Symantec Enterprise Vault™
Reducing E-Discovery Cost and Risk with Discovery Accelerator

Chris Dooley
December 15, 2006
Symantec Enterprise Vault
Reducing E-Discovery Cost and Risk with Discovery Accelerator

Contents

Introduction .............................................. 6
Purpose of paper ...................................... 6
Target audience ....................................... 6
The discovery challenge .............................. 6
Enterprise Vault Discovery Accelerator solution .............................................. 8
EDRM model ............................................. 9
Preservation and collection .......................... 10
Data at rest: Archiving product application content .... 10
Data at rest: Recovering deleted historical content ...... 12
Data in transit: Journal archiving .................... 12
Processing .................................................. 14
Discovery Accelerator architecture and installation .... 14
Creating multiple discovery databases ................ 16
Setting up cases ........................................ 16
Setting up roles (access control) ...................... 17
Running preliminary searches ........................ 19
Running an official search ............................ 19
Previewing results and refining criteria ................ 22
Combining multiple searches into a single review queue .... 22
Scheduling searches .................................... 24
Handling encrypted and ERM-protected content .......... 26
Integration with Automatic Classification Engine .......... 27
Contents (cont'd)

Review ...................................................................................................................... 29
Review assignment ................................................................................................. 30
Setting up default and case-specific marking schemes ........................................... 32
Dispensation marks versus extra marks ................................................................. 34
The review interface ............................................................................................... 34
De-duplication in the review interface ................................................................. 37
Improving the speed of review with prefetch cache .............................................. 37
Improving the speed of review with keyboard shortcuts ..................................... 38
Allowing external access from the Internet .......................................................... 39
Analysis .................................................................................................................. 39
Tiered review .......................................................................................................... 39
Segregating privileged information ...................................................................... 40
Applying and releasing legal hold ........................................................................ 40
Production ............................................................................................................ 43
Differences between export and production ......................................................... 43
Running the export or production ......................................................................... 44
Data formats for export and production ............................................................... 46
Improving production performance with prefetch cache .................................... 48
Importing output into third-party applications .................................................... 48
Symantec Enterprise Vault: Reducing E-Discovery Cost and Risk with Discovery Accelerator

Introduction

Purpose of paper
This white paper is intended to provide a comprehensive source of information about Symantec Enterprise Vault™ 7.0 Discovery Accelerator and guidance on how to use this powerful application. It is designed to:

• Explain the need for the Discovery Accelerator application
• Differentiate Discovery Accelerator from a typical search application
• Explain how and where Discovery Accelerator fits into the electronic discovery workflow
• Describe how Discovery Accelerator users can quickly and efficiently collect, process, review, analyze, and produce archived content
• Explain how Discovery Accelerator meets industry standards for chain of custody and audit logging
• Provide example usage scenarios in which Discovery Accelerator is commonly used

This paper is not a step-by-step installation or configuration guide and does not replace official product documentation or formal training curriculum that is available separately.

Target audience
The primary target audience for this white paper is prospective or existing Discovery Accelerator users who require detailed information about the capabilities and features of the application. It is assumed that the reader has some familiarity with Symantec Enterprise Vault content archiving. For more information about the core Enterprise Vault archiving platform from which Discovery Accelerator retrieves its content, refer to the Enterprise Vault product documentation.

The discovery challenge
In today's electronic world, organizations are increasingly communicating and storing valuable business-related information in unstructured content repositories such as corporate email systems, file servers, Web portals, and personal archives. Even though these repositories were not initially designed for long-term storage or easy search capability, they are often included in initial search requests from attorneys, auditors, internal HR personnel, law enforcement, and other interested parties look for information on internal or external communications. In addition, an increasing number of external regulations include specific or implied long-term storage and retrieval requirements, even if an organization is not currently involved in litigation or an investigation.
On December 1, 2006, in the United States, revisions to the Federal Rules of Civil Procedure became effective that codified standards regarding how organizations should deal with electronic evidence (electronically stored information, or ESI) in the course of litigation. The rules specified standards related to the requirements for parties to store, preserve, search, retrieve, and produce electronic information in litigation. They even formalized a requirement to disclose data management practices to the other side early in the course of litigation. These rules now apply to any participant in U.S. Federal Civil Court, whether a large, regulated business, a government organization, an educational institution, or a small, private company. Other U.S. court systems and even other countries are now following suit in accepting some of these standards. Please consult “2006 Federal Rules of Civil Procedure E-Discovery and Archiving Impact” by Greg Buckles. Available on www.symantec.com/enterprisevault.

Without a robust archiving solution and a fully audited enterprise search application, manually finding and producing electronic information can be a severe drain on human and capital resources. Organizations must either reassign internal personnel or spend an inordinate amount of money to outsource the project. Even if an organization chooses to outsource e-discovery processing and legal counsel, internal resources usually must be applied to collect information from the internal technology infrastructure, including application servers, enterprise storage devices, personal computers, and even PDAs or portable memory sticks.

With manual search, there is rarely a sense of certainty that all relevant information has been found and produced. In fact, some companies have been penalized for finding additional information after the initial set was delivered. Even without such penalties, however, a lack of information can lead to uninformed decisions about whether to settle or defend against a suit. Many corporate legal professionals want to assess the merits of their case and act accordingly after receiving notice about a pending investigation or litigation. Even though the findings may not be favorable, it is much less costly to settle quickly rather than engage in a prolonged legal process.

Another problem with e-discovery is the amount of duplicate content that is found when an investigation or other legal matter spans multiple employees of the defending organization. Once the IT staff finds and produces the relevant information from each source, it is practically impossible to de-duplicate the identical content, necessitating additional time and resources for IT as well as legal counsel. Furthermore, as many external law firms base the cost of discovery services on the volume of information to review, duplicate content often is directly responsible for the high costs of outsourced discovery and legal counsel.
While most attorneys and investigators have historically focused on the facts uncovered during the discovery process, one current trend in the legal community is to attack a responding organization’s process for managing discovery requests. Without a consistent retention policy, a centralized archiving solution, a discovery application, and a full audit trail for all content archived, reviewed, and distributed, an organization can be at risk for an attack on its process for managing electronic records, whether or not a “smoking gun” is found.

Finally, a common request by opposing legal counsel and government investigators is to temporarily suspend the deletion of content related to a pending legal matter or investigation. This can be a difficult challenge without a fully featured enterprise archiving solution that has the flexibility to suspend the disposition of specific messages and files. Many companies resort to an infinite retention schedule for their application and file server backup tapes. This can be extremely costly given the need to purchase an unknown quantity of tapes and the lack of precision to retain only the relevant information. Furthermore, such a solution provides little assurance that all communications are captured, as it may be possible for some email messages to be sent, received, and deleted before a backup solution can secure that information from the email application server.

Note that almost all of the challenges identified in this section apply to organizations of any size, in any industry, from any part of the world. While government regulations have specific jurisdiction, litigation in general is constantly evolving and expanding in industrialized nations across the globe. Legal procedures worldwide are evolving to deal with the differences between electronic information and traditional paper records. The scope, standards, and risks of outside discovery requests vary tremendously between countries, but every company needs to quickly and securely resolve internal investigations while preserving the authenticity of potential evidence. Plus, in today’s global economy, e-discovery can often span international borders when information must be collected from international colleagues, customers, partners, and suppliers.

Enterprise Vault Discovery Accelerator solution
Symantec Enterprise Vault 7.0 Discovery Accelerator, available in English and Japanese, extends the basic search functionality of Enterprise Vault email and file archiving to help lower the cost of data collection and facilitate the search and recovery of archived items. Discovery Accelerator can collectively search all types of indexed content residing in Enterprise Vault, including approximately 370 file types. It de-duplicates redundant archived items found in multiple searches in both review and production, provides a hierarchical review workflow, applies legal holds to relevant items found in a search, exports some or all of that relevant content using native formats, and audits case activity performed in the application to illustrate due diligence.

This white paper examines in detail the powerful and highly configurable capabilities of the Discovery Accelerator application.
EDRM model

The Electronic Discovery Reference Model (EDRM) project was launched by a nonpartisan industry group in May 2005 to address a lack of standards and guidelines in the e-discovery market. This problem was identified in the 2003 and 2004 Socha-Gelbmann Electronic Discovery surveys as a major concern for consumers and vendors alike. The EDRM model, illustrated in Figure 1, provides a common, flexible, and extensible framework for the development, selection, evaluation, and use of e-discovery products and services.

