Chapter 6

Defining Data Retention & Destruction Policies
Retention of data in protected storage as well as production data is a critical component of an overall data management policy of an organization. Data retention policies and data destruction policies in some organizations are well defined and it can be a challenge for administrators to meet those requirements. In other organizations, policies are not well defined or not defined at all. In some cases, determining how long data should be held on falls squarely in the administrators' hands. This is not the way it should be, but the truth is that all too often it is. This chapter is designed to assist administrators in getting the proper people to define the policies for data retention and data destruction and to explain how to implement those policies with Simpana® software.

There are three primary reasons data is retained:

- Disaster recovery
- Compliance and archiving
- Data Recovery

**Disaster Recovery** policies should always be implemented based on how many complete set of backups should be kept. A set is referred to as a cycle and it will include all data protection jobs required to restore an entire system to a specific point in time. In this case, the number of cycles should be used to determine retention policies.

**Compliance and Archiving** copies are usually point in time copies of data that will be retained for long periods of time. Month, quarter, or year end point in time full backups are usually retained for months, years or indefinitely. In this case, the length of time is the key factor in determining retention, not the number of cycles.

**Data Recovery** is the ability to go backward in time to a certain point to recover specific data at that point. It could be a file a user deleted two months ago, or an Email message. Data recovery policies may include all data for a specific length of time so keeping full and incremental backups may be required. In this case, which copies and the length of time they will be kept will determine retention policies.

**Destroying Data** when retention policies expire is a critical component of some organizations. Unfortunately, this aspect of retention policies is often overlooked. Moving data throughout its useful lifecycle and then destroying beyond that period can not only help an organization from legal and compliance standpoints, but from a technical aspect old data being destroyed will free up space on production storage.

**Key Requirements in Determining Retention**

The primary motive driving retention policies should be business. Retention decisions should not be made solely by backup administrators. Meetings with all key decision makers, auditors, and any outside consultants should be conducted. They should be educated on the basics of what types of protection can be provided. Realistic RTOs and RPOs, realistic retention periods, and costs associated with meeting goals should all be addressed. It can sometimes be difficult for non-technical people to really understand what is best for their needs. When they think backup, they think nothing will ever be lost and anything can be recovered at any time. We know that is not the case. You may not have the power to make the final decision, but you do have the power to educate and influence good decisions.
In a perfect world all data would be kept forever, but in the real world this is not always possible. A key element to remember is the advantage of logically addressing data within an environment. This allows businesses to determine retention strategies based on business systems rather than physical servers.

There are several key issues that must be discussed when planning a retention strategy:

1. Business and government regulations
2. Business reputation & customer confidence
3. Current capacity and planned growth
4. Budgetary limitations
5. Risk assessment

Business and Government Regulations
Regulations such as Sarbanes-Oxley, HIPAA, and Gramm-Leach-Bliley have forced industry to look closer at how they protect information. This can be a difficult task, especially when there are no clear cut rules for retaining information. I have seen many businesses that infinitely retain critical data such as e-mail and financial records. They have no choice since government regulations provide guidelines that can be interpreted differently depending on which auditor you ask.

In explaining retention to a customer they asked “How long should we retain data for?” My response was “As long as your company tells you to!” Their response was “They don’t know!” The problem they had was there was no top driven data retention policy within the company and retention decisions were being made by the IT staff based on storage space and budget. This is definitely the wrong way to approach retention. Business and government requirements will most likely be the most critical variable in determining data retention needs.

Business Reputation and Customer Confidence
Depending on the type of business, reputation and customer confidence could be a huge determining factor in setting retention requirements. If you used a free E-mail service and one day you logged on to find only one message in your inbox you would probably be concerned. On opening that e-mail you realize that it is an apology letter that all mail had been lost and they are sorry. Would you continue to use their service?

I like to qualify disasters as sympathetic and non-sympathetic. Disasters such as Katrina, the 2003 Northeast power outage, and 9/11 all qualify as sympathetic disasters. Customers are more likely to understand the situation and accept data loss or interruption in service. On the other hand, if your building catches fire and burns to the ground, and you never properly protected data or maintained off-site copies, then customers may not be as sympathetic because you were not proactive in being prepared for such an event.
98 - Defining Data Retention & Destruction Policies

In both cases, determine your customer base and user base and consider how data loss may affect your ability to continue to do business and retain customers.

**Current Capacity and Planned Growth**

You can only retain information based on the capacity to hold the data. Analyzing your current ability to store information will give you a starting point for determining retention capability. If storage capacity does not meet retention requirements you’ll need to purchase more storage or change retention policies. Considerations must be paid to expected data growth. This includes the following determining factors:

- Incremental rate of change for existing data.
- Projected trends based on historical data.
- New projects which may implement new business systems.
- Number of copies of protected data and locations for the copies.

