

# Archiving Microsoft® Office SharePoint® Server 2007 Data with the Hitachi Content Archive Platform and Hitachi Data Discovery for Microsoft SharePoint

Implementation Guide

*By Art LaMountain and Ken Ewers*

December 2009



## Summary

The unchecked growth of Microsoft® Office SharePoint® Server 2007 content databases is one of the hidden costs of rapid SharePoint adoption. As more data is added to SharePoint, backup and recovery times of the content databases increase, and infrequently accessed SharePoint files consume expensive tier one storage resources.

Hitachi Data Discovery for Microsoft SharePoint automatically archives selected SharePoint files to the Hitachi Content Archive Platform, a highly scalable, highly available digital archive. Files ingested to the archive are replaced by a small stub file. This gives users access to the original content and reduces the size of SharePoint databases.

Each archived file has flexible settings for retention period, hold-release, and digital shredding. This satisfies many regulatory compliance requirements for retaining SharePoint data in a secure unaltered state for a predefined amount of time. Data Discovery for Microsoft SharePoint also provides a full context search for direct access to files in the Content Archive Platform to support legal discovery processes.

**For best results use Acrobat Reader 8.0.**



## Contributors


The information included in this document represents the expertise, feedback and suggestions of a number of skilled practitioners. The author recognizes and sincerely thanks the following contributors and reviewers of this document:

- Keiko Harada
- Ron-An Lee
- Lisa Pampuch



# Table of Contents

<b>Solution Overview .....</b>	<b>2</b>
Security .....	3
Ingestion.....	3
Mass Ingestion .....	4
File Path Names.....	4
Stub Files in SharePoint.....	4
Single File Recovery .....	4
Metadata and Search .....	4
Litigation Hold and Release .....	5
Content Archive Platform Delete .....	5
Installation .....	6
<b>Solution Components .....</b>	<b>6</b>
Microsoft Office SharePoint Server 2007 .....	6
Hitachi Content Archive Platform .....	6
Hitachi Data Discovery for Microsoft SharePoint.....	6
<b>Tested Deployment .....</b>	<b>7</b>
<b>Deploying the Solution .....</b>	<b>8</b>
Configuring Hitachi Content Archive Platform .....	8
Installing Hitachi Data Discovery for Microsoft SharePoint.....	9
Deploying Hitachi Data Discovery for Microsoft SharePoint.....	9
<b>Using the Solution.....</b>	<b>11</b>
Reducing Database Size.....	11
Archiving Documents for Compliance .....	12
Performing Legal Discovery .....	14



# Archiving Microsoft® Office SharePoint® Server 2007 Data with the Hitachi Content Archive Platform and Hitachi Data Discovery for Microsoft SharePoint

## Implementation Guide

Because Microsoft® Office SharePoint® Server 2007 is a hugely popular platform with users, most SharePoint environments host a constantly growing variety of applications, workflows and documents. For many users, SharePoint is now considered a standard part of the IT infrastructure. Like other components of the infrastructure, SharePoint has demanding availability service levels and often hosts essential corporate information that is subject to mandatory compliance requirements.

SharePoint relies on many back-end Microsoft SQL Server databases to support the configuration and administration of the platform and to store user content. As more content is added to SharePoint, the content databases grow. Left unchecked, this database growth can become a problem.

Although Microsoft states that SharePoint has no hard and fast limits on the application's database sizes, larger databases take longer to backup, longer to restore and can slowly degrade performance. This places a strain on the backup window and recovery time objectives (RTO) and affects response time service levels. Maintaining infrequently accessed files on tier one storage resources can also raise the overall cost of the SharePoint infrastructure.

SharePoint's solution to combat content growth is to use quotas to monitor database size. When capacity nears the quota limit, SharePoint sends an alert and you must either reduce SharePoint content or resize databases.

Hitachi Data Discovery for Microsoft SharePoint solves the problem of continually expanding SharePoint databases with integrated SharePoint workflows that archive content to the Hitachi Content Archive Platform. This also addresses the need for secure, long-term storage to satisfy compliance requirements and provides a single source for legal discovery searches. This solution allows you to achieve the following goals:

- **Reducing the size of production SharePoint databases** — Data Discovery for Microsoft SharePoint is installed as a SharePoint Web application and automatically transfers selected files to a secure external digital archive. The process of moving files to the Content Archive Platform is transparent to end-users. Each archived file is replaced by a small 1KB stub file. This reduces the size of the SharePoint content databases, alleviating stress on the backup window, speeding SharePoint recovery, maintaining optimal end-user response times and freeing tier one storage for the most frequently accessed SharePoint files.
- **Archiving for compliance** — The active archive provides secure digital storage for structured, unstructured, and semi-structured digital content. Data Discovery for Microsoft SharePoint integrates into existing SharePoint workflows to archive critical corporate files for compliance. Retention period, hold-release, and digital shredding attributes are set for an entire SharePoint library, with the option to customize settings for individual files. The archive also uses robust content authentication and supports digital signatures for long-term data integrity.