In Figure 1, records management represents the entire technology infrastructure from which relevant information may be retrieved in its native format. The first step in e-discovery is the identification phase. In this phase, the organization identifies the “custodians” (or parties involved), date range, data types, potential sources for the information, and so on. The next step is the preservation of content that might be related to pending litigation or investigations, even before the formal discovery request is filed. This phase would also be when content potentially related to pending litigation is placed on hold to prevent the alteration or disposal of potentially critical evidence. Then the collection process consolidates the requested content into a central location for processing. During the processing phase, content is indexed and searched so a significant portion of the collected data can be set aside as likely not relevant due to factors such as type, origin, or date. Also, duplicate content can be discarded, as the same information is retrieved from multiple sources. The review process allows authorized parties (internal or external to the organization) to evaluate content found by targeted searches for relevance and confidentiality, or “privilege.” Reviewers often utilize some mechanism to annotate or mark individual records according to how they should be handled and produced later. When attorneys mark an individual
record as privileged, they claim that the information contained in the record is protected from discovery under attorney-client communication or attorney work product privilege, even if it is relevant to the discovery request. During the analysis phase, the discovered information is summarized to re-create a series of events that expose truth in the matter at hand.

Subsequent searches may be requested based on the information originally found. Once the content is categorized appropriately, it passes through production, which involves the assignment of a unique Bates number to each electronic record and delivery of the electronic information to various recipients (such as a law firms, corporate legal departments, and service providers), other systems (such as automated litigation support systems and Web-based repositories), and other media (such as CD, DVD, tape, hard drive, portable storage devices, and paper). In some cases, the original native files are converted to a printed format, usually TIFF or PDF, and unique page numbers are assigned. Finally, presentation involves the use of the prepared content as evidence at depositions, hearings, and trial.

Discovery Accelerator, together with the underlying Enterprise Vault archiving platform, can automate many steps in the e-discovery process, from preservation to production. In some cases, however, the review, analysis, and production steps may be performed by outside legal counsel using case management or analytics software. For the purposes of this white paper, it is assumed that the organization will take full advantage of the many capabilities within Discovery Accelerator. The paper reflects the organization of the EDRM framework, beginning with preservation and ending with production.

Preservation and collection

Data at rest: Archiving production application content

Symantec Enterprise Vault is a flexible content archiving platform that reduces the costs and risks associated with retaining unstructured information such as email and end-user documents. Eight years ago when Enterprise Vault was launched, the main driver for archiving was to relieve the burden of long-term content storage from the primary application infrastructure. Archiving the dormant information to a separate technology infrastructure designed for long-term storage provided direct cost savings and application performance improvements. A few years later, after a series of corporate scandals, new regulatory requirements such as the Sarbanes-Oxley Act in the United States and e-discovery, companies sought an archiving solution to enforce new retention schedules for their business-related email and file records. The goal was, and still is, to eliminate
the use of personal archives that are outside the control of the organization and therefore not in accordance with the organization’s official retention policy. Legal departments, records managers, and corporate compliance officers now advise that content be retained for a consistent period of time—no more and no less.

Enterprise Vault eliminates the need for personal archiving through its policy-driven archiving process, which moves older content out of the primary application store into a scalable and secure archive repository. Enterprise Vault can archive content from a number of content platforms—including Microsoft® Exchange, Lotus Domino®, Microsoft Windows® File Systems, and Microsoft Windows SharePoint® Services—through third-party connectors for instant messages, database records, SAP records, faxes, and more. A retention policy is assigned to each item as it is archived, and under most policies, the end user cannot tamper with or dispose of the archived content until the end of the configured retention period. At this point, Enterprise Vault would automatically delete it permanently (assuming there are no legal holds placed on the item). This retention period can be as short or as long as the organization requires, and it can vary by user, mailbox folder, and even message content. An added benefit of Enterprise Vault is that content, including over 285 file types, can be fully indexed and made available to the end user and authorized personnel who need to search the enterprise archive.

Because many organizations have historically allowed the use of personal archives, an optional component of Enterprise Vault called the Enterprise Vault PST Migrator is available that can locate, collect, and migrate a small or large set of personal archives (PST files) into the Enterprise Vault. This process offers IT and retention enforcement benefits, as information is de-duplicated in the archive, deleted from its original storage location, assigned customizable retention policies, and indexed for inclusion in end-user and/or enterprisewide searches.

It should be noted that end-user archives as described in this section are typically a subset of an organization’s unstructured information. Content is usually archived out of an end user’s repository (such as a mailbox or a file system folder) when it reaches a certain age, or when the application approaches its configured storage limits or quotas. Optional methods to capture and archive all email communications as they are processed by the primary email application are discussed. See section titled, “Data in transit: Journal archiving” on page 12.

Data at rest: Recovering deleted historical content

The scope of e-discovery often reaches well beyond the primary day-to-day application servers that contain email messages and end-user documents. Investigators and legal personnel frequently want to search any residual information that can be retrieved from personal archives or

Syantec Enterprise Vault:
Reducing E-Discovery Cost and Risk with Discovery Accelerator
backup tapes. Based on precedents in case law, organizations are required to perform an exhaustive search across all potential sources of information if deemed necessary to prove a plaintiff's case. In one recent case, an organization was fined heavily for discovering an additional set of backup tapes after the original set of information was produced.1

Many organizations that use Enterprise Vault have had to periodically recover information from backup tapes and archive it into Enterprise Vault. This can be done in-house by setting up a nonproduction lab with servers on which the backup tapes can be restored, and separate Enterprise Vault servers that are not linked to the production Enterprise Vault archive with which end users interact. A separate, discounted Enterprise Vault license called the Enterprise Vault Historical Vault is available to organizations that already have a licensed production Enterprise Vault system.

Some companies, however, prefer to outsource the process of recovering backup tapes and archiving the information into a separate Enterprise Vault system. Symantec has partners that provide the valuable service of taking an organization’s set of backup tapes and delivering a fully indexed, de-duplicated Enterprise Vault archive in return. For more information about these partners, please contact your local sales representative.

Data in transit: Journal archiving
As stated previously, when an end user’s data at rest is archived to Enterprise Vault, it is only a subset of the information that may have existed in the past. Historical information may have been created, stored temporarily, communicated, and deleted by the end user beforeEnterprise Vault archived the information. Organizations that need a complete collection of all email communications must implement a feature in their Microsoft Exchange or Lotus Domino messaging application called server-based journaling. The journaling process automatically creates a copy of every email message that is sent or received through the messaging application and places it into a separate mailbox that end users cannot access or tamper with. Enterprise Vault is then configured to continuously sweep the content out of these system mailboxes, fully index the content, and store the content in an access-controlled logical archive that authorized personnel can access as needed.

In many environments, Enterprise Vault is designed during implementation to de-duplicate the content between the messages that were archived through the journaling process and the messages that were later archived through the policy-driven archiving process. Each unique message can then be physically stored only once in Enterprise Vault, even though it entered the archive through two different methods at two different times.

This journaling process enables compliance by organizations that are governed by external regulations requiring that all email communications be archived. In addition, journaling all email messages while they are in transit might be desirable for organizations that do not wish to produce email messages that have been deleted from the messaging server but that are still present on backup tapes. Journaling is also advantageous for companies that want to be as educated as possible about the facts, positive or negative, and possibly even put controls in place to take corrective action soon after policy violations take place, before they wind up in a subpoena or discovery request. An organization can only be certain about what was communicated via email if journaling was implemented during the date range in question.

Figure 2 illustrates four of the Microsoft Exchange-focused Enterprise Vault product options and the type of content each archives.

Figure 2. Enterprise Vault product options
Discovery Accelerator architecture and installation

Discovery Accelerator is a Web-based application that utilizes Active Server Pages running on Microsoft Internet Information Services (IIS) for the end-user interface, and Microsoft SQL Server to store the configuration, permissions, and case information on the back end. Discovery Accelerator can search and retrieve content from one or more Enterprise Vault directories, and it is typically installed on a dedicated Windows 2003 Server (or Windows Server 2000) for optimal performance.

Before installing Discovery Accelerator, you must first install Enterprise Vault 6.0 SP3 or higher. You only need to run the Enterprise Vault setup program; do not run the configuration program or configure any services or tasks after the setup is completed. Please refer to "Installing and Configuring Discovery Accelerator.pdf" for more details on the full installation process, including software that must be installed prior to Discovery Accelerator. Table 1 lists the primary Discovery Accelerator components.