You need to consider how much data and where it will be stored. Historical information could be placed on tape to be archived off-site. User data may be required to be kept on fast disks for easy recovery. Capacity must be thought of not only as a total but broken down based on location of media, ease of access, and speed of recovery. Do not underestimate the amount of protected data that will need to be managed or you may find yourself running out of space.

**Budgetary Limitations**

Determining retention requirements may force a company to invest in more equipment to accomplish retention goals. However, budgetary limitations can affect the overall retention strategy and force you to readdress retention issues. Legacy hardware that has not reached the end of its lifecycle and strict budgets may force you to make do with what you have. This may ultimately force you to readdress retention strategies.

**Risk Assessment**

There are many different definitions of a disaster. For smaller companies, losing even one server can be catastrophic. Larger companies that cluster servers and store data on RAID arrays may think on a broader scale and consider a building loss as a disaster. Companies implementing site replication technologies could possibly sustain a site loss but with a larger cost of implementing and maintaining this type of storage infrastructure.
Placing data off-site for most companies is essential. Where to locate off-site data is a key issue. I was working for a company in southern California and after several meetings I convinced them to store data off-site. The office manager would take a tape home each night. The problem was that she lived five minutes from the building. I explained that with the potential for earthquakes, storing our backup data several miles from the office was unacceptable. Then I asked them what they would do if they lost all their data. Their response was “We work in a closely tied industry where we know everyone personally that we do business with. We would simply call them and reconstruct the information that was lost.” Their risk assessment was low since information could be recreated if lost. So they kept the data several minutes from the office.

Several years later I was talking to someone from the company. They told me they implemented a new database system using replication and implemented a very detailed disaster recovery strategy. I asked him what changed their mind and he responded “Sarbanes-Oxley!”

Risk assessment should be determined by perceived risk, value of the data, and the cost to reconstruct data if lost.

---

**Defining Retention and Destruction Policies**

Determining retention is a difficult task in many organizations. The problem with assessing requirements based on the criteria previously discussed is that retention requirements are based on guidelines and not rules. The point is, no one really knows. You can ask 10 people and get 10 different answers.

When defining retention policies the focus is usually on how long to keep the data for. An often overlooked point of discussion is “When does the data have to be destroyed?” Data destruction policies can be just as critical as retention policies. When the shredding truck comes to destroy documents and old tapes, a company may feel safe. With modern data management, the data being destroyed may only be a portion of what really needs to be destroyed. When planning retention policies, careful consideration should be paid to destruction policies as well.
100 - Defining Data Retention & Destruction Policies

Who Determines Policies

Business NOT backup administrators! That should be all that is written for this topic, but unfortunately it is not. Getting management to specify retention policies can be incredibly difficult. But the truth is it’s not always their fault. Non-technical people don’t really understand what is going on behind the scenes. They need to understand there is a limit to the scope of protection. Explaining what can be protected, how long it can be protected for, and the cost associated with protecting data can assist them in making intelligent decisions.

On the other hand, there are also the business managers that do not want to accept the responsibility for any decisions that can affect their career path. This presents a difficult situation for CommVault administrators who could become the scapegoat if data is lost and there is no accountability. Unfortunately there is not much that can be done in these cases, but the next two sections may provide some guidance to get reluctant business managers to commit to retention and destruction policies.

Document Policies

Arbitrarily configuring retention settings in a storage policy copy is a bad policy. Not sharing those settings with business managers is even worse. It is by design that when you create a storage policy the default retention is infinite. All retention policies should be well documented and shared with owners of the business data. The decisions of retention and destruction policies should of course be determined by management. These policies should be documented by them at least through Email, but preferably as official company documents. On the administrators end, reports should regularly be run and archived so they can later be referenced if needed. The Data Retention Forecast and Compliance report can be used to show data in storage, retention settings, and expected date of aging.

What the retention policies should be is part of the equation, but equally as important is documenting what those policies are. If archived data such as Emails are requested for investigation from six years ago, but your documented Email retention policy is five years, you will be in better shape than if you had no documented policies.

Default Policies

The major issue regarding retention policies is the lack of cooperation administrators get from the business side of an organization. Many business managers will rely on administrators to determine policies. From the company’s perspective, potential loss of data could cost millions. Consider point in time archive copies of financial records that must be maintained for extended periods of time. In the event that the data is needed for investigation and cannot be produced, the company may receive stiff penalties from regulating bodies. From the administrator’s perspective, he becomes the scapegoat. The final result is the company loses money and you lose your job.