- **Supporting legal discovery** — Data Discovery for Microsoft SharePoint provides direct search access to documents in the Content Archive Platform. The search capability is intended for use in legal discovery processes. Search provides access to all archived SharePoint documents, no matter which site collection originally archived the data. It also provides access to contents of the archive ingested by other applications, such as Microsoft Exchange e-mail messages.

This document provides a solution overview and deployment information for Data Discovery for Microsoft SharePoint. It is intended for use by SharePoint administrators, storage administrators and any IT personnel involved with governance and compliance of SharePoint data.

The processes described in this document assume that the Content Archive Platform and SharePoint are installed and available.

## Solution Overview

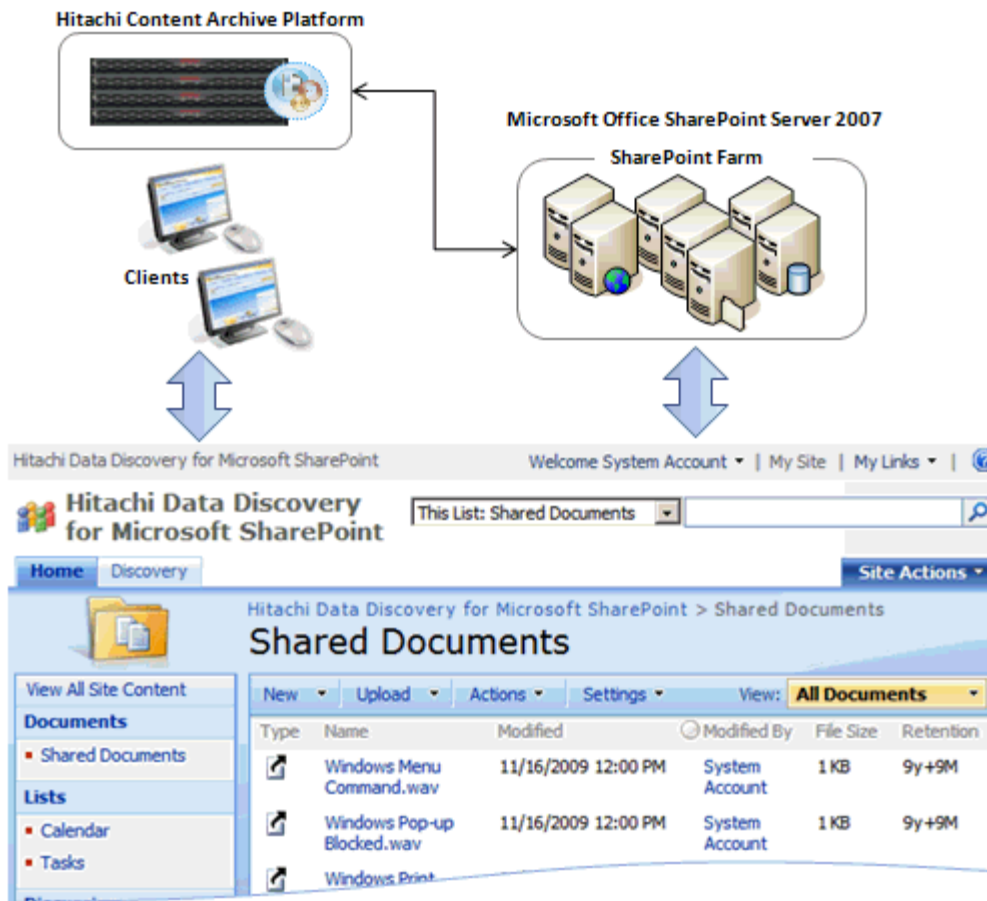
Data Discovery for Microsoft SharePoint archives files from SharePoint libraries to the Content Archive Platform. It supports assigning retention policies, enforcing litigation holds, in the event that content must be retained in an unaltered state, and disposition or shredding of archived files. Files and metadata are ingested to the archive individually or in bulk.

In addition to archiving, Data Discovery for Microsoft SharePoint advanced option provides a full context search capability. This enables direct search of all content in the Content Archive Platform from within SharePoint. This powerful functionality overrides SharePoint security and gives the user access to content from SharePoint and non-SharePoint applications ingested to the archive. It is intended for use by corporate legal and compliance departments during legal discovery and compliance audits. Full context search supports RSS and ATOM feeds for real-time publication and subscription to saved archive searches.

For full context search, the Content Archive Platform must be configured with at least one search node. During Data Discovery for Microsoft SharePoint deployment, you are prompted for interface ports to the Content Archive Platform search nodes.

Figure 1 shows the infrastructure components that require coordinating when deploying Data Discovery for Microsoft SharePoint: Microsoft Office SharePoint Server 2007, Content Archive Platform, and the Data Discovery for Microsoft SharePoint Web application.

Figure 1. Hitachi Data Discovery for Microsoft SharePoint Solution Architecture



## Security

Data Discovery for Microsoft SharePoint uses SharePoint's standard security concepts and structures. No external requirements or overhead, above and beyond the standard SharePoint site authentication and permissions, are needed to manage libraries configured for Data Discovery for Microsoft SharePoint archiving. When a site library is provisioned with Data Discovery for Microsoft SharePoint, each user permission is implemented to each site and respective library.

