Chapter 12

Compliance, Records Management & eDiscovery

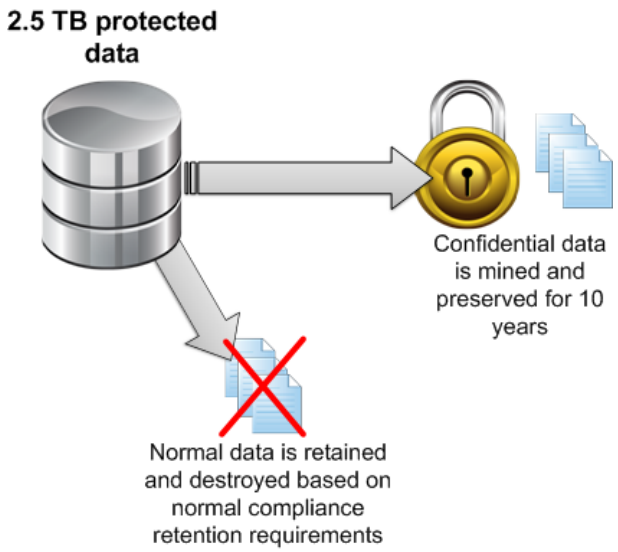
Compliance requirements in many organizations are forcing IT to readdress and redesign much of its infrastructure. Fortunately with Simpana software, compliance requirements can be more easily attained with the addition of a few key software features. Many compliance requirements can be addressed by transitioning from a data management concept to an information management concept.

Since Simpana manages all protected data within an environment it is a natural progression to expand capabilities into *Information Management*. Information Management is the intelligent management of data within an environment based on the content and ownership of the data. This is the difference between data management and information management. Data management addresses data in large pieces regardless of content such as a file share or email messages. Information Management digs into the files and messages and allows the information to be managed based on the specific needs of an organization.

When it comes to compliance regulations it is common to require special protection and preservation of specific information owned by groups or individuals. This information may contain keywords such as ‘confidential’, ‘top secret’, ‘patent’, or ‘classified’. Simpana software has the capability of drilling into data and extracting relevant information which can then be protected and preserved separate from other data.

Example: Approximately 2.5 TB of data exist in a file share. For security reasons certain documents marked as ‘Confidential’ must be maintained in a separate location where they will be preserved for 10 years. Without intelligent information management, all data would be required to be kept for 10 years. Using Simpana software all files marked as ‘Confidential’ could be mined and preserved for 10 years allowing non confidential data to be preserved based on normal compliance requirements.

The following diagram illustrates the use of intelligent information management to separate critical information from other data being protected. This allows the preservation of the critical information for long term retention while the rest of the data can be aged based on normal retention requirements.



Understanding Compliance and Information Management Concepts

This chapter will address several compliance methods that can be implemented using Simpana software. These concepts can be summarized as follows:

- Compliance
- Data Lifecycle Management (DLM)
- Records Management
- eDiscovery

Compliance

The term 'Compliance' can mean many different things. There are standard regulations such as SEC, SOX and HIIPA rules. There are also internal and industry specific compliance requirements many companies have to follow as well. It is quite difficult to design an IT infrastructure based on compliance regulations being that most of them are guidelines that do not spell out in black and white how information must be managed and preserved. However there are several Simpana features that can be used to meet various compliance requirements. Most of these features are covered in other parts of this book so this section will highlight relevant features that can be used to meet compliance requirements.

- **Retention Policies** configured within storage policy copies can be used to meet compliance requirements. Subclients associations to the policies can be used to protect specific data based on compliance requirements.

- **Encryption** is a critical aspect of compliance requirements, especially when sending data to off-site third party facilities. Inline encryption can be used when backing up over a WAN or into cloud storage. Offline encryption can be used when running auxiliary copy operations. Hardware based encryption for LTO4 / LTO5 drives can be configured for primary or secondary copies to tape media.

- **Wildcard Content for Subclients** can be used to define subclient content based on different data types. If Excel spreadsheets or PST files require special compliance requirements, separate subclients can be defined using wildcards *.XLSX or *.PST as the content.

- **Vault Tracker and Reporting** can be used to automate the process of exporting media, tracking media, and knowing when tapes must be returned and recycled. This is especially important when strict data destruction policies are required.

- **Secondary Selective Copies** can be used for point in time compliance copies of data such as quarter end financial backups.

- **Custom Business Calendars** can be used to align point in time backups (quarter, half year, yearly) with fiscal calendars.

- **Erase Data** option can be used to selectively mark specific data within data protection jobs as unrecoverable. This allows the jobs to be preserved based on standard retention rules while specific data that must be destroyed based on compliance requirements will be unrecoverable.

- **Simpana Security** can be used to delegate search rights for specific user groups. This can be an extremely effective method to provide delegation of power for compliance auditors when performing content searches since it can limit the scope of what the auditor can search for.

Data Lifecycle Management

For compliance reasons certain data may require specific retention and destruction requirements. Strict data destruction policies are sometimes defined to ensure data has a specific length of time in which it will be kept. Beyond that point it must be destroyed. Think of all the shredder trucks driving around. They are heading somewhere and they are intended to destroy physical information. Much of that physical information is also maintained digitally. Addressing the physical destruction of data but not addressing the destruction of the digital data does not provide a comprehensive solution for the data's lifecycle. Using Simpana features, digital destruction of data can be achieved.

Data Lifecycle Management (DLM) is the concept of managing digital data throughout its useful lifecycle and then destroying the data once it is no longer needed. Traditionally this has been done by end users deleting files and messages based on vague company policies. Using Simpana archiving and Storage Policies these procedures can be automated and taken out of the hands of the end user.

Using Simpana archiving agents, archiving policies can be defined to move data from production storage to CommVault protected storage when their relevance diminishes. The storage policy used to manage the data can be configured with specific retention policies that correspond to DLM compliance requirements. Based on these requirements, the destruction of the data will correspond to the retention defined in the policy. Since Simpana archiving agents move the data and manage it as well, the user would no longer be responsible for their own data destruction.

Records Management

Records Management is the intelligent management of information based on content and ownership. Information can be mined based on detailed search criteria and relevant information can be managed separately from non-relevant data.

Records management operates by first by running content index jobs on protected data. Then auditors can manually discover data or *Content Director* operations can be configured to automate this process. When information meeting the defined criteria is discovered it can have separate retention policies defined or it can be moved into a SharePoint repository for further processing and preservation.

Implementing Records Management strategies can ensure information is being properly protected and preserved. It can also reduce media storage requirements since non-relevant information can be pruned from storage based on normal retention requirements while relevant data can be preserved for long term compliance requirements.

eDiscovery

eDiscovery is the process or proactively or retroactively conducting content searches on information within an environment. This is most commonly used during litigation cases to discover relevant information for an investigation.