One way to avoid this and help guide business managers into making a decision is to have documented default policies. Basically this would be presented as a multiple choice question by providing several retention policies a business unit can choose from. These default policies should be established with guidance from IT, cooperating business units, executive cooperation, and auditors if involved. The business unit can choose which policy best suits their needs. If a business unit requires customized retention policies, they can be worked into an existing storage policy by adding additional secondary copies, or a new policy can be created. It would then be a requirement for the business unit owners to sign off on the policy and this of course would be documented. This method will provide guidance to business units in making their choice, result in documented policies with responsibility on business managers and not Simpana administrators (which is how it should be), and could greatly simplify management by limiting the number of storage policies within a CommCell® environment.

**Adverse Inference**

In an investigation, if evidence from the defense cannot be presented as a result of intentional or unintentional data destruction by the defendant, the jury can infer that the evidence would have been adverse to the defense. This means with the lack of documents to prove innocence in a situation, the jury can interpret the destruction of the documents as intentional as it would have harmed the defense. This will allow the jury to adopt the plaintiff’s reasonable account of what happened. Spoliation of evidence can make a defendant appear guilty and in certain situations even if not legally wrong, could sway a jury’s decision.

When people think about preserving electronic data, they think about Enron or Martha Stuart and how electronic evidence led to guilty verdicts. But in most cases the preservation of data is used to prove the innocence of an individual or a company. This should be known by management and if not should be explained to them. It should also be explained that the Simpana software provides a wide range of methods to protect data. You don’t have to keep everything for ten years or twenty years, you don’t even have to keep specific data types for that length of time, you could keep specific user data based on retention policies. In this case, Simpana software provides incredibly granular levels of protection to meet legal needs of an organization.
102 - Defining Data Retention & Destruction Policies

Understanding Retention for Standard Backup iDataAgents

With the introduction of new Simpana® 9 features such as deduplication, DASH-Full, DASH-Copy and SILO tape storage, the philosophy and approach to configuring retention has changed significantly. Where organizations would traditionally conduct full backups on weekends when resources were not being used, Client Side Deduplication and DASH-Full now allows Full backups to run incredibly fast and use less network bandwidth. DASH-Copy makes copying data to secondary disk locations on or off site significantly faster using minimal bandwidth. The SILO to tape feature makes it possible to not even bother with retention and keep everything forever. These features are changing the way CommVault promotes configuring retention policies. In this section the focus will be on understanding retention and how these new features can allow Simpana administrators to think outside the box when implementing retention strategies.

Retention Rules

Policy based retention settings are configured in the storage policy copy Retention tab. The settings for backup data are Days and Cycles. For archive data the retention is configured in Days. Retention can also be set through schedules or applied retroactively to a job in a storage policy copy.

Retain are configured for a storage policy copy in the Retention tab. For standard backup data retention is defined as cycles and days, and archive data retention is defined as days.

Cycles

A cycle is traditionally defined as a complete full backup, all dependent incremental, differential, or log backups; up to, but not including the subsequent full. In real world terms a cycle is all backup jobs required to restore a system to a specific point in time. To better understand what a cycle is we will reference a cycle as Active or
**Complete.** As soon as a full backup completes successfully it starts a new cycle which will be the active cycle. The previous active cycle will be marked as a complete cycle.

*A full backup is completed on the Friday of week 1 starting an active cycle. The active cycle grows as the Incremental jobs run Saturday through Thursday. When the week 2 Friday full backup completes successfully it becomes the active cycle. The full and all incrementals from week 1 become a complete cycle. In this case there are 2 cycles*

An active cycle will only be marked complete if a new full backup finishes successfully. If a scheduled full backup does not complete successfully, the active cycle will remain active until such time that a full backup does complete. On the other hand a new active cycle will begin and the previous active cycle will be marked complete when a full backup completes successfully regardless of scheduling.
A full backup failed on Friday of week two when it was scheduled. The active cycle which started on the Friday of week 1 will continue to be active. Then a full backup was manually run on Tuesday of week 2. At this point, the Tuesday full would become the active cycle and the active cycle which started on Friday of week 1 will be marked complete.

In this way a cycle can be thought of as a variable value based on the successful completion or failure of a full backup. This also helps to break away from the traditional thought of a cycle being a week long, or even a specified period of time. In the above example, over a two week time period there are two cycles. The first started on Friday of week 1 and finished 11 days later on Monday of week 2. This cycle was marked complete after the Tuesday full from week 2 completed successfully.
The number of cycles that are currently being retained is the total number of complete cycles and the currently active cycle in protected storage. In the following diagram there are three cycles, two complete and one active.
Days
A day is a 24 hour time period defined by the time on the CommServe server. Each 24 hour time period is complete whether a backup runs or not. In this way a day is considered a constant.