Each SharePoint site within a Site Collection can have unique permissions to restrict access to authenticated users only. If using Kerberos or NTLM security, the users must reside in the Active Directory to access SharePoint. Access permissions can be set to Read Only, Contributor (Read/Write) or higher level of collection administration. Sub sites within a higher level site inherit permissions from the parent site. Unique permissions can also be configured

## Ingestion

Files are ingested to the digital archive in accordance with each SharePoint library's Data Discovery for Microsoft SharePoint action schedule settings. Settings dictate how often library files are ingested, including options for one-time immediate ingestion, how long files are retained in the archive and how the files are disposed of after the retention period expires.

Sub folders automatically inherit the archive properties of a library. The ingestion process mirrors the SharePoint folder hierarchy in the Content Archive Platform.

## Mass Ingestion

Data Discovery for Microsoft SharePoint gives you an efficient mechanism for performing one-time ingestion of large numbers of SharePoint files to the Content Archive Platform. The Web application can be deployed to existing SharePoint libraries and custom settings can trigger the immediate ingestion of all library files.

Alternatively, Data Discovery for Microsoft SharePoint can be configured to mass ingest files uploaded to SharePoint. Uploaded files automatically inherit the retention and shredding attributes of the library. Data Discovery for Microsoft SharePoint archive frequency can then be configured to ingest files to the archive on a predefined schedule.

## File Path Names

Documents archived to the Content Archive Platform are stored in folders with the name /fcfs\_data[/<part of the website address>]/Archive/<folder structure of the Document Library>.

For example, the SharePoint document path `http://HDD-MS.com/Archive/dev/readme.doc` has the Content Archive Platform path `/fsfc_data/Archive/dev/readme.doc`.

When Data Discovery for Microsoft SharePoint is deployed to the site `http://HDD-MS.com/sites/HR`, the SharePoint document path is `http://HDD-MS.com/sites/HR/Archive/dev/readme.doc`, and the path to the archived document in the Content Archive Platform is `/fsfc_data/sites/HR/Archive/dev/readme.doc`.

Data Discovery for Microsoft SharePoint will reject ingestion of files from the same location with the same name, unless SharePoint versioning has been enabled.

## Stub Files in SharePoint

After a SharePoint file is copied to the Content Archive Platform, Data Discovery for Microsoft SharePoint sends the original file to the SharePoint site recycle bin and replaces it with a 1KB stub file. The stub file retains the metadata associated with the original file and has a pointer to the archived file in the Content Archive Platform. Users can access the archived file from the stub file and can perform single file recovery.

## Single File Recovery

Individual files can be recovered from the Content Archive Platform using single file recovery. The recovery process returns the file to its original location in the SharePoint library.

If you intend to return the recovered file to the Content Archive Platform, SharePoint versioning must be enabled. If versioning is not enabled, the ingestion process fails with a duplicate file error.

## Metadata and Search

Data Discovery for Microsoft SharePoint is integrated with SharePoint search to enable users to locate archived files from within SharePoint. Archive folders support predefined and custom metadata. Metadata is implemented as standard SharePoint library columns. Each library can have unique columns and custom metadata. All metadata values are ingested to the Content Archive Platform with file content.

You can configure SharePoint to crawl the archive libraries and index the stub file names and metadata values. This enables users to locate archived files using SharePoint search.

Table 1 shows the predefined metadata attributes and possible values set for all Data Discovery for Microsoft SharePoint archive files.

**Table 1. Data Discovery for Microsoft Sharepoint Default Metadata**

<i>Metadata Attribute</i>	<i>Description</i>
Shredding	<p>Shredding setting for the file. The setting stores the following values:</p> <ul style="list-style-type: none"><li>• <b>Yes</b> — The file is set to Shred on Delete.</li><li>• <b>No (default)</b> — The file is set to Do not shred on Delete.</li></ul> <p>On ingestion, the file in the Content Archive Platform is updated with this value.</p>
Retention	<p>File cannot be deleted from the Content Archive Platform until the retention period has expired. The retention period metadata column can have the following values:</p> <ul style="list-style-type: none"><li>• <b>Retention period</b> — Choose a period years and months.</li><li>• <b>Do not use retention</b> — File is archived for an unspecified term.</li></ul>
Archival Schedule	<p>Determines when a document is archived.</p>
Archival Status	<p>Status of the file being archived, as follows:</p> <ul style="list-style-type: none"><li>• <b>Archived</b> — File is successfully archived</li><li>• <b>Duplicated File Failure</b> — Same file exists in the same folder in the archive.</li><li>• <b>Archive Failed (xxx)</b> — File is not archived. "xxx" is the reason code.</li><li>• <b>Not Processed (default)</b> — File is not processed for archive.</li><li>• <b>Licensed Expired</b> — Archival status shows a "Licensed Expired" status when documents are archived with an expired Data Discovery for Microsoft SharePoint license.</li></ul>
Litigation	<p>Acts on the archived file and stores the following values:</p> <ul style="list-style-type: none"><li>• <b>Hold</b> — The content cannot be deleted after the retention date.</li><li>• <b>Hold Failure</b> — Content might have been deleted from the archive.</li><li>• <b>Release</b> — Held content has been released. The content is deleted after the retention date, or if the retention date is not specified, users can delete the contents in the archive.</li><li>• <b>Release Failure</b> — Content release failed. The content might have been deleted from the archive.</li><li>• <b>Deletion Failure</b> — Deletion of the content in the archive failed. The cause of deletion failure can be one of the following:<ul style="list-style-type: none"><li>○ Litigation hold is set.</li><li>○ Retention date has not been reached.</li></ul></li></ul>

