and shredding per library.  
                                    | • Set capabilities through Microsoft SharePoint interface for files on Hitachi Content Platform. | • Documents within Hitachi Content Platform can be put on litigation hold based on search results. |
| **“Stub” Files in Microsoft SharePoint Database** | • Stubs preserve access to archived data.  
                                    | • Stubs are a fraction of the size of the original file. | • Access to offloaded content is seamless and transparent to end users via Microsoft SharePoint.  
                                    | | • Microsoft SharePoint backups are reduced significantly. |
| **Ingestion**                    | • Ingest individual file or bulk data.  
                                    | • Conduct policy-based ingestion with versioning (retention or shredding).  
                                    | • Employ full audit logging for object ingestion. | • Simplify archival of large data volumes.  
                                    | | • Retain versioning within archive.  
                                    | | • Improve object monitoring, security and provenance. |
| **Custom Metadata**              | • Ingest Microsoft SharePoint metadata into Hitachi Content Platform metadata.  
                                    | • Future changes are synched. | • Retain original Microsoft SharePoint metadata. |
| **Metadata Policies**            | • Ingest data into the Hitachi Content Platform according to policies based on object metadata. | • Automate ingestion according to desired policies.  
                                    | | • Gain greater storage efficiency.  
                                    | | • Improve compliance. |
| **Broad Content and Language Support** | • 370 File Formats  
                                    | • 77 Languages  
                                    | • Full text, metadata and system data indexing  
                                    | • 2 byte character support | • Work with a wide range of content types.  
                                    | | • Work with a variety of user languages. |
## Hitachi Content Platform: Feature Overview

<table>
<thead>
<tr>
<th>Feature</th>
<th>Capabilities</th>
<th>Value</th>
</tr>
</thead>
</table>
| **Archiving and Compliance** | • Move static data to an active archive tier.  
• Data integrity checks, WORM and security.  
• Automated retention management  
• Up to 40PB of usable capacity in a single cluster | • Reduces consumption of SQL Server and primary storage resources.  
• Store and preserve data for long term storage.  
• Ensure compliance. |
| **Business Continuity** | • RAID-6  
• Disk block repair, bad block mapping and loss prevention alerting.  
• Data integrity validation at replica  
• Read failover from replica | • Data is always available.  
• System alerts operators to possible impending disk failure. |
| **Data Protection** | • Specifically designed to prevent data loss and corruption.  
• Low write overhead, which allows for aggressive flushing.  
• Object reference protection  
• Continuous data integrity checking  
• Digital fingerprints for validation  
• Proactive data repair capabilities | • Data does not get lost or corrupted.  
• Data is known to be safe and accessible.  
• Data can be readily tested for accuracy. |
| **Backup Reduction** | • Stubs preserve access to archived data.  
• Stubs are a fraction of the size of the original file. | • Replicated Hitachi Content Platform deployments do not need to be backed up to tape.  
• Experience less risk of data loss or theft from tape transport. |
| **E-discovery** | • Search both migrated SharePoint content and other data on the Hitachi Content Platform.  
• Retention and litigation hold enforcement.  
• Data disposal mechanisms  
• Audit logging | • Enable more detailed search.  
• Ensure adherence to legal and regulatory requirements.  
• Ensure that data is disposed of as allowed.  
• View a detailed record of key data management activities. |
| **Multitenant** | • A single content platform can be split into a number of virtual content platforms or “tenants.”  
• Subdivide tenants into namespaces.  
• Tenants and namespaces are logically separated from each other.  
• Tenants and namespaces can each have unique configurations. | • Each SharePoint server, site, line of business or other grouping can have its own unique archive.  
• Maintain security of content within a shared infrastructure.  
• Use the Hitachi Content Platform for more than just SharePoint archiving. |
Appendix B

Techniques to Improve Performance of SQL Database within Microsoft SharePoint

Set Growth Limit
The SharePoint Operations team set a 100GB growth limit. The reasons for this choice are related to administration ease and SQL Server blocking.

In terms of administration, backing up and restoring databases of a moderate size is faster and less error prone than it is on larger databases. The SharePoint Operations team must sometimes restore lists on a specific database, and dealing with smaller files is easier when the team is copying over the network, mounting the databases and reattaching them to farms.

In terms of performance and impact of user behaviour, smaller database sizes also help. When users delete large lists that are stored in SharePoint databases, the SQL Server-based server must process those requests and complete the deletion. This can create performance issues through SQL Server blocking. Users sometimes experience long render times for other sites that use the same database. From a practical standpoint, smaller database sizes tend to house fewer sites, and if a user deletes a large list, fewer sites are affected if fewer sites are housed on the database.

Limit Database Size to Enhance Manageability
When databases grow they can become less manageable for backup and restore operations, or for troubleshooting. The SharePoint Operations team used a 100GB limit for this reason.

Appendix C

Recommendation from Microsoft on Sizing of Storage Environment for Microsoft SharePoint

Storage
SharePoint Portal Server stores data in SQL Server and full text indexes in the file systems on the search and index management servers. In general, the most important characteristics for determining the amount of storage space required are the total size of the documents stored on the portal site and the total size of the documents included in the portal site index. The following table illustrates the storage requirements for the server roles in a SharePoint Portal Server solution.

<table>
<thead>
<tr>
<th>Server Role</th>
<th>Required Storage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Database</td>
<td>200% of the total size of all documents stored on the portal site</td>
</tr>
<tr>
<td>Index</td>
<td>60% of the total size of all documents stored on the portal site</td>
</tr>
<tr>
<td></td>
<td>The index size is about 30 percent of the size of all documents in its catalogues. Because a copy (a snapshot of the content indexes) is always present, the required space is doubled to 60 percent.</td>
</tr>
<tr>
<td>Search</td>
<td>Number of index servers multiplied by 60% of the total size of all documents stored on the portal site</td>
</tr>
<tr>
<td></td>
<td>When the index propagates from the index server to the search server, there are two copies of each index (each 30 percent of the total document size) on the search server for each index server.</td>
</tr>
</tbody>
</table>

For example, a portal site that stores one million documents with an average document size of 100KB stores 100GB of document data and, thus, requires 200GB of storage space. Adding new portal sites or team sites does not in itself consume much disk space. Each new portal site (without content) consumes approximately 20MB of disk space (in the database), whereas a new site, personal site or portal site area (without content) consumes less than 200KB of disk space (in the database).

Recommendation
Use the preceding table to compute your storage needs. Multiply the results by a factor of 1.5 to 3 to accommodate for future growth.

Corporate Headquarters
750 Central Expressway
Santa Clara, California 95050-2627 USA
www.hds.com

Regional Contact Information
Americas: +1 408 970 1000 or info@hds.com
Europe, Middle East and Africa: +44 (0) 1753 618000 or info.emea@hds.com
Asia Pacific: +852 3189 7900 or hds.marketing.apac@hds.com

Hitachi is a registered trademark of Hitachi, Ltd., in the United States and other countries. Hitachi Data Systems is a registered trademark and service mark of Hitachi, Ltd., in the United States and other countries.

All other trademarks, service marks and company names in this document or website are properties of their respective owners.

Notice: This document is for informational purposes only, and does not set forth any warranty, expressed or implied, concerning any equipment or service offered or to be offered by Hitachi Data Systems Corporation.

© Hitachi Data Systems Corporation 2010. All Rights Reserved. WP-375-A DG July 2010