*In the following example there are 17 days and 3 cycles.*

<table>
<thead>
<tr>
<th>F</th>
<th>S</th>
<th>S</th>
<th>M</th>
<th>T</th>
<th>W</th>
<th>T</th>
</tr>
</thead>
<tbody>
<tr>
<td>FULL</td>
<td>INC</td>
<td>INC</td>
<td>INC</td>
<td>INC</td>
<td>INC</td>
<td>INC</td>
</tr>
<tr>
<td>FULL</td>
<td>INC</td>
<td>INC</td>
<td>INC</td>
<td>INC</td>
<td>INC</td>
<td>INC</td>
</tr>
<tr>
<td>FULL</td>
<td>INC</td>
<td>INC</td>
<td>INC</td>
<td>INC</td>
<td>INC</td>
<td>INC</td>
</tr>
</tbody>
</table>

3 Cycles
17 Days

Why Cycles and Days?
The method CommVault software uses to define retention can be confusing. Consider the reasons why data is protected and how that effects retention planning and configuration.

Retaining Data for Disaster Recovery
When considering disaster recovery retention the key is determining how many complete sets of data is required to meet retention goals. This means the cycle criteria should be the key setting. Industry minimum standard for retention for DR is two cycles. CommVault has a minimum recommendation of two copies of two cycles, one copy for on-site and one for off-site.

Retaining Data for Data Recovery and Backup Data Preservation
When managing data for long term storage, days should be the determining criteria. If you want to give users the ability to recover home folder data for a period of 90 days, then the days retention should be set to 90. If you want to preserve data off-site for a period of seven years the days retention should be set to 2555 (give or take a day for leap years).
Cycle Criteria Guarantees Data in Storage
By setting the cycle setting to at least 1, some data will always be guaranteed to be available for restore. It might not be the latest data but it will be there. This is because the cycle variable is dependent on another full completing before a previous one ages. The day criteria is a constant and not effected by whether jobs run or not. If you retained data for 14 days and 0 cycles, and you don’t run jobs for 14 days, then no data will be in protected storage. For this reason CommVault strongly recommends never setting 0 cycles retention. Setting 0 cycles will be discussed later.

Notating Retention
Traditionally retention was notated as (cycles, days). (2, 14) represented 2 cycles and 14 days. In the storage policy copy’s Retention tab is shows days then cycles. So what is the best way? Whichever way works best for you. Retention is one of, if not the most important setting within the Simpana software. Notating retention in names of copies and documents is very common. I personally am a creature of habit and still tend to refer to it in the old notation style of (cycles, days). But with the concept of cycles changing because of newer technologies, and data being able to be stored for extended periods of time, days seems to be what everyone is interested in. So the following are a few examples for notating retention.

1. (2, 14) 2 cycles and 14 days
2. (C4/D30) 4 cycles and 30 days
3. Days=90 / Cycles=12 – 90 days and 12 cycles

The notation method is up to you, just keep in mind that it should make sense to you and others that may read it.

Days and Cycles relation
A rule of thumb that has been followed for years was that cycles and days should directly or indirectly equal each other. 2 cycles and 14 days with weekly full backups. 4 cycles and 30 days being approximately 1 month. 12 cycles and 365 days for month end fulls being retained for a year. But what about 52 cycles and 365 days? In situations like this it is rather irrelevant how many cycles are set. The truth is, 2 cycles and 365 days is good enough. You will meet your retention requirements since you will be keeping data for one year and if backups don’t run for over a year you are still guaranteed to have at least 2 cycles of data in storage based on the aging entire cycles rule.

When setting retention in the policy copy, base it on the primary reason data is being protected. If it is for DR ensure the proper number of cycles are set to guarantee a minimum number backup sets for full restore. If you are retaining data for data recovery then set the days to the required length of time determined by retention policies. If the data recovery policy is for three months, 12 cycles and 90 days or 1 cycle and 90 days will still meet the retention requirements.

Extended Cycles and Intermediate Jobs
With data deduplication, the traditional cycle is not as necessary as it once was. Weekly cycles were adopted because it would limit the number of jobs required for a restore to seven if running weekly fulls and daily incremental jobs to tape. Tape restores would sequentially restore data by linearly reading data off of each tape. The random access of disk allows data to be read from disk by randomly seeking right to each required block.
108 - Defining Data Retention & Destruction Policies

With deduplication, data can be stored on disk for much longer periods of time. These factors allow longer cycles to be run. Instead of a weekly cycle, a monthly or even quarterly cycle can be used.