Table 1. Discovery Accelerator components

<table>
<thead>
<tr>
<th>Component</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Discovery Accelerator Web site(s)</td>
<td>One or more Web sites hosted in Windows IIS through which Discovery Accelerator is presented to the end user. Each site utilizes a separate IIS virtual directory, URL, customer database, and application task with distinct configuration and security. The URL for the first site is <a href="http://servername/EVBADiscovery">http://servername/EVBADiscovery</a>.</td>
</tr>
<tr>
<td>Accelerator administration Web site</td>
<td>Used by the system administrator to create and manage multiple customer databases to partition case information separately. The URL is <a href="http://servername/EVBAAdmin">http://servername/EVBAAdmin</a>.</td>
</tr>
<tr>
<td>Discovery Accelerator Manager Service</td>
<td>Windows service that handles requests from the Discovery Accelerator Web interface and works with the Enterprise Vault server(s) to access archives, perform searches, and so on.</td>
</tr>
<tr>
<td>Configuration database</td>
<td>SQL database that specifies the locations of the customer databases and stores details of the SQL Server, database files, and log files to use.</td>
</tr>
<tr>
<td>Customer database</td>
<td>SQL database in which Discovery Accelerator stores details of cases, user roles, search results, marking schemes, and more. You can set up multiple customer databases.</td>
</tr>
<tr>
<td>Customer task</td>
<td>Server process (per customer database) that performs actions requested by the end user, such as running a search, viewing a message, exporting items, and so forth.</td>
</tr>
</tbody>
</table>
As part of the setup process, the Discovery Accelerator installation program installs Discovery Accelerator Website, Accelerator Admin Website, and Service, as shown in Figure 3. (Note that choosing Custom Setup allows individual components to be de-selected.)

![Figure 3. Discovery Accelerator setup screen](image)

The configuration database is created after you connect to the Discovery Accelerator Web site for the first time and follow the prompts for IIS server name, virtual directory name, SQL Server, path for SQL database and log files, Active Directory user or group that will have full administrative access, and so on. Once the configuration database is in place, you are prompted to create the first customer database. Again, you are prompted for IIS, SQL, and security information. Note that the Enterprise Vault Service Account will need temporary database creator permissions in SQL to create all databases. Please refer to the product documentation for more information on the installation process and specific permissions required.

By default, customer tasks are installed on the local Discovery Accelerator server, but you can choose to process them on another server, separate from the Manager Service. However, note that all tasks for a single customer database are performed by only one server, as discussed in more detail in the next section. Multiple Discovery Accelerator databases can be created to enable different groups to have their own sites, such as HR, audit, investigations, and litigation databases, as detailed in the section titled, "Creating multiple discovery databases" on Page 15 and, "Setting up roles (access control)" on page 17.
Creating multiple discovery databases

Even though access control is extremely flexible within a Discovery Accelerator database to allow case-specific roles and permissions, there are a number of reasons to implement multiple databases.

From a technical perspective, dividing cases among multiple databases helps the solution scale over time, especially as the size and scope of cases grow. As stated in the previous section, each customer database can utilize its own separate server for tasks such as searching, viewing items, producing, and so on. As a result, there are also performance benefits associated with separating cases into multiple databases. Within one customer database, all searching or production requests are serially queued to perform their tasks one request at a time, even if the commands originate from different cases within the database. When searches or productions are executed simultaneously from different customer databases, they perform in parallel because each customer database has its own dedicated set of tasks.

Separate customer databases also provide a separate security schema, with different administrators. This could be useful in environments where there are multiple business units, or when, for example, HR needs its own private database for internal investigations that have nothing to do with external litigation.

From a legal perspective, performing discovery searches in the same customer database as initial legal hold searches could lead to unnecessary exposure if a judge were to request documentation to prove due diligence for legal hold actions. More important, hold searches are, by definition, overly broad. This means that the legal hold database may contain many millions of search hits, which would eventually slow down review and production actions.

To create additional databases after the initial customer database, you must connect to the Accelerator administration Web site using Internet Explorer. As described previously, the administrator will have an opportunity to specify any IIS server, SQL Server, task server, and so on. Each database will have an assigned administrator (or group) who will be able to access the customer Web site, create cases, and assign additional access permissions as needed.

Setting up cases

The person or group designated as the administrator of the customer database can create new cases after logging in to the customer Web site. The administrator can give the case a name, select which Enterprise Vault stores should be included in the search scope, and assign an owner who can fully administer the specific case and assign additional permissions specific to it. Figure 4 shows the case creation screen. The case owner can later modify the case properties, including the name and vault stores, and even relinquish ownership of the case to another person.
Setting up roles (access control)
When the case is first created, only the case owner can access it. When other users log in to the customer Web site, they will not see the newly created case until the case owner grants at least one permission related to the case. There are two types of Discovery Accelerator roles: application and case. Application roles grant permissions related to the structure and configuration of the entire customer database, while case roles grant permissions specific to cases. Roles provide a logical way to group multiple specific permissions according to job function. Discovery Accelerator security requires that the administrator specifically grant a role permission. It is best not to grant group rights to avoid the risk of inadvertently giving someone access to sensitive legal matters.

For example, one person in an organization may be responsible for managing the schedules for searches and selecting which Enterprise Vault archives are included in the scope of the searches. In this scenario, you could create a role called "schedule and archive administrator" and only grant permissions to manage search schedules and select archives to include in Discovery Accelerator.
A different person in the organization may create the searches and assign the search results for others to review, but may not actually have permission to review and mark the actual messages into a category. You could create a role called “search administrator” and only grant the specific permissions for searching and assignment.

Table 2 lists a few of the most frequently used application and case permissions that can be grouped together into as many roles as an organization requires. For a complete listing with descriptions of all application and case roles, please refer to the online help on the Discovery Accelerator Web site.

Table 2. Application and case permissions

<table>
<thead>
<tr>
<th>Application permissions (partial)</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Create and configure cases</td>
<td>Create new cases and assign owners to them, and edit the properties of existing cases.</td>
</tr>
<tr>
<td>Manage administration security</td>
<td>Assign application roles to users (cannot assign case roles to users)</td>
</tr>
<tr>
<td>Manage marks</td>
<td>Create and edit the marks available to all cases (cannot manage case-specific marks)</td>
</tr>
<tr>
<td>Manage schedules</td>
<td>Create, edit, and delete search schedules.</td>
</tr>
<tr>
<td>Manage vaults</td>
<td>Select which vault stores and archives are available for searches by default. (Note: Case administrators can modify the list of archives that are used for particular searches.)</td>
</tr>
<tr>
<td>Case permissions (partial)</td>
<td></td>
</tr>
<tr>
<td>Perform ad hoc searches</td>
<td>Create, delete, and modify personal folders, search for items to store in the folders, and review those items. (See “Running preliminary searches.”)</td>
</tr>
<tr>
<td>Search</td>
<td>Perform searches for items to store in the case review set.</td>
</tr>
<tr>
<td>Assign</td>
<td>Assign search results in the case or folder to individual reviewers.</td>
</tr>
<tr>
<td>Review</td>
<td>Review items and assign marks and comments.</td>
</tr>
<tr>
<td>Manage legal holds</td>
<td>Place holds on the items in a case to prevent manual or automatic deletion.</td>
</tr>
<tr>
<td>Production</td>
<td>Copy items out of Discovery Accelerator so they can be reviewed offline.</td>
</tr>
</tbody>
</table>
Running preliminary searches
During the identification phase of e-discovery, legal departments often hurry to gather facts and perform a quick risk assessment to create a defense or settlement strategy. This process usually occurs between the time the organization's legal department is given notice of pending litigation and the time the organization is given a specific discovery request. Because the official discovery request has not been received yet, many organizations want to perform a series of preliminary searches with an audit log separate from the main audit log that they may be asked to produce during formal discovery. Search results and entire preliminary search result folders can be deleted in this temporary folder area, unlike the main search interface and result set attached to a particular case.

To provide this separate work area for searches, Discovery Accelerator displays a feature called Folders for users who have the permission to perform ad hoc searches. This can be compared with an artist’s scratch pad that is used before an actual painting is started on the canvas. As necessary, folder creators can give other users access to these special folders so they can collaborate on the review and decision-making process. Additionally, items can be exported from these folders in much the same way as items can be exported from the official case review set.

Note that because the audit trails are designed to be separate, search results cannot be transferred from these ad hoc folders to the official case review set. Each review set must be populated by a fully audited search process. Technically speaking, the two different audit logs are stored in separate SQL database tables.

Running an official search
Once the organization is ready to begin running their permanent searches and populating a case’s main review set, a user with search case permission can access the case administration page and perform one or multiple searches for specific content. Many times, an outside party (such as opposing legal counsel or an auditor) will provide a detailed list of search terms such as date range, parties (or custodians) involved, and specific words or phrases.