This works by first running content index jobs on data. A huge advantage of this is the ability to index jobs retroactively. As long as the job is retained in CommVault storage and the data is indexable, indexing jobs can be run and subsequent search operations can be conducted by legal teams. For cases involving past events historical data such as email messages can be indexed and searched. For ongoing investigations data can be indexed after it has been protected by CommVault. Online information for Windows file data, SharePoint documents and Exchange messages can also be indexed and searched.

Simpana® Components of Compliance & Information Management

In order to fully understand how the records management, eDiscovery, and DLM processes work, there are several Simpana components that must be understood.

Data Archiving

Archiving data is the process of moving information to less expensive secondary storage. This shrinks the size of primary storage which results in less space required to hold the data and smaller backup and recovery windows. Data Archiving is also referred to as *Hierarchal Storage Management* or HSM.

Archived data can still be retrieved through the use of stub files. Stub files contain information required to retrieve the data from CommVault protected storage. Depending on the OS or application being used, stub files may work differently but the general concept will be the same. When a user wants a file they open the stub file which initiates a recovery of the original file from secondary storage.

Stub files are not the only way to recover archived data. Files can be content indexed and users can search for files based on the contents of the file itself.

Example: A user is looking for a word document that is several years old. There is a specific section they want to use in a new document. The user does not remember the name of the file so they perform a content search. He does remember content from the file so they use what they remember as the basis of the search. Search results return the correct file, they recall it, and extract the information they need to create the new document.

Since the Simpana software moves the data into protected storage it will have the ability to preserve and destroy the data based on storage policy defined retention. This allows DLM policies to be established and automated within an organization. From an eDiscovery standpoint this can allow users to discover information which they might now be aware of that can assist them in performing their job more efficiently.

Compliance Archiving

Compliance archiving works with Microsoft Exchange and Lotus Notes Domino. The concept is the preservation of all messages that flow through the mail server. The messages are preserved in a special journal mailbox which can then be backed up by CommVault using compliance archiving agents. The information can then be content indexed for eDiscovery and Records Management purposes.

Using journaling and content indexing allows all relevant data to be discoverable. Traditional message protection methods may result in certain messages falling through the cracks. If end users delete messages before they are protected by CommVault, the message may be lost forever. Using compliance archiving and journaling ensures all messages are preserved.

Content Indexing

All object level data managed by CommVault that contains text information can be content indexed. Indexed information can then be searched through a web interface or automatic proactive searches can be conducted through Content Director Policies. For the web search component, there are two search methods, End User and Compliance Search. *End User Search* allows users to search for data that they have at a minimum read only access. *Compliance Search* allows auditors and legal teams to search all indexed information.

For a complete list of all data that CommVault can content index refer to Simpana Online Documentation.

Content Director

Content Director allows for the automation of search and preservation of information. Policies can be designed to move relevant information to CommVault legal hold, review sets, Enterprise Records Management systems, or data can be tagged for further review.

The biggest advantage of Content Director is with records management. Various policies can be created to proactively seek out information by keywords, data type, and ownership. From a compliance standpoint, this ensures records are preserved based on defined company policies.

Enterprise Records Management (ERM)

Enterprise Records Management connector allows information to be moved from CommVault protected storage to a SharePoint document repository. These processes can be automated using Content Director or relevant information can be moved into an ERM through the compliance search web interface.

Using the ERM connector is very useful when specific lifecycle policies are required. SharePoint has powerful workflow processes built into the product that allows the automation of records processing. Using the Simpana software to discover that data will greatly reduce discovery time and integrating with SharePoint allows discovery experts familiar with SharePoint to process information in an environment they are comfortable with.

EDRM XML Support

To provide a complete end-to-end eDiscovery solution, the Simpana software supports the EDRM standard XML output of search results. This output can then be ingested into advanced analysis tools to further analyze and produce relevant information regarding an investigation.

Data Lifecycle Management (DLM)

Data Lifecycle Management is a concept which allows organizations to intelligently manage data throughout its useful lifecycle. CommVault® software is one of the leading technologies to allow for the intelligent management of data. This is accomplished by logically addressing business data based on the relevance of the data to the business and the expected lifecycle of the data. DLM is implemented through Simpana archiving agents which will automatically move the data through different layers of an HSM model and then destroy the data based on defined data destruction policies.

Data Lifecycle or Information Lifecycle?

The term *Data Lifecycle Management* has been used throughout this book when discussing storage policies and data protection. This term is used because data being grouped into subclients can be managed independently within a backup environment. *Information Lifecycle Management (ILM)* however focuses more specifically on the type of data, content, and ownership of the data that is being managed rather than where the data resides. ILM policies can be applied in conjunction with DLM policies through the use of content indexing.

DLM Concepts

Traditionally files created by a user were small documents that would not take up much space on storage. As the capabilities of applications and the power of computers increased, users would create more elaborate documents embedding pictures, diagrams, and even video files into a document. Over time these files sitting in home folders became difficult to manage and backup times increased.

"Manual DLM" is the idea of forcing users to clean out old files they no longer need. Problems came about since many people would not delete files, and those who did would often ask for them to be recovered later on because "It's very important and I need it back now!"

Data Lifecycle Management automates the archiving of older data through the use of policies. These policies are determined by business needs. As files become older they are moved or *archived* to less expensive media. Archived data can still be accessible to the user either through stub files which point to the actual data, by conducting index searches, or both. Simpana administrators can also recover archived data by using the *Find* or *Browse* features.

Service Level Agreement

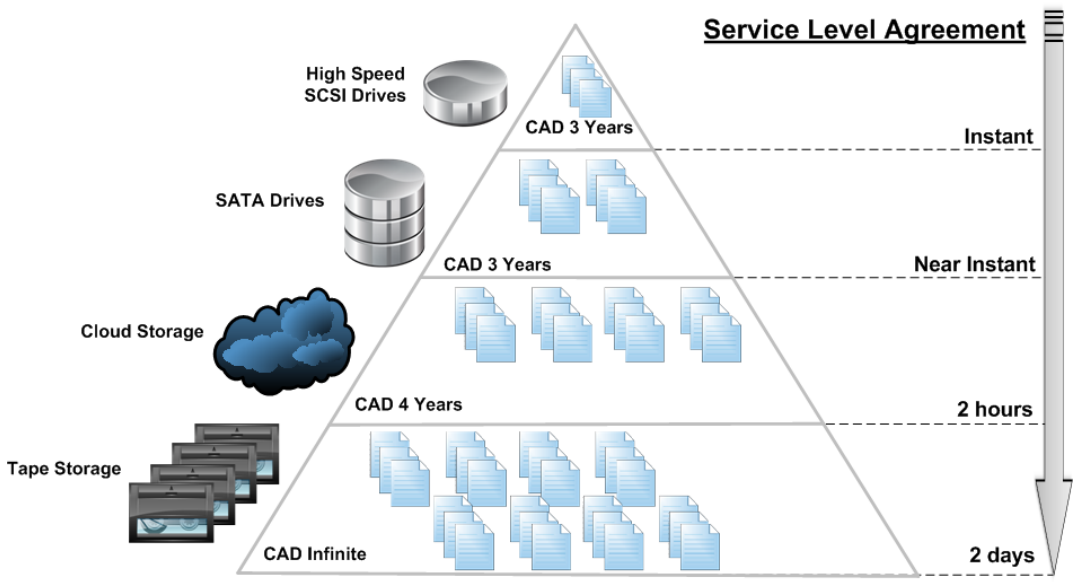
As data becomes older the chances of it being needed diminishes. A *Service Level Agreement* or SLA can be defined stating that the older the data is the longer it will take to recover. Traditional data archiving would define policies to archive all data within specified folders. The idea of DLM is to set SLA's for different data owned by different people within an organization.