**Example:** Data is being backed up to a deduplicated disk library. A daily DASH-Copy is run to copy data to a disk library at a DR site. Once a month a selective copy to tape will be performed which will be retained for five years. Since data is being copied to two disk locations, on and off site disaster recovery protection is provided. The month end copy to tape is for long term archiving of data. So instead of running weekly full backups, monthly fulls can be used which correspond to the point in time that tape copies are being made.

With deduplication after the initial full is run, only changed blocks and metadata will be committed to disk regardless of the type of backup performed (full, incremental or differential). When running extended cycles more incremental jobs will be retained using more disk space. This of course is dependent on the incremental rate of block change. A new feature in Simpana 9 is to remove incremental backups by performing a periodic full backup, daily incremental jobs, and periodic intermediate differential jobs. In the *Retention* tab of the storage policy copy the option *Do not retain intermediate incremental and differentials before a new differential* can be enabled to prune intermediate jobs that are not needed to restore the entire cycle.

Example: Monthly full jobs, weekly differential jobs, and daily incremental jobs are being run. When DIF 1 runs it will back up all data that has changed since the full backup ran. As a result all the incremental jobs from the full to DIF 1 are not needed and will be pruned from disk. When DIF 2 runs it will also backup all changed data since the full was run. As a result DIF 1 and all the incrementals between DIF 1 and DIF 2 will be pruned. The same applies when DIF 3 is performed. The result at the end of the cycle would be the full, DIF 3 and all incremental jobs after DIF 3 making up the cycle. This allows for more efficient use of disk media by pruning blocks for data that frequently changes.
The following diagram illustrates the pruning of intermediate jobs. When a differential is complete, all preceding incremental and differential jobs will be pruned leaving just the full and differential.

The drawback to this option is point in time restores of specific objects may not be possible. For example, during week 3 of the cycle a user requests a file to be restored to a point in time from Tuesday during week 1. If the file was not modified since that Tuesday it will be available in the DIF 2 job. If it was modified after the Tuesday from week 1, DIF 2 would contain the newer version of the file. So an advantage of keeping all the incremental jobs would be the point in time restorability of an object to any backup point within the cycle.

Traditionally the use of differential backup jobs was to improve the performance of restores by reducing the amount of jobs required to restore a cycle. With the nature of deduplication, the use of differential backups will have no impact on restore performance, only disk consumption.
110 - Defining Data Retention & Destruction Policies

**Data Aging for Non-Deduplicated Data**

There are two processes that will be performed during a data aging operation. **Aging** simply marks jobs that have exceeded retention as aged. **Pruning** will physically delete eligible disk jobs or recycle a tape when all jobs on it have been marked aged.

The *Data Aging* process will compare the current retention settings of the storage policy copy to jobs in protected storage. Any jobs that are eligible to be aged will be marked aged. By default the data aging process runs every day at 12PM. This can be modified and multiple data aging operations can be scheduled if desired.

**Pruning** is also part of the data aging process. How Pruning occurs depends on whether jobs are on disk or tape. For disk jobs if Managed Disk Space is disabled and no auxiliary copies are dependent on the jobs, they will be pruned. This will physically delete the data from the disk. If deduplication is being used, job blocks that are not being referenced by other jobs will be deleted. If Managed Disk Space is enabled, the jobs will remain until the Disk library reaches the upper watermark threshold defined in the Library Properties.

For tape media, when all jobs on the tape have been marked as aged, and there are no auxiliary copies dependent on the jobs, the tape will be moved into a scratch pool and data will be overwritten when the tape is picked for new data protection operations. In this case the data is not deleted and can still be recovered by browsing for aged data, until the tape label is overwritten.

**Rules for Aging Data**

There are several rules that are applied during the data aging process

1. Both days and cycles criteria must be met for aging to occur.
2. By default, data is aged in complete cycles.
3. Days criteria is not dependent on jobs running on a given day.

**Rule 1: Both CYCLES and DAYS criteria must be met before Data will age**

Simpana software uses AND logic to ensure that both retention parameters are satisfied. Another way of looking at this is the longer of the two values of cycles and days within a policy copy will always determine the time data will be retained for.

**Example:** Retention for a storage policy copy is set to (2, 3) or 2 cycles and 3 days. Now this example doesn’t make sense in the real world but it is being used to logically prove the statement that both days and cycles criteria must be met for data to age. By Monday three full backups have been performed. If we were to age Friday’s full there would be two 2 fulls left meeting our criteria of two cycles. However, the days criteria calls for three days and if the Friday full was aged only two days would be counted. The Friday full would therefore age on Tuesday.
Monday at 12 PM the data aging operation runs and determines no data can be marked aged.