In Figure 5, the search terms are as follows:

- Date range: January 1, 2004, to March 31, 2005
- Custodians (people) possibly involved with either sending or receiving the captured email messages: Tom Miller, Jane Pearce, or Marcus Sizemore
- Content 1: Subject line must contain either "Confidential" or "Secret."
- Content 2: Subject or content must contain the exact phrase "Patent pending."
- Content 3: Subject or content must contain one of the following three terms: "Capacitor," "2954," or "dielectric."
Figure 5. Search screen

Note that all search fields on the Web page are joined by an AND. Within the text boxes, all terms entered on a single line are joined by an OR. Phrases can be specified by either placing the words in quotation marks or by inserting a period between the words in the phrase. For example, apple.tree equals “apple tree.” To indicate a required term in a line of optional terms, place a “+” sign in front of the mandatory term, without a space between the sign and the term. This is a Boolean AND function within the line. The order of the terms is not important. To exclude messages because they contain a particular term, place a “–” sign in front of the mandatory term, without a space between the sign and the term.
For example, to search for messages that mention "tree" and either "apple" or "pear" but not "rotten," you could enter the terms on a single line as follows:

Apple pear + tree - rotten

One important note is that you cannot use the exclusionary term ("–" sign) as the only criterion in the search argument. In order to exclude items, there must first be a positive search result set from which to exclude the items containing the restricted term. Therefore, you cannot search the entire Enterprise Vault for every message that doesn't contain a certain term. You must specify at least one positive search term and then exclude the restricted term.

When multiple lines are entered within a text box, you can specify whether the lines should be joined by an AND or OR by selecting "All of" or "Any of," respectively, from the drop-down box.

As shown in Figure 5, a single search can use up to four separate text boxes. With the drop-down boxes on the far left side of the screen, you can choose any combination of the following index fields:

- To
- From
- To or From
- Subject
- Content
- Subject or Content
- All

Finally, note that Discovery Accelerator supports the question mark (single character) and asterisk (multicharacter) wildcard characters, provided that at least three characters are supplied before a wildcard character is used. For example, to look for any word beginning with "play" (such as playground, playscape, or playmaker), you could type "play*". You could not type "pl*". To search for five-letter words starting with "the" (such as there or their), you could type "the??". You could not type "th???".

Previewing results and refining criteria
It is common for two parties involved in a litigation or investigation to negotiate the search criteria in the discovery request based on initial findings. For example, if the initial search terms are too broad and the number of search results is overwhelming, additional or different search terms may be requested by either party to create a more manageable set. The Discovery Accelerator search feature was designed with this process in mind. While the search is still processing, a preview list is displayed as soon as a single search result is returned. The user can read messages from the preview display and cancel (reject) the search that is still in process. This conserves system resources and valuable time, as the user can then refine the search criteria and perform the search again.

Combining multiple searches into a single review queue
There are various reasons why multiple searches may need to be performed, and accepted, within a single case. The most common is that search terms can be fluid throughout the discovery process. It is not uncommon to receive a new list of search terms and/or custodians in the latter stages of an e-discovery project. This poses a time-consuming and expensive task for many organizations that do not have a centralized and fully indexed archive, especially when historical content (such as backup tapes) must be re-restored and re-searched.

With Discovery Accelerator, each case can have an unlimited number of searches that contribute to the combined review set. Each search is fully audited, and a log of previous searches and respective search criteria is displayed in addition to the hidden audit log in the SQL database. Figure 6 shows an example of a Discovery Accelerator case in which multiple searches were performed.
In Figure 6, there were a different number of “hits,” or search results, for each of the six searches. Four of the result sets have been accepted and added to the overall case review set, but two of the searches still need to be previewed and accepted before those 1,015 items will be added to the case set.

Discovery Accelerator users often choose to display or export the specific results from one search at a time. Because the same message can be returned in multiple searches and reviewers typically want to see all items that matched a particular set of search criteria, it is important to select the box called “Include items already in review” when running subsequent searches, as shown in Figure 7. Refer to the sections titled, “Differences between export and production” on page 43 and “Running the export or production” on page 44 for further detail on the export and production processes.
Many users mistakenly assume that selecting this box will create duplicate messages in the review queue and the production set. In fact, it will help ensure that overlapping search results are identified as a result from each of the individual searches, without redundantly storing or displaying the content. So, when you later display or export the messages from a particular search, all results matching the specified criteria will be included. If you want to review only the new items and not a comprehensive set of results, do not select this box.

Scheduling searches

While most users typically want to run searches immediately, there are reasons why a search might need to be scheduled, either once or on a recurring basis. Because searches affect the performance of Discovery Accelerator, Enterprise Vault, SQL Server, and the network bandwidth connecting these servers, it may be appropriate to run the very large searches during off-peak hours. Also, if the date range for a case indicates it is still in progress, it may be necessary to...
repeat the same searches on a regular basis, each time adding results to the case set. An example of this would be items subject to a legal hold, as described in the section “Applying and releasing legal hold.” If the legal hold affects current data, the organization would typically implement Enterprise Vault (Exchange or Domino) journaling to capture all electronic communications in a way that end users cannot tamper with, and it would perform a daily search for the specified criteria in Discovery Accelerator to show due diligence. Performing the search once a day is often adequate since the Enterprise Vault journal can be configured with strict access control (see “Chain of custody”).

Note that two different permissions are required to fully configure the Discovery Accelerator scheduling feature. As stated in the “Setting up roles (access control)” section, there is an application-level permission to create and manage schedules for searching. Then, those who have permission to perform a case-specific search can select one of the predefined search schedules.

Behind the scenes, the scheduled searches in Discovery Accelerator utilize Microsoft SQL agents. Figure 8 illustrates some of the options presented when an application administrator creates a predefined search schedule:

For more information about how to configure scheduled searches, please refer to the online help in the Discovery Accelerator application.
Handling encrypted and ERM-protected content

Some organizations today use encryption in their messaging and file systems to control who can see content, particularly when it is in transit. For example, encryption technology allows users to determine whether a given email they are about to send can be opened only by the addressees, with the content encrypted based on each addressee’s public encryption key (which generally must be known in advance). Example encryption systems for internal email include Public Key Infrastructure (PKI) solutions such as Entrust and PGP.

Similarly, many companies are adopting advanced enterprise rights management (ERM) systems that use encryption to enforce rights on what a given user can do with an email message or document. For example, ERM systems allow the sender of an email message to specify whether it should be kept within a certain group (for example, executives) and never forwarded, whether it can be printed, whether it can be sent outside of the company, and so forth. Microsoft produces the leading ERM system, called Microsoft Windows Rights Management Services (RMS). Adobe, EMC, and Liquid Machines also produce solutions in this space.

In PKI encryption as well as ERM systems, item properties (such as To, From, Date, and Subject) are not encrypted and therefore are indexed and searchable by Enterprise Vault. In addition, organizations can use the Enterprise Vault Web search application to identify all records that have been encrypted (and hence cannot be indexed) so legal staff can attempt to decrypt these email messages or files.

Enterprise Vault 7.0 includes an option called the Symantec Enterprise Vault Adapter for RMS that can decrypt RMS-protected message and attachment content before indexing and storing it in the Enterprise Vault journal archive. This adapter also allows Discovery Accelerator to search protected content within the journal archive.

Symantec offers a fully documented custom filter application programming interface (API) to allow other encryption and ERM vendors to integrate with Enterprise Vault. Similarly, Entrust sells an adapter for decrypting PGP and S/MIME, protecting email for storage and retrieval by Enterprise Vault.
Integration with Automatic Classification Engine

The Enterprise Vault Automatic Classification Engine is an optional component that extends Enterprise Vault capabilities to include intelligent email message categorization through predefined and custom smart tagging rules. For organizations that have incorporated this component, the Enterprise Vault journal archiving process sends each incoming message through the intelligent classification engine for content analysis. The classification engine then renders a verdict for the message that includes a policy (or tag) name and Enterprise Vault retention category name. The tag name is stored in the message index and can be searched using Discovery Accelerator.