Example: The engineering department in a company generates large CAD files containing schematics for various projects. Once the projects are over the CAD files must be kept for legal reasons. Five years later if something goes wrong with the project, engineers must have the ability to go back to old diagrams to find out what went wrong. Through business meetings with the engineering department it is determined that the probability of CAD files being accessed drops after three years. However, there have been emergency cases within a period of six years that CAD files had to be accessed. All CAD files must be kept forever. Being that CAD files are usually quite large it would be very expensive to maintain all CAD files forever in production disk storage.

The DLM Pyramid

DLM policies are usually diagrammed in a pyramid. The top level of the pyramid represents production storage. It usually the most expensive storage and has least amount of capacity due to cost. Subsequent layers represent the Hierarchal Storage Management (HSM) of the data.

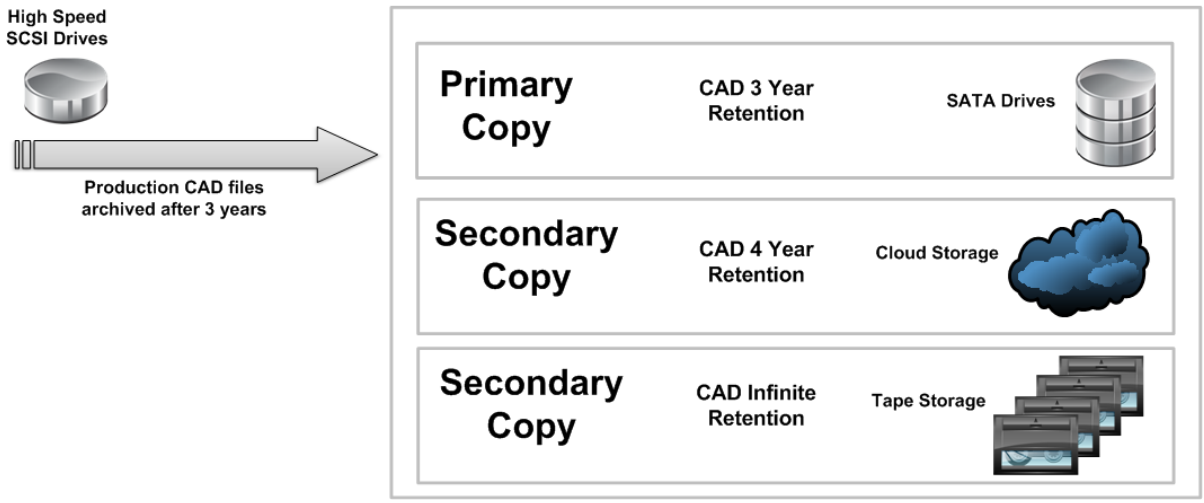
The following diagram illustrates the DLM model for the engineering CAD files. The top tier represents production storage and subsequent tiers represent the HSM layers. Note the use of different media to meet on site and off site protection requirements.



Storage Policies and DLM

DLM policies are implemented with the Simpana software through storage policies. Each layer in the HSM pyramid is a storage policy copy. The first HSM layer is the primary copy and subsequent layers are secondary copies.

The following diagram shows data being protected based on the DLM model. In this case the CAD files will be stored in production storage for three years. Beyond that they will be archived into protected storage based on DLM policies for the CAD data.



Using DLM concepts and CommVault storage policies, elaborate SLA designs can be implemented. Multiple copies within each policy can be created representing different SLA's for business data. Custom subclients can be defined for different data locations and data types.

Since storage policies create multiple copies of data, as the data ages the time to recover (SLA) may increase. Recovering from disk will always be fast for the user. Data can also be copied to tape. An emerging method for data archiving is the use of cloud storage. This actually serves two purposes. Moving data to the cloud for long term archiving allows storage to be allocated on demand through third party vendors. It also provides a level of disaster recovery since the data is being moved off-site. A tape tier is also used to preserve data. The advantage to tape is that it can scale infinitely. It also provides a level of disaster recovery by locating the data off site. An important note is that though cloud also gets the data off site, being that it is managed by third party vendors and requires WAN connectivity, having the tape tier is an important safety net.

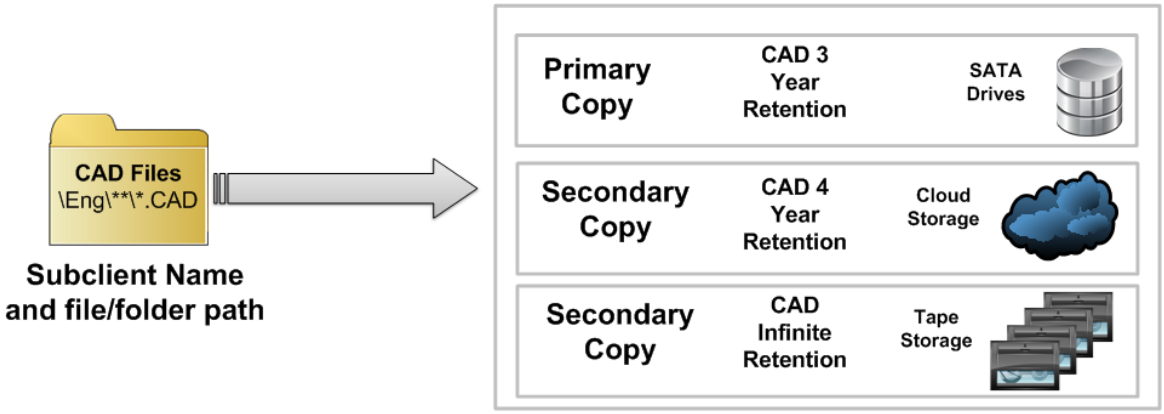
Defining Subclient Content for DLM

Subclient content can be predefined through a traditional archive set and subclients, the turbo agent, or on-demand archive sets. Traditional archive agent allows for data to be defined within the subclient contents and rules can be set to determine what data is archived. The turbo agent integrates backup and archive into the same process resulting in a single pass backup and archive job. On Demand Archive Sets can use scripts to dynamically define content to be archived.