Retention (2,3)

<table>
<thead>
<tr>
<th>F</th>
<th>S</th>
<th>S</th>
<th>M</th>
<th>T</th>
<th>W</th>
<th>T</th>
</tr>
</thead>
<tbody>
<tr>
<td>FULL</td>
<td>FULL</td>
<td>FULL</td>
<td>Cycles met but not days</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Tuesday at 12 PM the data aging operation runs and determines the Friday full can be marked aged.

Retention (2,3)

<table>
<thead>
<tr>
<th>F</th>
<th>S</th>
<th>S</th>
<th>M</th>
<th>T</th>
<th>W</th>
<th>T</th>
</tr>
</thead>
<tbody>
<tr>
<td>X</td>
<td>FULL</td>
<td>FULL</td>
<td>FULL</td>
<td>Both criteria have been met</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Rule 2: Data is aged in complete cycles by default

Backup data is managed within a storage policy as a cycle or a set of backups. This will include the full which designates the beginning of a cycle and all incrementals or differentials. When data aging is performed and retention criteria allow for data to be aged, the entire cycle is marked as aged. This process ensures that jobs will not become orphaned resulting in dependent jobs (incremental or differential) existing without the associated full.

Example: This is another retention example used to prove the rule. Retention is configured for (2, 7) 2 cycles and 7 days. Fulls are being performed on Fridays and Mondays, and incrementals on all other days. On Saturday the cycles criteria of two has been met meaning that if I removed a cycle there would be two left, a complete cycle (Monday – Thursday) and the full on Friday night. However, since we prune entire cycles we would have to age the Friday full and the incrementals from Saturday and Sunday. This would result in only five days which does not meet our days retention requirements of seven. So on Monday when the data aging operation runs (default 12PM daily) there will now be 2 cycles and 7 days which will allow the first cycle to be aged.

Retention has been defined for 2 Cycles and 7 Days. When the data aging operation runs on Saturday, the cycles criteria has been met but not the days criteria.
Retention has been defined for 2 Cycles and 7 Days. When the data aging operation runs on Monday both cycles and days criteria have been met and the first cycle will be marked as aged.

**Retention (2,7)**

<table>
<thead>
<tr>
<th>F</th>
<th>S</th>
<th>S</th>
<th>M</th>
<th>T</th>
<th>W</th>
<th>T</th>
</tr>
</thead>
<tbody>
<tr>
<td>FULL</td>
<td>INC</td>
<td>INC</td>
<td>FULL</td>
<td>INC</td>
<td>INC</td>
<td>INC</td>
</tr>
</tbody>
</table>

Week 1

Week 2

Both criteria have been met

**Rule 3: Days Criteria is not Dependent on Jobs Running for a Given Day**

A day will be measured as a 24 hour time period from the start time of a data protection operation. Days are considered constants since regardless of a backup being performed or completed successfully the time period will always be counted. If a backup fails, backups are not scheduled or if power goes out a day will still count towards retention. This is why it is so critical to measure retention in cycles and days. If retention was just managed by days and no backups were run for a few weeks all backup data may age off leaving no backups.

**Example:** During a Friday night backup power is lost in the building. Power is restored on Sunday resulting in two days elapsing and counting towards retention. Note that since the Friday full failed the cycle continues into the next scheduled full (following Friday). This illustrates the importance of defining retention in both cycles and days.
A failure of a full backup on Friday due to a power outage results in a cycle continuing until a valid full is completed.

Interesting note:
The default retention for a storage policy copy is infinite which is set by a check box. After deselecting the check box the default retention is 2 cycles and 15 days. The 15th day results in another full being performed before aging will occur for the first cycle. Now although this may be thought of as a good thing that the extra full is performed, what if the full fails to complete? In this case it is the 15 days setting that is forcing data to be retained here, not the cycles. The result of a failed full backup will have no effect on the aging of the first cycle in the diagram.
Defining Data Retention & Destruction Policies - 115

Retention configured to 2 cycles and 15 days results in another full being performed. The fact that the cycles criteria is set to 2 does not guarantee that the full must be completed successfully.