The Automatic Classification Engine ships with 50 default policies:

1. Attorney-Client Privilege: Labels
2. Attorney-Client Privilege: Secondary Privilege
3. Legal Documents (attachments)
4. Anti-Money Laundering
5. Customer Complaints—Legal
6. Customer Complaints—Service and Support
7. Customer Complaints Responses
8. Auto Generated Messages
10. Attachments—Faxes
11. Attachments—PST, NSF
12. Attachments—PowerPoint
13. Published Information (for example, Research, Marketing)
14. Identity Theft
15. Social Security Numbers
16. Personally Identifiable Information
17. Account Numbers
18. Language—English
19. Language—Non-English
20. Language—French
21. Language—Spanish
22. Language—German
23. Language—Chinese (Chinese character set)
25. Language—Korean (Korean character set)
Symantec Enterprise Vault:
Reducing E-Discovery Cost and Risk with Discovery Accelerator

26. Attachments—Video Files
27. Attachments—Audio Files
28. Attachments—Audio and Video Files
29. Attachments—Graphic Files
30. Internal Use Labeled
31. Draft Documentation (Attachment)
32. Offensive Language—7 Deadly Words
33. Offensive Language—Extended Level 1
34. General Harassment
35. Sexual Harassment
36. Discrimination
37. Religion
38. Jokes
39. Chain Mail
40. Compensation Discussions
41. Solicitations—Political
42. Solicitations—Charities
43. Solicitations—Private Investment
44. Betting
45. Gaming
46. Contact Information Lists
47. Resumes
49. Financial Attachments—Invoices, Bills, POs
50. Financial Communication—Firm Financial Information

In addition, policies can be customized based upon keywords, users and groups, and other criteria. For example, organizations can use the Automatic Classification Engine to tag email messages that contain Social Security Numbers and then use Discovery Accelerator to search based upon this tag.

Application administrators can extend the search configuration page to include additional fields for including or excluding messages with particular policies. The System Configuration area of the application administration interface contains a category called Policy Integration that enables or disables these advanced search fields.
Policies can be specified with a single-selection drop-down box or by typing multiple policy names in a text box. When multiple policies (or tags) are specified on the same line, they are joined during the search using the logical OR operator. Criteria specified in multiple search fields are joined by the AND operator.

Figure 9 illustrates the additional search fields that can be displayed and used to include or exclude messages with policy tags assigned by the Automatic Classification Engine. In this example, the search results will be limited to only the archived items that were automatically assigned the tags Sexism or Racism.

Review

After a search is performed and potentially relevant items are consolidated, selected individuals (internal or external to the organization) typically review the search results to determine relevance and privilege. This review phase is sometimes outsourced to outside legal counsel and thus skipped entirely by the internal organization. In those scenarios, organizations typically create cases, run one or more searches, export the results from Discovery Accelerator, and proceed to the export procedure, as discussed later in the section titled, “Running the export or production” on page 44.

Organizations that perform a significant amount of review during e-discovery are subject to two requirements. First, they must divide large sets of content among multiple reviewers in a way that avoids duplication of effort. For example, if there are 10,000 items to be reviewed for relevance and 10 reviewers available, an organization might assign each reviewer 1,000 items. In this case, each reviewer, by default, would see a subset of the total review set when they access the application and display the items needing review.
Second, during this process, reviewers need some method to annotate or mark the reviewed messages with a status flag, but in a way that does not modify or tamper with the original content. This is very important when the review is conducted over several days or weeks, especially if the information must be exported according to its dispensation or confidentiality. By marking records and producing only a subset of information, costs can be significantly reduced, and confidential information can be retained inside the organization. Note that this review process can be of great benefit even if the organization plans to export the data and have outside counsel perform an additional review.

**Review assignment**

To assign search results to different reviewers, a case administrator can click a link called Review Assignment to display the number of items needing review and to specify the number of items each reviewer should be assigned. Figure 10 illustrates an example in which there are three different reviewers with a different number of items assigned to them.
Note that items can be moved from one reviewer to another, for example, when a reviewer is on unplanned sick leave. Simply enter a negative number in the Assign box for the reviewer who cannot perform the review, and enter a positive number for the substitute reviewer. Click Apply. The numbers are recalculated and displayed accordingly. The new amount entered in the Assign box is always added to or subtracted from the number of items previously assigned to each reviewer.

When reviewers enter the Review feature of Discovery Accelerator, they can choose to display and mark only their assigned items or all items in the entire case, as shown in Figure 11.

![Figure 11. Display Items screen](image-url)
Setting up default and case-specific marking schemes

It is important to have a well-planned marking scheme so that reviewers can assign dispensation to items based on relevance and privilege. The marking scheme may vary by case and by the reviewer’s role. Each Discovery Accelerator customer database has a default marking scheme that defines what marks are available to the reviewers and which roles can access each specific mark. From the home page, authorized users with the application permission to manage marks can create or edit marks available throughout the customer database. An unlimited number of marks can be created and made available for use.

Three levels of categories can be assigned to items: status, mark, and extra mark. Status is a basic indicator of what should be done with the item. There are three statuses, which usually indicate that an item will be included and associated with the case, excluded from the case, or evaluated later. When users create a new mark and give it a unique name, they must select one of the three statuses to apply to the items that are assigned that mark. For example, you could exclude messages from a discovery request for a variety of reasons such as privilege, nonrelevance, and spam, but each of those marks would share the same exclusionary status flag.

Extra marks are covered in the “Dispensation marks versus extra marks” section.

Once marks are created, administrators with the application permission to manage scheme templates can create or modify a default marking scheme that enables specific marks for use and specifies which marks each user role can see and use. At the case level, the case administrator can leave the default marking scheme as it was inherited from the application level or modify it to include different marks and/or different roles that can access them.

Figure 12 displays the screen for selecting which marks should be included in the marking scheme. The box on the left shows the marks that are available but that have not been included in this marking scheme. The box on the right shows the four marks that are included in this scheme.
Figure 12 illustrates the screen for selecting which marks a particular user role has the ability to see and use. Note that in this example, the Reviewer role cannot designate an item as Priv (or Privileged).
Dispensation marks versus extra marks

The main purpose of marks is to assign categories that indicate relevance and privilege. Therefore, reviewers typically have a short list of marks from which to choose. There are situations, however, that might warrant having additional marks or categories that can be assigned after the primary mark has been selected. For example, a general mark for an item may be “Relevant,” but an extra mark on the same item may be “Requires follow-up,” “Smoking gun,” or “Hot.” These extra marks are defined within the case settings as an additional scheme. As they can with primary marking schemes, case administrators can choose which marks are available in the extra scheme and which roles may use which extra marks. As described in the next section, an icon in the review interface displays a list of extra marks that the user is permitted to use. Note that the selection box is dimmed until a primary (dispensation) mark has been selected on that message.

The review interface

Users who have been granted Review case permission on at least one case can see a column on the home page called Review Messages, with a link to all cases to which they have been granted permission. When the user clicks the link with the case name, a page is displayed that allows the reviewer to filter which messages will be displayed once the review interface is opened. The user can make selections such as content type (Exchange, Domino, file, and so forth), current mark, and search name. There is also a small icon on the home page next to each case name that places the reviewer into the review interface using the default content filter. Figure 14 shows an example review interface.
The interface is divided into four main sections. Along the top are several icons that provide on the highlighted item, viewing legal holds that are placed on the highlighted item, and accessing the extra marks selection screen.

The middle section of the review interface is divided into two panes. The left pane lists a configurable number of search results with various metadata about the content in the review queue, such as the sender (if it is an email message), subject or file name, date, current primary mark for the item, and current extra mark for the item. There are also checkboxes next to each item for bulk marking, if permitted. The reviewer can sort this list by simply clicking the column title and can select which metadata columns are displayed by right-clicking any column title and checking/unchecking the options displayed in the context menu. The right pane displays an HTML rendering of the currently highlighted message. To save time for the reviewer, search terms are highlighted in yellow. Attachments are also displayed in the preview pane, and the reviewer can view the file in native application format or in an HTML rendering using the Web browser. When the HTML rendering is viewed, search terms are highlighted in yellow, as they are in the preview pane.
The bottom section of the review page provides buttons for the available primary marks (“Relevant,” “Query,” and “Not relevant” in Figure 14), navigational controls, and an area to type free-form comments related to the item currently displayed (which is therefore an unstructured third way to annotate an individual item in the case review set). Hot keys can be assigned to primary marks, and there are Next and Prior hot keys to speed up review.

For reviewers who prefer to see as much of the message rendering as possible on the review screen, a single-item view is available that eliminates the content list and uses the entire width of the browser window to display an HTML rendering of the currently selected item. Selecting a mark button at the bottom of the display automatically advances the display to the subsequent item. Figure 15 illustrates the review interface in single-item mode. Note that the top and bottom areas of the review interface remain unchanged compared with the example in Figure 14.