Defining Wildcard Subclient Content

With the introduction of Simpana v9 SP3, content can now be defined using wildcards. This will allow data to be defined within a subclient based on file extension. This can be used in an DLM based strategy to move specific data throughout its useful lifecycle.

In the following diagram, a file share on the network contains engineering files. These files consist of DOC, DOCX, PDF, and CAD files. We will use the File System Archive agent to archive data. We will create a subclient to the share location and define \Eng.CAD as the data path and file type.***



Data Destruction Policies with DLM

By defining DLM policies and implementing them through archiving agents and storage policies, data can be automatically moved to secondary tiered storage and then destroyed when its expected life expires. Expanding on the concept of DLM, Records Management, which will be discussed next, can incorporate content indexing to mine data more granularly and extract relevant data. The data mined can be associated with a storage policy just like traditional DLM strategies and managed throughout its useful lifecycle. The main difference is where DLM uses traditional archiving methods of folder and file type, the concept of records management allows data to be mined based on content and ownership.

Data Classification v8 and v9

In Simpana v8 the Data Classification Engine (DCE) allowed archive rules to be defined not only on content but also ownership of the content. This required the DCE to do a complete scan and build a database that would be queried during an archiving operation. The issue with the complete scan method was that the construction of the database in large volumes with millions of objects would take to long to complete. In Simpana v9 the DCE has been reengineered to use the metadata of the volume being monitored to construct the DCE database. This greatly reduced the time to build the database but resulted in the v8 feature to archive content based on ownership nonfunctional in v9. The main purpose of the DCE in v9 is to expedite scan processes during backup and archive operations.

The ability to use wildcards to define content has in part made up for this. However archiving content based on ownership cannot be done with the normal archiving agent. This is still possible though through the use of content indexing and search. The approach to information management is now greatly enhanced using this method. Now not only can information be searched and preserved based on data type and ownership, it can also be searched based on content within the data. This concept is referred to as *Records Management* and will be discussed next in this chapter.

Records Management

Records Management provides the ability to intelligently mine data based on content, data type, and ownership and manage that data through CommVault Storage Policies or Enterprise Records Management system such as SharePoint. These processes can be automated to proactively manage information as it is discovered within an environment.

Once data is backed up or archived it can be content indexed. Once the data has been indexed, searches can be conducted on the information. Information meeting the search criteria can be moved into a CommVault Legal Hold or moved to an ERM.

CommVault Legal Hold

Legal Hold Storage Policies can be configured with specific retention requirements to meet records management policies. Movement of the data can be accomplished in two ways:

1. Auditors can use a web search console to conduct content searches. Results of the searches can be moved into a legal hold. The selected objects will be copied from existing data protection jobs into a new job associated with the legal hold storage policy.

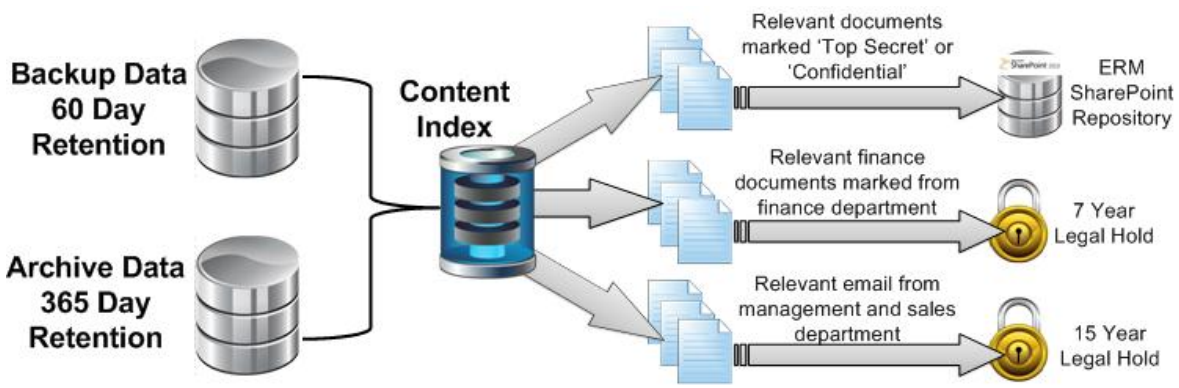
2. Content Director Policies can be created to automate the process of searching through content and automatically copying the data into the legal hold policy. These operations can be scheduled to run at regular intervals.

Enterprise Records Management (ERM)

ERM operations work the same as the legal hold processing except instead of moving discovered items into a legal hold policy they are exported through an ERM connector into a SharePoint repository.

Example: backup and archive data are being managed within CommVault protected storage. The data is content indexed and search operations are conducted to find the following: items marked ‘Confidential’, relevant items from the finance department, and email messages from the Sales and Management groups. Confidential documents are automatically moved to an ERM SharePoint document repository. Finance documents are moved to a seven year legal hold and email from Sales and Marketing groups are moved to a 15 year legal hold.

The following diagram shows backup and archive data being indexed and moved to legal hold storage policies based on compliance and legal requirements. Content searches are performed and relevant information is preserved based on retention and destruction requirements.



Automating Records Management with Content Director

Simpana software allows the use of a Compliance Search web interface where auditors can query content indexed data and manage the data based on compliance requirements. As new data is protected by CommVault, these operations must be conducted over and over again by the auditors. Content Director is a method to automate this process.

Content Director Policies are created in the CommCell Console under *Content Director* in the CommCell Browser. The Content Director Policy wizard can be used to define and schedule workflow operations to proactively mine and manage information. The following list describes some of the search criteria that can be specified for Content Director Policies:

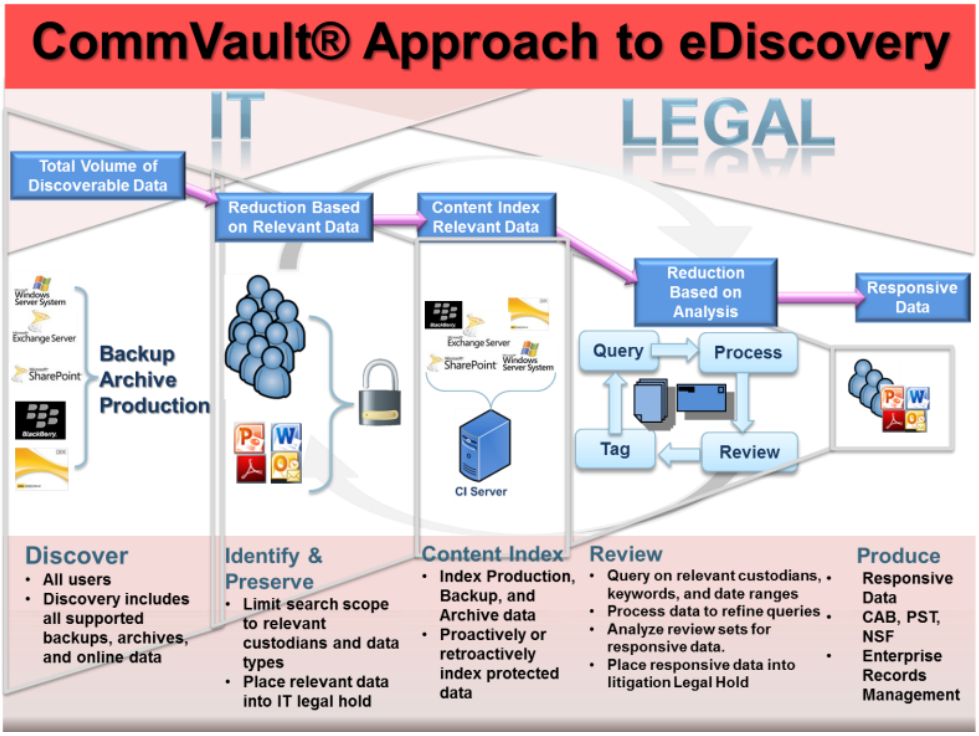
- Keywords within searchable items.
- File names and patterns within specified folders.
- Email messages can be searched by: subject, From, To, Cc. Bcc, attachment content, or Received Time.
- Lemmatization search which use variation of words within search criteria such as: run, ran, running.
- Synonym search which uses a synonym dictionary to find similar words.
- Entity search for custom defined content strings such as: SSN, Phone number, or credit card number.
- Search specific Client or Client groups.
- Search for data accessible or owned by Active Directory users and user groups.

Within the workflow relevant items defined by the search criteria can be moved into a CommVault legal hold, moved to a review set for further review, Moved to an ERM SharePoint repository, or tagged for further processing.

Once the workflow is complete the policy can be scheduled, run immediately or saved as a script. Defining search criteria and scheduling the policy to run at regular intervals allows proactive records management of data on a recurring basis.

eDiscovery

eDiscovery is the process of discovering specific relevant information typically required for an investigation. This process requires several steps to identify, preserve, review, and produce relevant information. The eDiscovery process is illustrated in the following diagram.

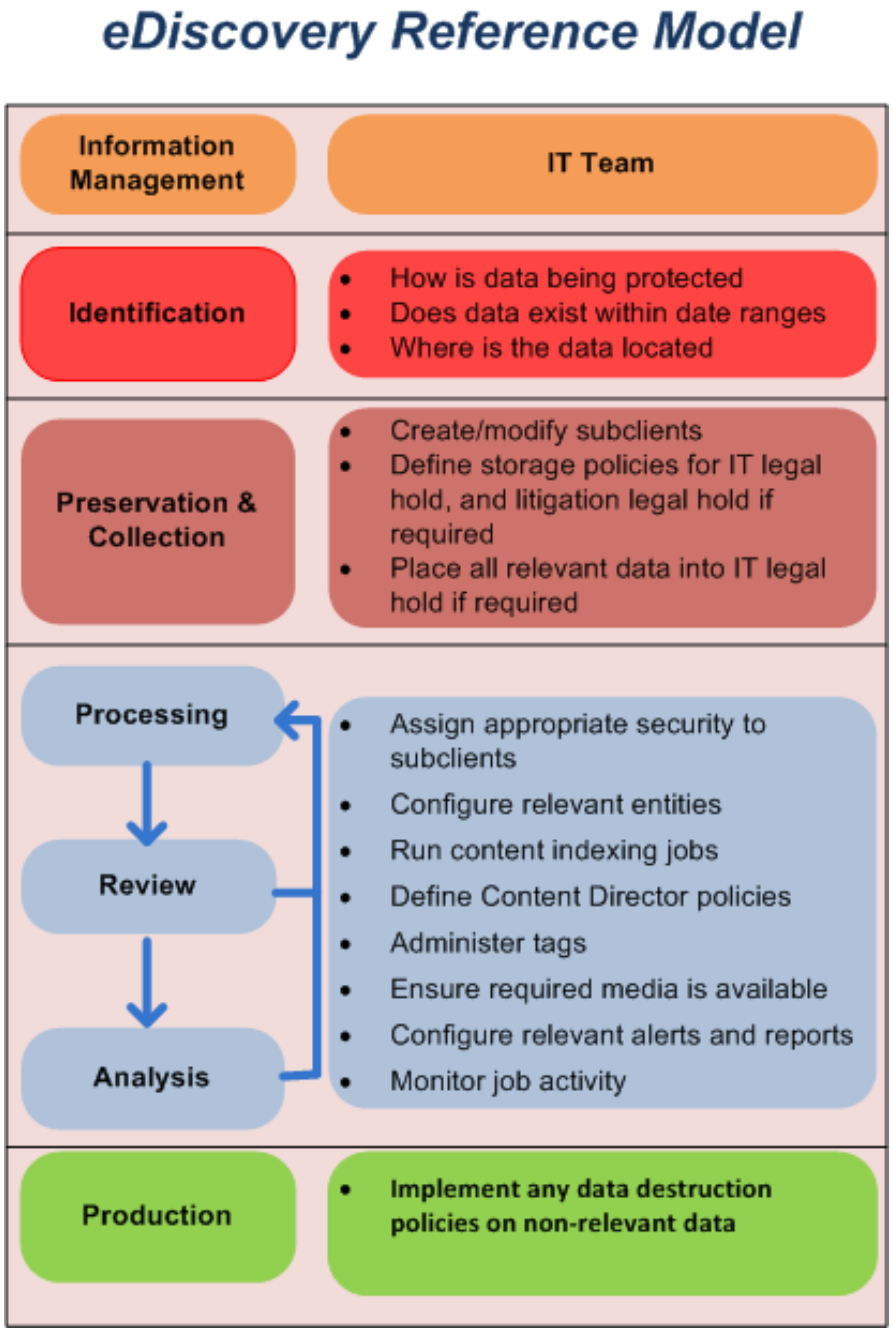


The eDiscovery process requires IT and legal to closely communicate to ensure a successful investigation is conducted. The beginning phases (discover, identify, and index) require CommVault administrators to work with legal to determine the scope of the investigation. Once relevant information has been gathered and processed then legal takes over digging into the information to review and produce responsive data for the investigation. This provides a separation of powers and removes IT from the legal processes (review, analyze, and produce).