Retention (2,15)

<table>
<thead>
<tr>
<th>F</th>
<th>S</th>
<th>S</th>
<th>M</th>
<th>T</th>
<th>W</th>
<th>T</th>
</tr>
</thead>
<tbody>
<tr>
<td>FULL</td>
<td>INC</td>
<td>INC</td>
<td>INC</td>
<td>INC</td>
<td>INC</td>
<td>INC</td>
</tr>
<tr>
<td>FULL</td>
<td>INC</td>
<td>INC</td>
<td>INC</td>
<td>INC</td>
<td>INC</td>
<td>INC</td>
</tr>
<tr>
<td>FULL</td>
<td>INC</td>
<td>INC</td>
<td>INC</td>
<td>INC</td>
<td>INC</td>
<td>INC</td>
</tr>
<tr>
<td>FULL</td>
<td>INC</td>
<td>INC</td>
<td>INC</td>
<td>INC</td>
<td>INC</td>
<td>INC</td>
</tr>
</tbody>
</table>

Both criteria have been met

The key point here is this: If you want to retain 3 cycles of data, set the cycles to 3. Don’t try to figure out how many days there are in 3 cycles as it will not guarantee the protection you need. On the other hand if you need 2 cycles on hand at any given time changing the 15 to a 14 could potentially save quite a bit of media since the first cycle will age on Friday at 12 PM before the Friday full is performed. This will free up media to run that Friday full.

Setting retention to 2 cycles and 14 days results in the first cycles being aged on the Friday.

Retention (2,14)

<table>
<thead>
<tr>
<th>F</th>
<th>S</th>
<th>S</th>
<th>M</th>
<th>T</th>
<th>W</th>
<th>T</th>
</tr>
</thead>
<tbody>
<tr>
<td>FULL</td>
<td>INC</td>
<td>INC</td>
<td>INC</td>
<td>INC</td>
<td>INC</td>
<td>INC</td>
</tr>
<tr>
<td>FULL</td>
<td>INC</td>
<td>INC</td>
<td>INC</td>
<td>INC</td>
<td>INC</td>
<td>INC</td>
</tr>
<tr>
<td>FULL</td>
<td>INC</td>
<td>INC</td>
<td>INC</td>
<td>INC</td>
<td>INC</td>
<td>INC</td>
</tr>
<tr>
<td>FULL</td>
<td>INC</td>
<td>INC</td>
<td>INC</td>
<td>INC</td>
<td>INC</td>
<td>INC</td>
</tr>
</tbody>
</table>

Both criteria have been met

Data Aging for Deduplicated Data

During normal data aging operations all chunks referencing jobs that have exceeded their retention are marked as aged. With Simpana deduplication data blocks within chunks can be referenced by multiple jobs. If the entire chunk was aged then jobs referencing blocks within the chunk would not be recoverable. The Simpana software uses a different mechanism when performing data aging operations for deduplicated storage.

The data aging operation for deduplicated storage works by checking with the deduplication database to determine if the block is being referenced by any jobs. If the block is being referenced then it will be maintained in storage. If the block is not referenced then the block will be deleted from the disk. This means that when using Simpana deduplication data is not deleted from disk at the chunk level but instead at the block level.
In the following diagram three jobs are retained in storage. Job 1 has exceeded retention. Jobs 2 and 3 are referencing blocks from a chunk belonging to Job 1. When data aging runs it will determine if blocks from the chunk are being referenced from other jobs. In this case block C and D will no longer be referenced when job 1 is aged and they will be deleted. All other blocks will remain in deduplicated storage.

**Deduplication, Data Aging and Fragmentation**

CommVault stored data on disk in chunks which appear to the operating system of the Media Agent as a CHK file. When blocks are deleted from within the file white space results. Over time as the disk storage fills up this white space will be used for new chunk data which can lead to heavy fragmentation. This will result in slower restore and auxiliary copy operations. It is recommended to periodically defragment disk mount paths when they are storing deduplicated data. A fragmentation analysis can be run in the CommCell console by right clicking on the library and selecting **Fragmentation Analysis**. This operation will check for fragmentation but a 3rd party utility will be required to defragment the data.

**Zero Cycle Retention**

It is possible to configure a storage policy copy for a zero cycle retention. This can cause undesired results where data being retained can fall below the days threshold defined. It is strongly NOT recommended to set zero cycles for a policy copy unless another copy has been configured with at least one cycle defined.
Extended Retention

Standard retention allows you to define the length of time based on cycles and days that you want to retain data. Extended retention allows you to define specific retention in days that you want to keep full backups for. It allows you to extend the basic retention by assigning specific retention to fulls based on criteria configured in the extended retention settings. Basically it allows you to set a grandfather, father, son tape rotation scheme.

Example: You want to retain backups for 4 cycles and 28 days. You also want to retain a monthly full for three months, a quarterly full for a year, and a yearly full infinitely.