Figure 15. Review interface single-item view
De-duplication in the review interface

A particular item may be found multiple times through various searches that are run in a case. Discovery Accelerator, however, de-duplicates messages in the review interface according to the single-instance model in the underlying Enterprise Vault archive infrastructure. Enterprise Vault de-duplicates archived content from multiple sources within an archive storage device (referred to as a vault store partition), even if there were multiple copies of the content in its original location (such as multiple Exchange servers or PST files). The same principle holds true in Discovery Accelerator. Items stored in a single Enterprise Vault store partition are displayed only once in the review queue, regardless of how many users had a copy of the item or how many Discovery Accelerator searches found the same item. Refer to “De-duplicating and exporting data for outside counsel” to learn more about de-duplication of items during export and production.

Improving the speed of review with prefetch cache

The massive workload and tight deadlines inherent in e-discovery make speedy message rendering during review critical. Depending on the Enterprise Vault architecture and type of platform used to store the archived content, it may be desirable to implement a configurable amount of disk cache for items found in Discovery Accelerator searches. The cache can be as small as 1 megabyte or as large as 100 gigabytes, and the first in/first out method is used, as with most caching technology. The administrator schedules the process of building the cache from content in the Enterprise Vault. During review, the message rendering process first checks the cache for the requested item and retrieves it from the Enterprise Vault storage device if it is not in the cache. Organizations are advised to use very high-speed storage technology (SAN or direct-attached disk) for this cache to enhance performance.

Note that the option to implement a disk cache is at the organization’s database level. This may be another reason to create multiple customer databases, when you want only certain cases to use a cache. A cache would be unnecessary for databases that do not intend to utilize an internal review process. The cache is scheduled to fetch new search results from all cases within a database during certain hours, usually evening. A legal hold database, for example, would not be a good candidate for caching due to the large number of items and lack of review.

Application administrators can enable and configure this cache using the settings for Item Prefetch Cache in the System Configuration area of the application administration interface (see Figure 16). Remember that this cache is shared across all cases in the customer database. Advanced configuration settings for the cache are also available, such as encrypting the cache, maximum age of content in the cache, and maximum item size allowed in cache.
Improving the speed of review with keyboard shortcuts

For reviewers who prefer to use keyboard shortcuts rather than the mouse to mark messages, Discovery Accelerator provides the ability to establish an access key for individual marks. During the review process, the reviewer can press the Alt key along with the access key to assign the appropriate mark to the currently selected item. For example, to mark an item as Relevant, the reviewer might press Alt+R. Access keys are defined in the properties of each mark, during or after their initial creation.
Allowing external access from the Internet
Some organizations want to provide remote access to Discovery Accelerator from outside the internal network, for example, content access to outside legal counsel. This may help reduce the time and cost of printing search results on paper and paying outside counsel to review a large number of paper documents. Because Discovery Accelerator is a Web-based application, there is no client application to install or maintain on any workstation, internal or external. The main challenges from a technical perspective would be connectivity through the organization’s firewall and user authentication. The method to provide access to an internal Web server from the Internet is well known and relatively straightforward (like any public Web server hosted internally), and the hosting organization would need to create Active Directory user accounts and passwords for external user authentication. Discovery Accelerator or case administrators would then grant specific permissions to those special Active Directory user accounts. For example, the accounts may only be granted the permission to review existing search results and assign marks to each item. They do not necessarily need permissions to perform additional searches or to export content out of the application.

Analysis
Tiered review
Even though search results are often assigned to individual reviewers for initial review, many organizations require a tiered review process that allows an initial set of reviewers to cull content that is clearly not relevant to the case or investigation, such as junk mail and personal messages. The initial reviewers might use marks to flag messages as possibly relevant and possibly privileged, but let the legal experts review the reduced set of content to make the final determination of relevance and privilege.

For a tiered review process, all reviewers would be granted the case permissions to review and mark items in the particular case, but the second tier of reviewers would filter their review interface to display all messages in the case that already have a mark established, such as Rel-Relevant or PosRel-Possibly Relevant. They would not use the My Items filter because individual assignments are for unreviewed messages, not for previously marked messages. To conserve the interface space available for marks, use abbreviations for the mark names.

Once the second reviewer is in the review interface, the list of available marks to use will depend on the permissions for that reviewer’s role in the marking scheme. This list of marks could be completely different from that presented to the first reviewer. Regardless of which marks are available for use by the second reviewer the previous marks are displayed. Those marks can also be used as a filter to ensure that only relevant items are displayed.
Segregating privileged information

One of the main reasons to review the search results is to flag certain items as “Priv-attorney client privilege,” meaning there are proprietary or confidential elements to those items. In many legal situations, it is acceptable to exclude such items from the main production set as long as there is a full audit trail indicating how many items were set aside. There is always the chance that a judge or auditor may ask to see at least some of those records, but it is generally acceptable to remove them entirely from the first production set.

To provide this level of content protection in Discovery Accelerator, the authorized reviewer (often an attorney) marks the items as privileged, and then a case administrator with the permission to produce items “produces” the items with the privileged mark into a special storage location, which is kept separate from items that will be produced later, possibly by search or custodian. Once the privileged items are produced once, they cannot be produced again, and the designation mark is locked for that item in that case. For example, if someone later produces all results from a particular search, only the items that have not been previously produced (therefore all search results that were not marked as privileged) will be included in that secondary production run. In the analysis phase of e-discovery it is very important to flag items as privileged and to run the production of privileged items first to segregate those items before the general production run is executed for disclosure outside of the organization. This production run should use a separate prefix that helps designate the items as privileged. The production report contains the basic information needed to speed up the creation of a privilege log by counsel. For more details on production, see the “Production” section.

Applying and releasing legal hold

At any time, a case administrator with the permission to manage legal holds can set a flag in the properties of a case to apply or release a legal hold on all search results accepted into the case (whether they are reviewed and marked or not). One of the unique features of Enterprise Vault and Discovery Accelerator is that legal holds can be placed on the one existing copy of the relevant content already stored in the Enterprise Vault archive. Most legal hold solutions on the market today either store the relevant content redundantly in a separate storage location (on disk or indefinitely retained backup tapes) or suspend the disposition of the entire enterprise-wide archive, including content that is outside the scope of the case and thus the legal hold order. Discovery Accelerator targets the legal hold by adding search results to a case and applies the legal hold to the existing archive content, even managing multiple holds that are placed on a single item in the archive.
It is common for legal holds to be requested early in the e-discovery process—well before a formal discovery request is filed. To manage these types of cases, you can create a completely separate Discovery Accelerator customer database for the sole purpose of legal holds. This way, when an audit log, and possibly the entire SQL database, is requested to show due diligence and adherence to the legal hold order, there will be no trace of internal legal counsel’s “work product” in which counsel is running ad hoc searches and reviewing the data before the official discovery request is filed. Also, when it is time to produce information from targeted searches as specified in the negotiated discovery request, only that specific, targeted information will be produced—not the entire pool of content under legal hold. Legal holds are typically broad and far reaching, whereas formal discovery requests are often targeted to a specific date range, list of custodians/parties, and set of search terms.

Another reason to configure legal hold in a different database is to implement a different security model for the placement of legal holds and mining of content by attorneys. Various legal personnel, such as paralegals, may be tasked with running broad legal hold searches to ensure all data is retained in Enterprise Vault, but attorneys may run the targeted searches related to a discovery request. More important, very few people, if any, would be granted permission to review the content in the large and broad case review set in the legal hold case, but several people may be given permission to review the small, targeted case review set in the discovery case.

Keep in mind that having two different customer databases with the same case will not result in duplication of content storage because Discovery Accelerator is ultimately a set of organized pointers and metadata that relate back to the underlying (de-duplicated) Enterprise Vault content repository. Figure 17 illustrates the interface in which a privileged case administrator can apply or remove the legal hold status flag on a particular case in its properties page.
Figure 17. Placing a case on legal hold

Note that there is a comment box where the case administrator can specify a reason or details about the application of the legal hold status flag, and statistics are provided to indicate the number of items currently associated with the case (accepted search results), number of items on hold in the underlying Enterprise Vault archive, and number of errors in the legal hold process. Once the organization is released from the legal hold order, the case administrator can return to this case properties page and simply de-select the box next to “Put items on hold.” At that point, Enterprise Vault would release the hold established in the archive for those items and expire that content during the next scheduled expiry run, except for the items still under legal hold by other Discovery Accelerator cases.
Because the legal hold status affects performance during the acceptance of search results into the case (so it can notify Enterprise Vault to hold each individual item as part of the search acceptance process), an organization can apply legal hold on a case retroactively shortly after the searches are performed and the results are accepted in the case review set. The legal hold flag does not need to be in place during the search for items to be appropriately held—only before the next scheduled Enterprise Vault storage expiry process.