The eDiscovery Process

The Simpana approach to eDiscovery is based on the eDiscovery Reference Model (EDRM). Within the EDRM workflow, both IT and legal will have specific responsibilities. The following diagrams illustrate both IT and legal team's responsibilities in the workflow process.

IT eDiscovery workflow



Identification

- Coordinate with legal team regarding custodians, search scope, and data types.
- IT must assess how the relevant information is currently being managed in the CommVault protected environment. All data within legal's defined scope needs to be included in CommVault's protected environment (archive or backup).

Preservation & Collection

- Coordinate with the legal team regarding length of investigation.
- An IT legal hold would be required if processing and analysis of relevant data is going to be potentially performed beyond the scope of standard retention policies.
- IT legal hold may require creating custom backup/archive sets. This would allow separate protection of data based on legal's retention requirements, while still allowing the data to also be protected with normal retention policies.
- Configure subclients to define all relevant data (if necessary) and direct them to a Content Indexing (CI) enabled storage policy.
- Configure a new or existing CI enabled storage policy to content index relevant subclient data.
- Configure a legal hold storage policy for use by the legal team.

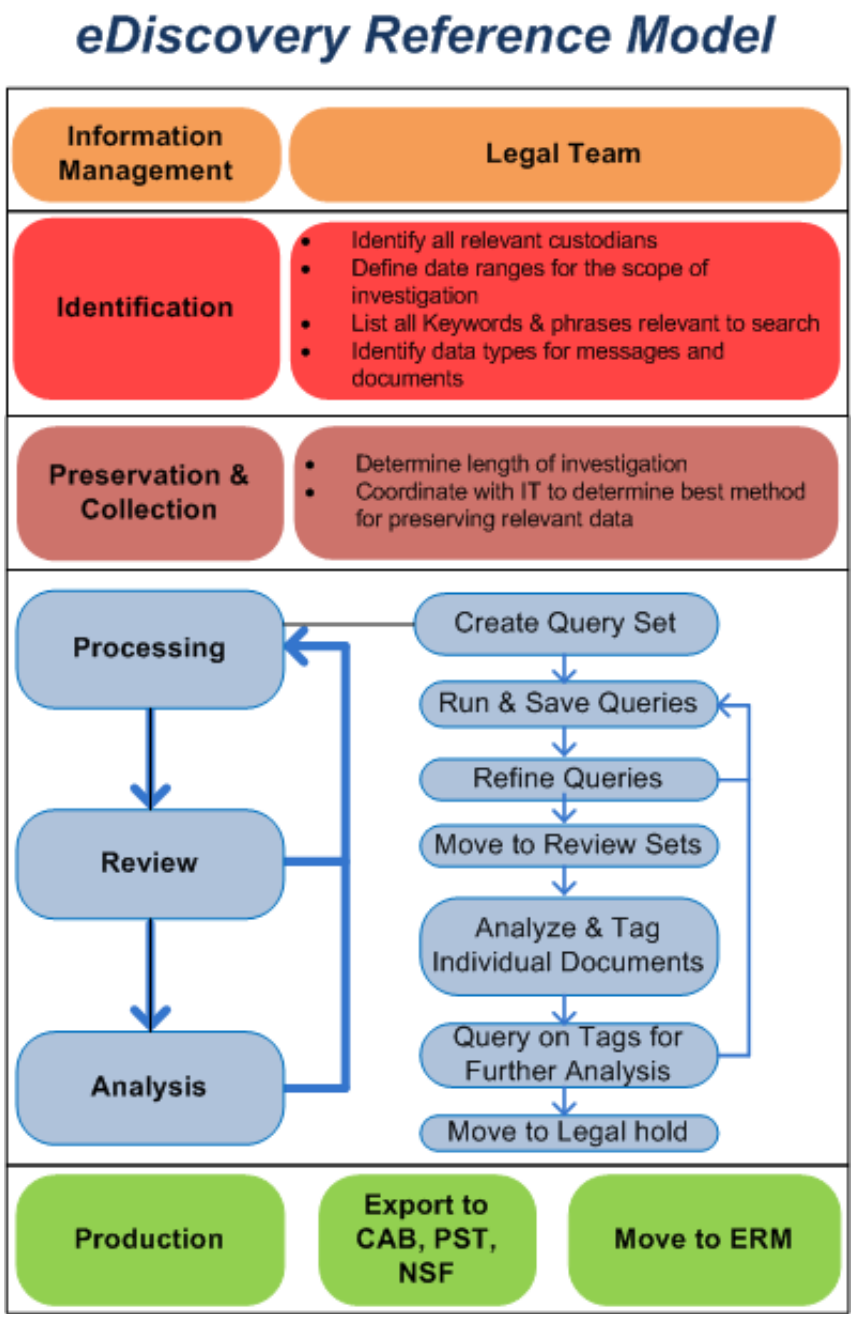
Processing, Review & Analysis

- Security can be defined to permit certain users to have rights to searching subclient data. Coordinate with the legal team to determine security requirements based on each member of the legal team.
- If legal will be using entities search (SSN, Phone, Zip, credit card), the storage policy needs to be configured to normalize these data types. **Note**: If entities are being used they must be configured prior to running content indexing jobs. If entities are enabled after CI jobs have been run the data must be re-indexed to capture entity settings.
- IT can define Content Director policies and schedule them to run which allows additional data being added to the investigation to be automatically processed. This can be extremely beneficial in ongoing investigations where custodians are being actively monitored or if additional data is discovered after the initial searches.
- Tags can be created by the legal team or a list can be provided to IT to create in the CommCell Console, but deleting tags must be done by IT in the CommCell Console.
- When jobs are submitted to legal hold or export, these jobs may take some time to run and there is the possibility of object failures if data cannot be retrieved from backups and archives. It is essential to monitor jobs, configure alerts, and reports. Determine what type of reports and alerts are required and which legal team members should receive them.

Production

- Once all relevant data has been processed it is important for IT to remove any IT legal holds to ensure data protection requirements are complying with standard data retention and destruction policies.

Legal Workflow



Identification

- Legal must define the scope of the search including date ranges, relevant custodians, types of data required (Email, document types), and relevant search words and phrases.
- Present this information to IT so they can begin preparing data for collection, processing, and analysis.

Preservation & Collection

- Legal teams must determine the length of the investigation.
- Provide IT with the length of time data must be preserved so they can assess current data retention and destruction policies and determine whether an IT legal hold will be required.
- If the data is going to be processed and analyzed within the currently defined data retention policies then the legal team can perform any legal holds if required. Coordinate with IT so they can define legal hold policies within the CommVault® software that will be used by the legal team.
- If the length of the investigation is going to be potentially beyond the scope of standard retention policies, IT will have to place data into IT legal hold.