## Litigation Hold and Release

The Content Archive Platform provides the ability to place a records hold on archived data to prevent the modification or destruction of files. A hold and release of a hold can be issued on a specific file, from within successful content queries using the full content search of Content Archive Platform, or on an entire library from the Data Discovery for Microsoft SharePoint library settings. Data Discovery for Microsoft SharePoint full context search also supports locating files with a hold and release status.

## Content Archive Platform Delete

Data Discovery for Microsoft SharePoint can automatically delete documents archived in the Content Archive Platform that have no retention period specified or that have passed their retention date.

## Installation

Hitachi Data Discovery for Microsoft SharePoint application software is installed on the SharePoint Application server running the Central Administration service. After installation, the web application replicates across all web services front-end servers. The solution is deployed to individual sites by implementing the archive template to all new site libraries.

A reset of Microsoft IIS is necessary for the Data Discovery for Microsoft SharePoint installation to complete. The reset can be performed automatically, as part of the installation process, or deferred. It is important to follow all change control processes and rules for your site, as a Microsoft IIS reset, although only momentary, still disrupts access to SharePoint services.

## Solution Components

The following components are used in the solution described in this guide.

### Microsoft Office SharePoint Server 2007

Microsoft SharePoint Office Server 2007 is part of the 2007 Microsoft Office system. The program provides collaboration, workflow, Web application and content management functionality. Microsoft Windows SharePoint Services 3.0 is the base platform for SharePoint. Microsoft Windows SharePoint Services is a free download available to licensees of Microsoft Windows Server. The SharePoint Services object model is used by SharePoint when creating Web pages and application for collaboration and document and content management.

### Hitachi Content Archive Platform

The Hitachi Content Archive Platform is a digital archive providing highly scalable, highly available long-term storage for structured, semi-structured and unstructured data. It supports standard file and Internet protocols to automatically ingest data from multiple source applications, and offers policy-based authentication, protection, retention, and destruction to ensure long-term content integrity.

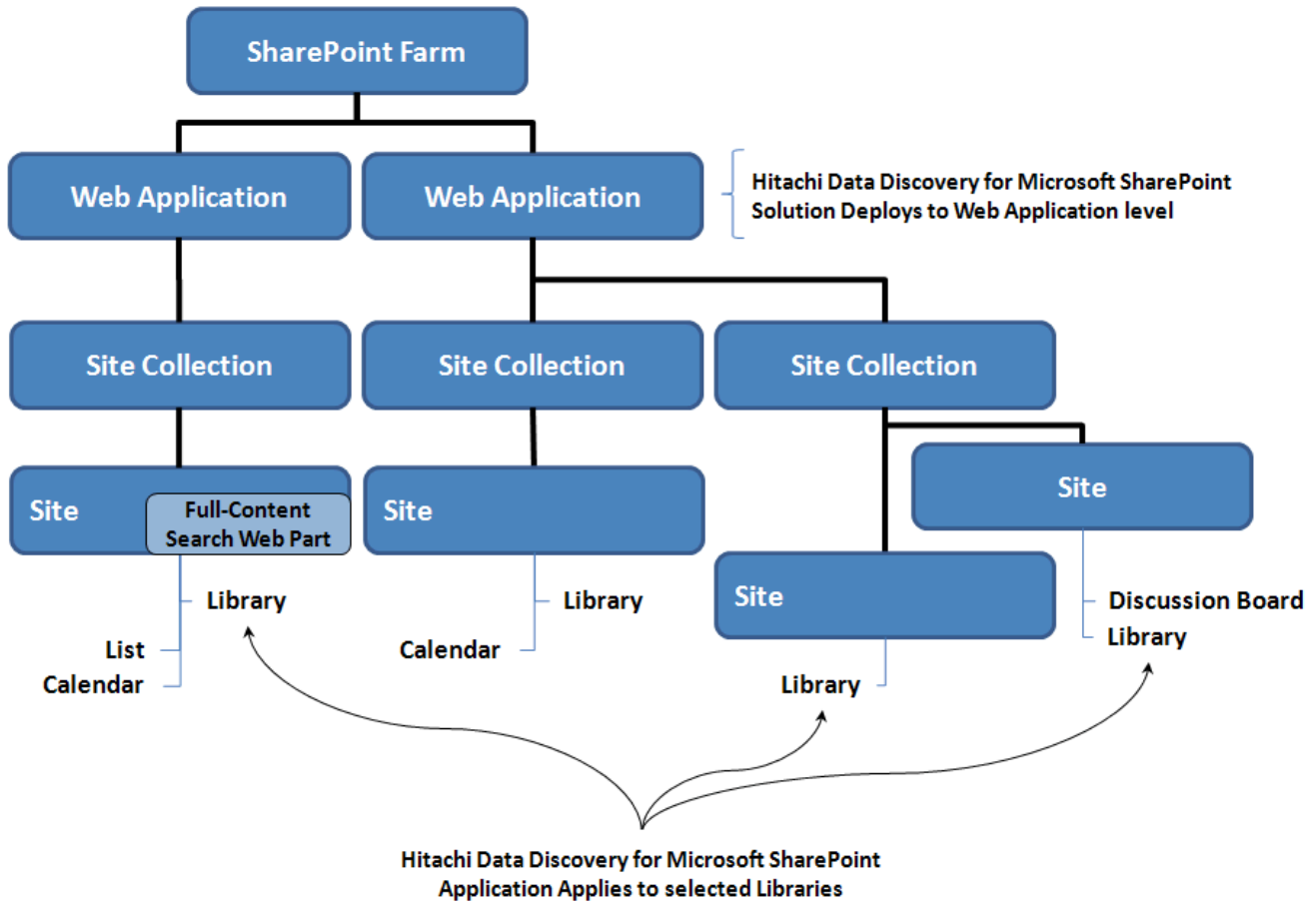
### Hitachi Data Discovery for Microsoft SharePoint