**Production**

The final step in e-discovery in which Discovery Accelerator plays a role is the production of items, generally in the original format from which they were archived. As stated previously, production might occur soon after the search is performed if the organization does not wish to cull the content by relevance and privilege before sending it out of the organization. Some example usage scenarios are described in the “Usage scenarios” section.

**Differences between export and production**

Discovery Accelerator provides two different ways to output content from Enterprise Vault: export and production. This can lead to confusion when an organization is trying to determine the most appropriate method. The main difference between export and production centers on whether this process should be an enforced final step in the handling of content in Discovery Accelerator or an opportunity to copy data while it is still being processed within the application.

The export facility enables items to be viewed outside of Discovery Accelerator before the review process is complete. Reviewers can continue to work on the items because the status is not locked, and an item can be exported as many times as necessary, even as marks are still changing. Note that items are stored on disk redundantly if they are exported multiple times, because each export is saved comprehensively in a separate folder.

The production process locks the status of the items so no further changes can be made to the status/mark. Additionally, once an item is produced once, it cannot be produced again, avoiding duplicate storage. This is helpful in terms of storage as well as the need to segregate privileged content.

Produced items are also assigned a unique (incrementing) Bates number so they can be tracked as individual items that are a part of a larger set of produced information. A produced item will be consistently referred to as an item number—51 out of 150, for example—whereas exports can contain a slightly different set and sequence of items each time they are run.

Exported items are numbered nonsequentially according to the item’s case-specific identification number that was assigned when the item was accepted as a new search result in the case review set. So, while productions have a logical numbering system, for example, from 1 to 1000, an
export might have files numbered 47, 64, 94, and so on. This numbering is helpful because exports are often performed to allow an outside party (such as legal counsel) to perform the review outside of Discovery Accelerator. If the organization wants to mark the items in Discovery Accelerator according to the results of the external review, an internal reviewer can navigate directly to a specific exported item by specifying its item ID number in the review interface.

Running the export or production
Discovery Accelerator users who are assigned the case-level permission called Production can export or produce information from the same link on the case administration page. A link called Production leads to the screen displayed in Figure 18.
As shown in Figure 18, the user can choose between export or production and specify the type of items that should be output. The output folder is configured on the case properties page and is used for both export and production. Note that in Figure 18, only instant messages will be exported because all other checkboxes were cleared by this user. Users can also choose to export only their assigned items, or all items from the case that match the other specified criteria, including those assigned to other users. If the organization is required to export data grouped by custodian (person), the organization can perform multiple searches according to custodian (with the person's name indicated in the search title) and running multiple exports according to those searches. Once the criteria are selected, the user clicks the Select Items button to display the number of items that match the specified criteria. The user can at that point refine the export or production criteria, which will affect the number of resulting items. Also, the user can limit the number of exported items by entering a specific number as shown in Figure 19.

![Figure 19. Restricting the number of items to export](image)

When items are produced, the numbering system will be incremented according to the items that may have already been produced. This helps ensure that items will be given a unique number even if there are multiple production runs. Alternatively, users can de-select the option to “Generate unique IDs” and specify their own starting sequence number for each production run, keeping in mind that this could cause multiple items to share the same number if not handled carefully.

Remember that an item can only be produced once within a case, regardless of the number of searches in which it was found. The item’s status and mark are locked, and the production number is logged within the database. An item can be exported many times, even if it has been previously produced. An exported item is tracked by the internal Discovery Accelerator tracking number.
Data formats for export and production

Discovery Accelerator supports a number of file formats for the export and production of content. All types of content can be output in their native application format or an HTML rendering, which are also stored in Enterprise Vault. The user performing the export or production can choose between the “Original type” or “HTML” as shown in Figure 20. If there is Exchange content and the original type is chosen, another radio button appears to specify whether the Exchange items should be written to individual MSG files or combined into PST files. If PST is selected, the user can specify a password for the file as well as a maximum number of megabytes for each file that is written. Multiple PST files will be written as necessary to accommodate all of the output items.

During the export or production process, the user can move to another page in the application or close the Web browser without affecting the process. A status page will be displayed in the application that indicates whether the process is still working and how many items have been output so far. When the process is finished, the application will report a status of Finished in the case’s production page. To increase the production speed for PST files, the system creates temporary MSG files and then recombines them into incremental PST files. It can take a while for a PST file...
to ingest all the MSG files after Discovery Accelerator has finished sending them. Always open the
folder and verify that there are no remaining MSG files if you are going to copy the production
folder immediately. After the production or export is finished, a new folder structure appears,
depending on the types of data output and formats chosen, as shown in Figure 21.

![Folder structure after export or production](image)

Figure 21. New folder structure after export or production

Note that there are folders for different content types, such as File System, Lotus (Domino),
and others not shown, such as Exchange, Instant Messages, Bloomberg, and more. Figure 21
shows that within the File System folder structure, the original folder hierarchy was re-created to
display the organizational structure in which the files were originally archived by Enterprise Vault.
The original file names are preserved, but additional metadata is added to the beginning and end
of each to indicate the ID number, modified date, and time stamp of the file from the time it was
archived to Enterprise Vault. There is a database-wide configuration setting that can turn off the
insertion of the date and time unless there are multiple versions of a file with the same name that
have been archived from the same location.
Improving production performance with prefetch cache

The rate at which messages can be produced or exported from Discovery Accelerator depends on the type and speed of Enterprise Vault storage device(s) from which you are producing, as well as the storage device to which you are producing or exporting. Any increase or decrease in the speed of either storage device will have a direct impact on production speed because most of the production process entails simply copying files from one storage location to another.

In the section “Improving the speed of review with prefetch cache,” item prefetch cache was described as an effective method to improve the performance of Discovery Accelerator review because the application first checks the cache location for the content before looking to the Enterprise Vault archive. In the same way, the speed of production and export is dramatically improved. Remember, however, that it is difficult to know precisely which items are currently located in the cache, so during targeted production runs some items may be copied from the cache and others from the Enterprise Vault store partition(s).

Symantec conservatively estimates that a well-tuned Discovery Accelerator server with item prefetch cache enabled will export, on average, 50,000 items per hour, but as described above, this varies according to the storage devices being used. Some organizations have reported rates as high as 80,000 items per hour in production, and during some internal lab tests Symantec engineers were able to achieve production rates as high as 150,000 items per hour (on a highly tuned system with high-speed direct-attached storage used for testing only).

Importing output into third-party applications

The vast majority of companies that use Discovery Accelerator export Exchange content in PST format because it requires handling fewer files, and the PST format is generally best considering the third-party case management applications into which content is sometimes imported. Several case management applications have historically encountered difficulty importing content as well as metadata from MSG files, but this technology appears to be advancing.

In the near future, direct integration of Discovery Accelerator with multiple market-leading case management applications will be announced. These integrations will make it possible to export or produce content directly from Discovery Accelerator into these third-party applications without temporarily exporting the content to a generic format such as MSG or PST. For now, however, PST files are the most common format used to transfer content between the two systems.
Chain of custody

Chain of custody is a frequently used phrase in law-related discussions. It refers to the integrity and handling of evidence and often is presented in the form of a comprehensive document or report. These documents generally contain topics including conditions under which evidence is gathered, identity of all evidence handlers, duration of evidence custody, security conditions while storing or handling the evidence, and manner in which evidence is transferred to subsequent custodians. While the scope of such documents goes beyond the role that Symantec’s software solutions play in content retention and e-discovery, Enterprise Vault and Discovery Accelerator contribute well-documented features that preserve the integrity of content, control access to it, and audit important activities.

Completeness of the archive

To determine the completeness of the archive, it is first necessary to identify how the content was originally collected into Enterprise Vault and how exhaustive was the capturing process. For example, was Exchange Server based journaling enabled to capture all messages sent or received? If so, during which time period was this feature enabled in Enterprise Vault? Also, did Enterprise Vault perform an exhaustive scan of all computers and servers in the internal network infrastructure to import all PST files into the archive? If so, when was this performed and which computers, files, and users were involved? Was the Enterprise Vault unique “safety copy” mechanism in place to help ensure that content was not deleted from the source application until the archive storage was successfully backed up?