Processing, Review & Analysis

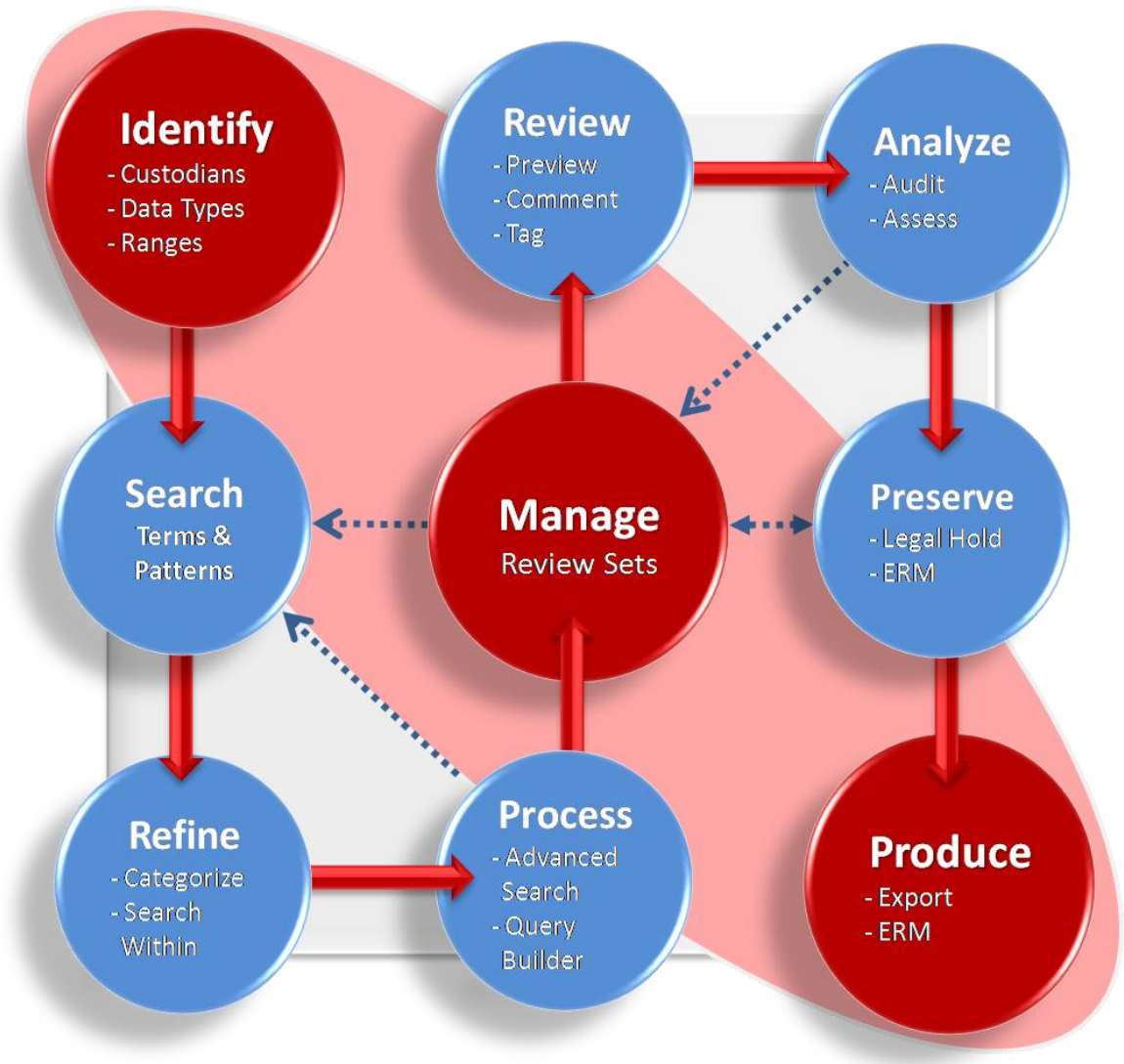
- Strategies should be developed on how to best craft queries. Many queries may be required so a query set for the project should be created and all relevant queries should be saved to the set.
- Refinements to queries should be made to eliminate non-relevant data. Use the advanced search options to exclude non relevant messages, date ranges, file types, and keywords. Strong knowledge of search and query language should be obtained through CommVault training and professional services. Refining queries will make review sets much easier to process. If the scope is determined to be too narrow, additional queries and relevant data can be added to review sets at any point in the review process.
- When data is moved to a review set, ensure as much non-relevant data has been excluded from queries. The purpose of the review set is to process documents and messages individually and tag items for relevance and follow up. Non-relevant data within the review set can be deleted if necessary. Permissions can be granted by the owner of the review set to users to permit or deny the ability to delete data.
- Additional queries can then be used to query based on tags. At this point additional review sets may be defined to hold relevant tagged data. This process also can be used to move data through the companies workflow process and share the review sets only with necessary legal team members.
- Once all relevant data has been processed it should be moved to legal hold. Although the data can be exported from the review set, it is more efficient to first move to a legal hold. CommVault software will locate all documents and messages and consolidate them into a separate backup which will make exporting the data considerably faster.
- When jobs are submitted to legal hold or export, the jobs may take some time to complete and there is the possibility of object failures if data can not be retrieved from backups and archives. Coordinate with IT so they can set up all required alerts and reports for the legal team to receive.

Production

- Data can be exported to CAB files, PST (Exchange), or NSF (Lotus Notes).
- Data can be exported to the Enterprise Records Management Center (ERM) for further processing and analysis or longer term preservation.
- Once all relevant data has been processed coordinate with IT so they can remove any IT legal holds to comply with defined data retention and destruction policies.

eDiscovery Search Components

The compliance search web interface contains several different components which will assist in the search process. The search process is illustrated in the following diagram.



Query Sets

Complex queries can be crafted, shared and saved. Multiple queries can be used to ensure the full scope of the investigation is being achieved. Relevant information can be moved to review sets, legal holds, ERM, or export sets.

Review Sets

When information is moved to a review set each object is typically reviewed in greater detail. At this point the items can be tagged, filtered, or comments can be made. Multiple review sets can be used to further process information. Items in review sets can be moved to legal hold, export sets, or ERM.

Tagging

Items can be tagged for follow up. Additional queries can be run on specifically tagged items for detailed review. Along with the built in tags, custom tags can be created by the reviewer.

Legal hold

When the legal team places items into legal hold they will be able to associate the items with a selected legal hold policy. The legal hold policy is actually a storage policies designated as a legal hold policy. When items are moved into a legal hold they will be retrieved from previous backup or archive jobs and a new job will write them to the legal hold policy. This means data is physically moved in CommVault protected storage which requires the media to be in libraries for a legal hold operation to complete successfully.

Export

Members of the legal team can also add items to an export set. The export can be in CAB (compressed files), PST (Exchange), or NSF (Lotus Notes) format. This is an operation that will retrieve the items from CommVault protected storage and save them in the export file. The exported file can then be downloaded and managed independently of the CommVault environment.