Hitachi Data Discovery for Microsoft SharePoint is installed as a SharePoint Web application and connects SharePoint libraries directly to the Content Archive Platform using HTTP and HTTPS protocols. The Web application mediates movement of documents to and from the archive using integrated workflows. All archiving takes place within the SharePoint environment and is transparent to users.

Data Discovery for Microsoft SharePoint is installed and integrated into the general management features of SharePoint. This gives you access to common SharePoint functionality, like security, and provides a familiar interface for embedded workflows. Data Discovery for Microsoft SharePoint resides in the SharePoint web application hierarchy and can be deployed to as many sites as required.

Figure 2 shows a sample SharePoint schema with Data Discovery for Microsoft SharePoint. The top level shows the SharePoint farm, a single instance of SharePoint Server 2007. The second level shows Web applications. Data Discovery for Microsoft SharePoint is deployed at this level, allowing subsequent integration to all levels below. The third level shows site collections and the fourth level shows individual SharePoint sites with Data Discovery for Microsoft SharePoint physically enabled. Figure 2 also shows optional Web part for full-content search that can be set on any site that has the Data Discovery Web Application deployed.

Figure 2. Hitachi Data Discovery for Microsoft SharePoint Schema



## Tested Deployment

Installation and deployment scenarios detailed in this document were tested using the Hitachi Content Archive Platform version 2.8 with storage and search nodes

Table 2 lists the tested server, browser and Microsoft Office components. Browsers must have JavaScript enabled.

Table 2. Server, Browser and Microsoft Office Components Used in Hitachi Data Systems Lab Tests

<i>Category</i>	<i>Software Product</i>	<i>Version</i>
Server	Microsoft Windows 2008 Server	SP2
Server	Microsoft Office SharePoint Server 2007	SP2
Browser	Internet Explorer	8

For a complete list of supported platforms, see the Data Discovery for Microsoft SharePoint administration guide and the Content Archive Platform administration manual.

## Deploying the Solution

To install Data Discovery for Microsoft SharePoint you need the following information:

- Product license key (temporary or permanent)
- Content Archive Platform configuration, including fully qualified domain name (FQDN)

For complete instructions and guidelines for Data Discovery for Microsoft SharePoint installation, see the installation guide. Hitachi Data Systems and the client support team will review the installation prior to implementation.

### Configuring Hitachi Content Archive Platform

A prerequisite to the final deployment of Data Discovery for Microsoft SharePoint is to setup the Content Archive Platform to allow communications between SharePoint and the Data Discovery for Microsoft SharePoint web application. Content Archive Platform configuration is performed from the system management console.

#### *Hash Type*

Hash type must be the same for the Content Archive Platform and Data Discovery for Microsoft SharePoint. Data Discovery for Microsoft SharePoint uses hash type SHA-256 by default. For more information about hash type, see Data Discovery for Microsoft SharePoint online help.

#### *Enable HTTP/WebDAV*

Perform the following steps in the Content Archive Platform system management console:

1. Log on to the Content Archive Platform **system management console** as administrator.
2. Navigate to the **Protocol Settings** panel.
3. Select the **HTTP/WebDAV** tab.
4. Click **Enable the HTTP and WebDAV protocols** check box.
5. Click **Enable SSL for the HTTP and WebDAV protocols** check box.

Note: If this box is not checked, you might be required to install an SSL certificate. For more information, see your local Hitachi Data Systems service team.

#### *Enable Custom Metadata*

Perform the following steps to enable custom metadata:

1. Log on to the Content Archive Platform system management console as administrator.
2. Navigate to the **Metadata Settings** panel.
3. Select the **Add/Delete/Replace** radio button in the **Custom Metadata** box.