Questions like these can be easily answered through the built-in Enterprise Vault auditing feature. This feature is configurable to log the necessary amount of detail as directed by the organization. The audit log is stored in an access-controlled SQL database called EnterpriseVaultAudit and is most frequently accessed when a chain of custody document is being prepared.

Integrity of the archive

It is also critical to ensure that the archive system preserves not only the content, but also the metadata surrounding the content, such as original location (down to the folder level), archiving date, date and time of entry into the archive, whether the email message had a “read” or “unread” status at the time it was archived, and whether the email message had been flagged by the user (or original custodian) in Microsoft Outlook for importance, reminder alarms, and so on. All of these attributes are preserved and access controlled in the Enterprise Vault archive.
In addition to the standard access control provided by Microsoft’s Active Directory security model, some organizations are required or wish to implement tamper-proof storage devices known as write-once, read-many (WORM) devices. Enterprise Vault supports a number of WORM storage devices, including EMC Centera® Compliance Edition, Network Appliance SnapLock®, IBM® DR550, and many others.

Finally, when content will be retained for an extended period of time, it is important to provide a method to display it in a generic, “future-proof” fashion without the need for the original application in which the content was created and saved. As described in the “Review” section, HTML content is stored in addition to the native content format in Enterprise Vault, and the HTML rendition is the default format in which Discovery Accelerator users display and review the content in the case review set.

Audit trail: Archive and application
In addition to auditing to determine the completeness of the archive, some organizations wish to audit the changes to archiving and retention policies, the archiving of each record, each end-user retrieval of the archived content, and the disposition of archived content that was not under legal hold at the time. The Enterprise Vault auditing feature can record events like these in an access-controlled database.

Within Discovery Accelerator, all activities are automatically audited and logged in the Discovery Accelerator customer database(s). In this database, authorized personnel can retrieve records relating to each step outlined in this paper, from case creation to production. Some of the most important activities to include in chain of custody documentation are:

- Preservation of legal holds
- Search terms used in response to formal discovery requests
- Record of who reviewed and marked each item (including a chronological log of multiple reviewers and marks over time)
- Record of the number of search results set aside and not produced (usually presented in a privilege log)
- Record indicating which items were produced or exported

The auditing features provided by Enterprise Vault and Discovery Accelerator allow the custodian of record to authenticate chain of custody documents or privilege logs using built-in tracking mechanisms. If a particular item is scrutinized during a hearing, the export ID number can be traced back through the entire time that it was under the control of the archive and discovery system, without requiring the manual documentation that most organizations without Enterprise Vault and Discovery Accelerator produce now.
Discovery Accelerator creates authentication reports in delimited text, HTML (with links to the files) and XML format. The fields and format of these reports can be modified. The text report contains the original Enterprise Vault item (saveset) ID, archive, and original location of the item.

Usage scenarios
Throughout this paper, it has been assumed that the organization is facing a large legal hold and discovery request and that it intends to perform an internal review of search results before producing them out of the application. There are also many other scenarios in which Enterprise Vault and Discovery Accelerator can help. The following three sections describe some common scenarios and explain the various ways that the software would be used.

Investigative search: Finding a needle in a haystack
Organizations often need to perform small-scale, targeted internal investigations for matters such as HR violations, equal opportunity employment complaints, and intellectual property leaks. In such circumstances, there may not be a legal hold order or external discovery request. Information may not even need to be exported out of the application. Because any sensitive investigation may turn up information that can result in legal action, Discovery Accelerator helps ensure that a complete history and audit trail are preserved.

Using Discovery Accelerator, an authorized individual creates a new case in a customer database that is completely separate from external cases. This prevents the commingling of information related to internal affairs and external requests.

Marks and review assignment are often unnecessary in this scenario, as typically a corporate security officer (or designee) will create the targeted search and review the small number of results, often looking for one “smoking gun.” The internal investigator typically knows the date range of the alleged incident, employees involved, and keywords to find the content quickly.

The goal in this scenario is to determine facts so that appropriate disciplinary action can be taken internally in a discreet fashion. An important requirement is access control for the discovered content. The corporate security officer wants to designate who will have access to the information that is found. For example, the case manager might open the item in native format (such as Outlook) and forward it to others who need to see it. Alternatively, case-level review permission might be granted so that a specific individual can open the review interface for the case and navigate directly to the item marked as Relevant.
De-duplicating and exporting data for outside counsel

Organizations with relatively small (or very busy) in-house legal departments often outsource as many phases of e-discovery as possible. The primary objective for Discovery Accelerator in this scenario is to gather a copy of the requested content and deliver it to outside counsel in a format that can be imported into their case management application (or their own implementation of Discovery Accelerator). Even though this organization does not take advantage of every feature of Discovery Accelerator, it benefits from the software in the following key areas:

- Simplified and comprehensive collection of email and file content without asking end users to manually copy their relevant data to a central storage location
- De-duplication of redundant content gathered from individuals and storage locations
- Reduction of time and cost associated with the use of paper-based records and review

In this scenario, an authorized individual creates a case in a customer database separate from any internal investigations and grants access permission to only a small number of individuals, if any, beyond the case owner. Marks and schemes are unnecessary in this scenario, and the majority of time in Discovery Accelerator is spent in the search interface. Even though the internal organization is not interested in marking the items for relevance and privilege (since outside counsel will handle that), it is still interested in reducing the volume of exported content because some outsourced legal costs are based on the volume of information to be reviewed. Therefore, the internal user performs a search, glances at the number of items returned, possibly peruses some of the item previews, rejects the search, refines the criteria, and runs the search again until a reasonable number of items are returned to export. The search criteria and number of hits can be used by counsel to negotiate with the requesting party to refine the initial request.

Once the search criteria and result set are approved or agreed upon, an authorized user then accepts the searches, produces the entire set of search results, and delivers this export to the outside counsel in a secure fashion.
Reviewing in-house for relevance and privilege

In this scenario, the organization has a substantial in-house legal department and has concluded that today’s e-discovery process is too time-consuming and/or costly. They would like to either conduct the review for relevance and privilege in-house or provide remote access through their network boundary so that outside counsel can access Discovery Accelerator directly. Many corporate legal departments have a review team that can also manage contract reviewers for larger cases.

The key requirement for these users is that the system be as fast and simple as possible because they are dealing with many records and potentially many reviewers. Features to improve performance, such as item prefetch cache and keyboard shortcuts, are critical.

This organization will still likely use a third-party case management tool to prepare the evidence for trial but will use Discovery Accelerator to cull content according to relevance and privilege. Using in-house or contract attorneys rather than outside counsel to conduct this review can provide dramatic cost savings.

More information

Administrator documentation

The Discovery Accelerator installation kit ships with a document titled “Installing and Configuring Discovery Accelerator.pdf.” This 72-page document details the system preparation and installation process and also includes a “Getting Started” chapter to provide the administrator with a general understanding of how the application can be configured and used. There is also a chapter on “Troubleshooting.” This document would be the first source for technical professionals to get more information on Discovery Accelerator.

End-user documentation

The Discovery Accelerator installation kit also includes a separate folder called “Reviewer” that contains multiple documents tailored for end users who will perform searches, review items, export data, and so on. This information is provided in Microsoft Word format so it can be customized as appropriate for users with varying roles in using Discovery Accelerator.
Discovery Accelerator release notes

The Discovery Accelerator installation kit ships with a document titled “DAReadMeFirst.htm,” which includes important tips, a description of the newest features, known issues, guidelines for improving performance, and an index of official product documentation.

Customer training course

Symantec Education Services offers a one-day Discovery Accelerator training course that focuses on application use rather than installation. For more information on Symantec Education Services course offerings, please visit www.symantec.com/enterprise/training.

The information contained in this document is provided by Symantec for informational purposes only and is not legal advice. The information may or may not reflect current legal developments. This information is not intended to constitute legal advice or to substitute for obtaining legal advice from an attorney licensed in your region, and is specifically not intended to provide advice or counsel on compliance recommendations for Sarbanes-Oxley or other regulatory requirements.

NO WARRANTY. The information contained in this document is being delivered to you AS-IS, and Symantec Corporation makes no warranty as to its accuracy or use. Any use of the information contained herein is at the risk of the user. This document may include technical or other inaccuracies or typographical errors. Symantec reserves the right to make changes without prior notice.
About Symantec
Symantec is a global leader in infrastructure software, enabling businesses and consumers to have confidence in a connected world. The company helps customers protect their infrastructure, information, and interactions by delivering software and services that address risks to security, availability, compliance, and performance. Headquartered in Cupertino, Calif., Symantec has operations in 40 countries. More information is available at www.symantec.com.