Security

Delegation of Search Capabilities

- IT can separate data and assign content search capabilities to specific legal teams. This ensures proper security and can prevent certain legal members from searching specific content.
- These security measures must be established during the Identification phase and coordinated with IT to establish search limitations, if any.

Initial Queries and Review Sets

- Legal teams with content search capabilities run queries on data and refine queries to include all relevant data. This process may require multiple queries and extensive query refinement.
- All queries relevant to search can be saved in a query set. Each investigation can have its own query sets.
- Team members running queries should be well trained and understand all query search options.

- Query sets can be shared with other legal team members. Permissions for sharing include: Add/Append, Delete, Execute, and View.
- Initial review sets are created and relevant data is saved to the queries for further review.

Delegation of Review Sets to Legal Teams

- Legal team creates review sets and Shares the set to legal team members to analyze and process data.
- The owner of the Review Set has capabilities to share review sets.
- Permissions for Sharing the review set include: Add/Append, Delete, Retrieve/Download, and View.
- This process can limit the scope of working with review sets to specific team members for tighter security.

Working with Review Sets

- Legal team members can review data for relevance and tag the data for later analysis and processing.
- Multiple review sets can be created for each investigation.

Query on Tags & Attorney Review Sets

- Once legal teams have tagged all relevant data a query can be run on the tagged items and placed into a new review set.
- The review set can be shared to specific legal team members and attorneys for final analysis and processing.

Legal Hold

- Once all relevant data has been analyzed and processed a legal hold can be placed on the data if necessary.
- Legal holds can be shared with other legal team members. Share permissions include: Add/Append, Retrieve/Download, and View.
- Data can be exported to present to opposing legal council or copied to an ERM for further processing and analysis.

Export

- Data can be exported directly from queries, review sets, or legal holds.
- Data can be exported as CAB, PST, NSF files.

Best Practices

- Identify custodians and level of confidentiality of data.
- Identify legal team members and the limit of their search capabilities.
- Coordinate during Identification phase with IT to limit search capabilities of legal team members if necessary.

Configuring Content Indexing

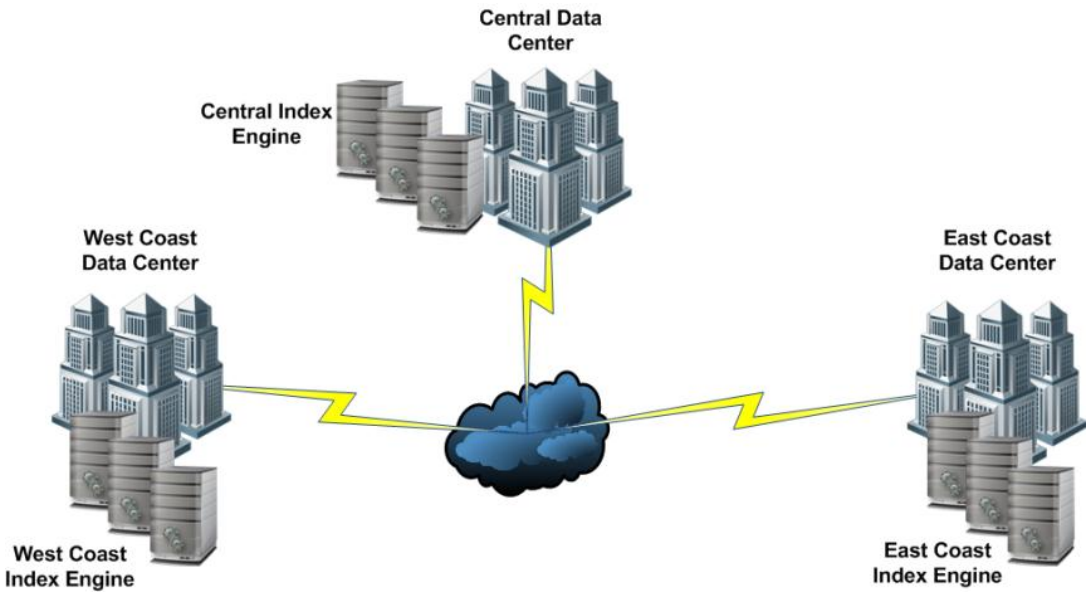
Note: This section is not intended as a guided step by step manual on installing and configuring content indexing. It is STRONGLY recommended consulting with Professional Services when planning and deploying content indexing within an environment.

There are two content indexing methods: Online and Offline. Online indexing is capable of indexing production data for Windows file systems and Exchange mailboxes. Offline indexing can index all indexable objects within CommVault protected storage.

Indexing Engine

The central workhorse for indexing data is the Content Indexing Engine. One or engines can be deployed for a truly scalable enterprise solution. Each engine can contain up to nine nodes and each node can index up to 60 million objects. It is required that CommVault Professional Services be consulted when deploying content index engines. A properly planned and scaled environment is essential for effective indexing and search.

The following diagram shows multiple distributed indexing engines



Configuring Online Content Indexing

Simpana online content indexing is supported on Windows file systems, Exchange, and SharePoint. Though this will provide the ability to index online data, since CommVault is not managing the actual files they cannot be preserved or moved into legal hold policies. Off line content indexing for backup and archive data provides more flexibility in information management.

Configuring Offline Content Indexing

Offline content indexing is configured in the storage policy Properties in the **Content Indexing** tab. A content indexed enabled storage policy must first be associated with an indexing engine. Multiple engines can exist within a CommCell environment allowing great scalability for large environments. This also allows indexes to be distributed and kept local to an environment when multiple datacenters exist.

Associating Subclients

A main feature of offline content indexing is the ability to select what data needs to be indexed. Use the Subclient Association section to select which subclient data will be indexed

Filters

Filers can be defined to include or exclude specific file types for indexing. Filters also allows the filtering of files based on file size. The maximum default size that will be indexed is 50 MB. The most common file type that will be larger than this size is PST files which may be crucial to an investigation. For PST files it is recommended to use the Exchange archiving agent to ingest the PST files and index the messages through the archive job.

Retention

The default retention for content indexes is based on the longest retention setting defined within all the copies of the storage policy. This can be changed and a specific number of days can be designated to retain content index data. If the index time has expired on a job and indexes are deleted, the job can be re-picked for content indexing as long as it is still retained in CommVault protected storage.

Entities

Entities are custom content of data that is typically in a specific form within indexable items. Social Security numbers for example, will typically contain nn-nnn-nnnn format. This information can be normalized during the indexing process and specific entity searches can be conducted through the web search console or Content Director to search for the entity item. Custom entities can be defined. CommVault Professional Services can assist in defining custom entities.

Part III

Designing & Implementing Storage Policies