#### *Enable Data Discovery for Microsoft SharePoint Full Context Search*

To enable Data Discovery for Microsoft SharePoint full context search perform the following steps:

1. Log on to the Content Archive Platform system management console as administrator.
2. Navigate to the **Configuration Link** panel.
3. Select the **Enable Search Console** check box.
4. Select the **Enable search facility API** check box.

## Installing Hitachi Data Discovery for Microsoft SharePoint

Follow these steps to install Data Discovery for Microsoft SharePoint:

1. Using CD media or network share, execute the Set up. exe file on the application server where the SharePoint Central Administration service is enabled.
2. Accept the terms of the License Agreement.
3. Enter the product license key.
4. If using a trial license, the license file is located in the License folder on the installation CD.
5. To immediately reset Microsoft IIS, check the box for **Enable IIS RESET** after installation.

You can also leave this unchecked and cycle Microsoft IIS at a later time.

6. Click the **Install** button.

The installation completes.

If the **Open Central Administration of HDD-MS Management** check box is selected, Data Discovery for Microsoft SharePoint Central Administration launches when you click the **Finish** button.

## Deploying Hitachi Data Discovery for Microsoft SharePoint

to the following sections provide instructions for configuring the Data Discovery for Microsoft SharePoint interface to the Content Archive Platform and setting up individual sites requiring deployment.

### *Solution Deployment*

Follow these steps to deploy Data Discovery for Microsoft SharePoint solution from SharePoint central administration:

1. After installing Data Discovery for Microsoft SharePoint, click the **Solution Management** link.
2. Click the **Hitachi Data Discovery for Microsoft SharePoint.wsp** link.
3. Click the **deploy solution** link.
4. On the **Deploy Solution** page, choose **Now** and pick a single Web application or all Web applications to deploy.

The solution takes several minutes to deploy depending on how many Web applications you chose. You can refresh this screen to get an updated status.

5. Manually perform an I I S reset /noforce to activate Data Discovery for Microsoft SharePoint on the application server.

An IIS reset is necessary only for the first deployment of Data Discovery for Microsoft SharePoint.

## *Managing Data Discovery for Microsoft SharePoint from SharePoint Central Administration*

Data Discovery for Microsoft SharePoint management functions are integrated directly with SharePoint central administration. After deploying Data Discovery for Microsoft SharePoint, you can manage the solution by clicking the following links under the **Hitachi Data Discovery for Microsoft SharePoint Management** heading in the **Application Management** tab of central administration.

- **Solution Management**
- **Target Storage Settings**
- **Action Schedule for Document Library**
- **License Management**

### *Solution Management*

The **Solution Management** link shows a list of Web applications in a four-column format. Information displayed includes solution name with port number, URL, zone and virtual directory location.

### *Target Storage Setting*

To connect a Web application and its Site Collections to an archive platform, follow these steps:

1. Click the **Target Storage Setting** link and click the **Add** button.
2. Enter values for the **Title**, **Storage Type**, and **Search Node** (if present) and **Hash Type** (derived from archive configuration) fields.
3. Choose **HTTP** or **HTTPS** from the **File Path** (rop-down menu for the data file path and optional search node).
4. Enter the FQDN for storage and search nodes.
5. Click **OK** to display the **Target Storage Settings** screen.
6. Using the **Target** drop-down menu for each site collection, select the correct target store for the archive interface.

If you are using HTTPS secured sockets access to maintain a secure connection between SharePoint and the Content Archive Platform, name the instance accordingly for ease of reference when performing the final configuration. You can set up one instance for HTTP and one for HTTPS if required. A single Web application can use multiple Content Archive Platform systems, but each site collection can only use a single Content Archive Platform.

It is not necessary to deploy Data Discovery for Microsoft SharePoint to libraries with no archiving requirement.

### *Action Schedule for Document Library*

After registering the target storage with the site collection, set the action schedule for each document library using the following steps:

1. Click the **Action Schedule for Libraries** link.
2. Click the **Setup** button on a corresponding document library.  
The document library setup page displays.
3. Select an appropriate action in the **Choose your action** drop-down menu, as follows:
  - **Archive to Storage** — Archives the entire library based on library retention and shredding values.
  - **Delete from Storage** — Deletes the entire library from the Content Archive Platform.
  - **Change Litigation of File on Storage** — Places the entire library on litigation hold or release.

4. Select an ingestion schedule in the **Schedule Time** section, as follows:
  - **Now** — Performs immediate ingestion for the entire library.
  - **At a specified time** — Performs a one-time ingestion at the indicated date and time.
  - **At every week** — Repeats the ingestion weekly on the specified day.
  - **At every month** — Repeats the ingestion monthly.
  - **At every year** — Repeats the ingestion yearly.
5. Select a retention period in years and months.

If **Do not use retention** is selected, files are saved to the archive until deleted.
6. In the **Shredding** field, select the **Yes** or **No** radio button.
7. For **Litigation action**, select a hold or release option.

### *License Management*

The last link on the **Application Management** tab of central administration is **License Management**. The license is based on a basic or advanced implementation of the application, where advanced includes the Data Discovery for Microsoft SharePoint full context search option.

If you are using a temporary license that is nearing expiration and you need more test time, make a request through your Hitachi Data Systems service or sales representative for an extension.

If your trial is over, you must update the license with a permanent key. Consult your Hitachi Data Systems contact to obtain a permanent key. You might be asked to run a simple command line interface (CLI) command to create a special CSV file, which is submitted through the Hitachi Data Systems key provider group.

After you receive a new key, enter the data in the license field of the License Management page and click **OK** to apply.

## Using the Solution

The following sections describe the steps necessary to use Data Discovery for Microsoft SharePoint for reducing the size of SharePoint databases, archiving SharePoint files for compliance and performing a Content Archive Platform search for legal discovery.

### Reducing Database Size

To perform mass ingestion of files in an existing library that is set up for archive, modify the action schedule for document libraries by performing the following steps.:

1. From the SharePoint Central Administration site, navigate through **Application Management > Hitachi Data Discovery for Microsoft SharePoint Management > Action Schedule for Document Libraries**.
2. Select a Web application.
3. Select the document library.
4. Click **Setup** button.
5. Select **Archive to Storage** from the drop-down menu.

6. Select an appropriate schedule, as follows:

- **Now**
- **At every week**
- **At every month**
- **At every year**

7. Click **OK**.

Data Discovery for Microsoft SharePoint automatically archives non-ingested files in the library in accordance with the selected schedule.

Adding Data Discovery for Microsoft SharePoint to an existing SharePoint library and setting the library's action schedule frequency to Now triggers immediate ingestion of all library files to the archive.

For SharePoint libraries that do not currently have the Data Discovery for Microsoft SharePoint Web application deployed, follow the steps for setting up **Target Storage Settings** in the "Deploying Data Discovery for Microsoft SharePoint" section. After the Data Discovery for Microsoft SharePoint Web application is deployed to the library, you can set up an action schedule to perform mass ingestion to the archive.

## Archiving Documents for Compliance

The procedures in this section describe how to archive SharePoint files for compliance from a new SharePoint library or an existing SharePoint library.

### *Create a New Site Library for Archiving*

From a SharePoint site collection with the Data Discovery for Microsoft SharePoint Web application deployed, use standard SharePoint workflows to create a new library and perform the following steps (permissions are required to provide this function, usually afforded to Site Collection or Site Administrators):

1. Click the **Site Actions > Create** link and choose **Site Library** in the Libraries category.
2. Enter a library name and description.

Use an intuitive naming convention. For example, If this library is to hold Logistical procedures that are to be retained for six years in the archive, you might want to use **Procedures-6Y** as the library name.

3. Click the **Enable Document Version History** check box.

Enable this setting if you need to be able to recover files from archive storage to the SharePoint library.

4. From the newly created **library** page, navigate to **Settings > HDD-MS Document Library Settings for Site Administration**.
5. Follow the steps for setting up **Target Storage Settings** and **Action Schedule for Document Library** in the "Deploying Hitachi Data Discovery for Microsoft SharePoint" section.

This requires input for setting retention and shredding values.

### *Add Data Discovery for Microsoft SharePoint to an Existing Library for Archiving*

For existing SharePoint libraries that do not currently have the Data Discovery for Microsoft SharePoint Web application deployed, perform the following steps:

1. From an existing **library** page, navigate to **Settings > HDD-MS Document Library Settings** for **Site Administration**.
2. Follow the steps for setting up **Target Storage Settings** and **Action Schedule for Document Library** in the “Deploying Hitachi Data Discovery for Microsoft SharePoint” section. (

This page requires input for setting retention and shredding values.

### *Add Custom Metadata*

Perform the following steps to add custom metadata to the archive library. Adding custom metadata is identical to adding columns for a conventional SharePoint library.

1. From the SharePoint library, click the **Settings** link.
2. Select **Document Library Settings** from the drop-down menu.
3. Scroll to the **Columns** section and click the **Create Column** link.
4. Complete the **column name** and **type of data** fields in the column and provide any additional column settings.

All column types are supported.

5. Click **OK**.

The new metadata column is displayed in the library.

### *Upload Files to an Archive Folder*

To upload files to an archive folder, perform the following steps:

1. Navigate to the library folder.
2. Click the **Upload** link.
3. Click **Upload Document** from the drop-down menu.
4. Click **Browse** to locate the files.
5. Select one or more files and click **Open**.
6. Click **OK**.

The file is scheduled for ingestion to the Content Archive Platform based on the schedule in the library’s **Action Schedule for Document Libraries** settings.

### *Manually Archive a File*

To trigger ingestion of a file that exists in a library with the Data Discovery for Microsoft SharePoint web application deployed, perform the following steps.

1. From the library, hover over the file name to display the drop-down menu of actions and select **Archive to HCAP/NAS**.
2. To set retention different to the library default, click the **retention link and set custom** the retention attributes for the file.
3. Click **OK**.

4. To set shredding different to the library default, click the **shredding** link and set custom shredding attributes for the file.
5. Click **OK**.

The file is scheduled for ingestion to the Content Archive Platform based on the schedule in the library's **Action Schedule for Document Libraries** settings.

## Performing Legal Discovery

If SharePoint is configured with the ability to perform Data Discovery for Microsoft SharePoint full context search, a sub site is created beneath the site determined appropriate for this function.

Hitachi Data Systems recommends creating this nested site within the top-level site, mainly for legal departments that require legal discovery functions, and giving it unique SharePoint permissions for discovery searches; that is, do not inherit permissions from the parent site. This limits search access to a group of individuals specifically granted authority to search the archive. Provide permissions to users based on corporate guidelines for legal discovery actions. For more information, see the Installation Guide for Data Discovery for Microsoft SharePoint.

### *Create a Sub Site*

To create the sub site and set up the Data Discovery Web Part for search, follow these steps:

1. Click the **Site Actions > Create** link and click the **Sites and Workspaces** link within the **Web Pages** category.
2. Provide a **Title** and **Description** of the new site.
3. Enter a **URL name** for the URL address.
4. Choose the **Team Site** template.
5. Select the **same permissions as parent** site or **unique permissions** check box as appropriate for this new site.
6. Choose navigational options whether the new site link is to be placed in **Quick Launch** or as a **tab** on parent site.
7. Click the **Create** button.

### *Deploying the Web Part*

Follow these steps to deploy the Web part:

1. Navigate to the new site and prepare to add the Data Discovery Web Part for full-content search.
2. Click the **Site Actions** and **Edit Page** links.
3. Close all existing Web parts.
4. Click the **Add a Web Part** link.

A list of available Web Parts appears.

5. Click Hitachi Data Discovery for Microsoft SharePoint check box under the **Miscellaneous** heading.
6. Click **Add**.

7. Click the **Exit Edit Mode** link.
8. Click the **Quick Launch** link to display a pane with the following options:

- **Simple Search**
- **Structured Search**
- **Advanced Search**
- **Saved Queries**

### *Simple Search*

To define simple search criteria perform the following steps:

1. Click the **Simple Search** tab.
2. Select the appropriate search criteria from the following:
  - **All of these words** — Returns files that contain at least one occurrence of each of the specified search terms.
  - **Any of these words** — Returns files that contain at least one occurrence of one or more specified search terms.
  - **This exact phrase** — Returns files that contain at least one occurrence of the search terms exactly as specified, including any embedded white space.
3. Enter the search term in the text entry box.

### *Structured Search*

To perform a structured search, perform the following steps:

1. Select the **Object Type** for the search.
2. Select the operator syntax as follows:
  - **AND**
  - **OR**
  - **Is**
3. Select a comparison value for the search query.

If you require multiple search criteria, click the **+** button to add another series of selection boxes.

Be sure to select the appropriate join criteria from the **Find files that match** drop-down menu.

### *Advanced Search*

To submit a free-form search query follow these steps.

1. Click the **Advanced Search** tab
2. Enter a search query in the text box
3. Click **Search** to submit the query.

The **Need Help** link provides additional information about searchable metadata fields, the query language syntax and a query parameter reference guide, with query semantics.



### *Saved Queries*

To save queries and execute previously saved queries, click the **Saved Queries** tab.

### *Litigation Process*

After identifying files using Data Discovery for Microsoft SharePoint search queries, use the litigation attribute to prevent deletion of libraries needed for litigation. Alternatively, set the litigation attribute on a per-file basis from the Content Archive Platform.

### *Search Files in the Archive to be Placed on Hold*

To search files in the Content Archive Platform to be placed on hold perform the following step:

1. After running a successful query, use the **Control Operations** drop-drop menu and select **Place Results on Hold**.
2. Click **continue** in the pop-up window.

Note: To release litigation hold, view files currently on hold and change the **Control Operations** option to **Release Hold on Results**.



---

**Corporate Headquarters** 750 Central Expressway, Santa Clara, California 95050-2627 USA  
Contact Information: + 1 408 970 1000 [www.hds.com](http://www.hds.com) / [info@hds.com](mailto:info@hds.com)

**Asia Pacific and Americas** 750 Central Expressway, Santa Clara, California 95050-2627 USA  
Contact Information: + 1 408 970 1000 [www.hds.com](http://www.hds.com) / [info@hds.com](mailto:info@hds.com)

**Europe Headquarters** Sefton Park, Stoke Poges, Buckinghamshire SL2 4HD United Kingdom  
Contact Information: + 44 (0) 1753 618000 [www.hds.com](http://www.hds.com) / [info.uk@hds.com](mailto:info.uk@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 mentioned in this document or Web site 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. This document describes some capabilities that are conditioned on a maintenance contract with Hitachi Data Systems being in effect and that may be configuration dependent, and features that may not be currently available. Contact your local Hitachi Data Systems sales office for information on feature and product availability.

© Hitachi Data Systems Corporation 2009. All Rights Reserved.  
AS-027-00 December 2009