To accomplish this you configure retention as follows:

- Standard retention is set for (4,28)
- Extended retention is configured for:
  - For 90 days keep monthly fulls
  - For 365 days keep quarterly fulls
  - For infinite keep yearly full

Extended retention rules are similar to selective copies in that they only apply to full backups. However, a selective copy creates an additional copy of a full and assigns it a specific retention. Extended retention applies retention to an existing full and does not create an additional copy. Determine which is the more appropriate solution when planning retention strategies. Note that extended retention rules will not work if the policy copy is using deduplication.

GRACE DAYS

In Simpana 9 grace days can be configured to specify the number of days to go backward to find an eligible full for extended retention if the full designated for extended retention fails. You can also select the option to Select jobs from a previous tier. This is used when multiple extended retention rules have been configured. Lets say you are applying extended retention rules to keep weekly full backups for a month and monthly fulls for a year. If there is no eligible full available for the month end rule, enabling this option will allow a weekly full to have monthly retention rules applied to it.

Managed Disk Space

Managed Disk Space is a feature used with disk libraries which allows data to reside on the disk beyond its retention settings. This allows you to increase the chances of recovering data faster from primary storage on disk without changing retention settings. Managed data on disk is treated the same as retained data for data recovery.

Managed data will be held on the disk beyond the standard retention settings until an upper threshold is reached. A monitoring process will detect data exceeding the upper threshold and then delete aged jobs from the media until a lower threshold is reached. It is important to note that only aged jobs will be pruned. If all aged jobs are pruned and the lower threshold is not met no more pruning will occur.

Managed disk thresholds are configured in the disk library properties and can be enabled in each storage policy copy.
Defining Data Retention & Destruction Policies

As a general rule of thumb the upper threshold should be set to allow one hour of backups to run after the threshold is reached. The lower threshold should be set so that the managed disk space pruning operation will not run more than once in a backup time period as the pruning operation will have a negative effect on the performance of backups. For more information on configuring library settings for managed disk space see CommVault’s online documentation.

**Example:** User data on a file server has a recovery window for a period of 90 days. Disk media can hold up to two months of data. To increase the potential of being able to recover data from disk media and not have to recall tapes, the managed disk space option will be used.

**Solution:** Create two storage policies. The first will write data to disk for a standard (2,14) retention. The second policy will also have a (2,14) retention but managed disk space will be enabled. All DR data will be directed to the first storage policy. After the (2,14) retention is exceeded the data will be pruned. User data will point to the second policy. User data will remain on the disk until the upper threshold is reached at which point jobs will be pruned. A secondary copy to tape will be created with a retention of (12,90) to ensure that user data is being protected for the full 90 days.

*Diagram illustrating the use of two storage policies to backup one file server. Managed disk space is only enabled on the second policy where user data is being managed. This increases the chances of recovering user data from disk.*

Managed disk space is not available when using deduplication. This is due to the fact that data stored on deduplicated disk is being referenced by multiple jobs. If deduplication is being used with SILO storage, managed disk space can be enabled for the SILO so when disk threshold are reached, infrequently used data will be placed in a tape SILO freeing disk space. SILO will be covered in more detail in this chapter.

**Spool Copy**

The *Spool Copy* option can be used to take advantage of fast disk read/write access and its multi-streaming capabilities when there is limited capacity available on the disks. A spool copy is a no retention copy. Data is spooled to disk and then copied to a secondary copy. Once the data is successfully copied to the secondary copy, the data on disk will be pruned and the space will immediately be available for new backups. The spool copy option is not available when using deduplication.

**Job Based Retention**

Typically retention is based on company policy and therefore managed through storage policy retention settings that effect all data being managed by the policy. There may be situations where jobs retention would need to be individually set. With Simpana 9 there are two methods for accomplishing this.

*Retention Set Through Schedules*

A new feature of Simpana 9 is to configure retention through a schedule or schedule policy. This is done by setting the *Extend Job Retention* options in the *Media* tab of Advanced Options. The default setting is to use storage policy retention settings. You can set schedule based retention for a specified number of days or infinitely retain the data.

*Retention Applied to Job in Policy Copy*

Retention for a job in a storage policy copy can be retroactively modified by going to the job history for the copy. Do this by selecting the storage policy copy where the job is located, right click the copy and select View → Jobs. Specify the time range of the job then click OK. Right click on the job and select Retain Job. The job can be retained infinitely or until a specific date. The job icon will change to reflect that the job has been pegged down.

**Modifying Storage Policy Copy Retention**

Retention settings are stored in the CommServe database and are logically applied to media. This means if retention settings are modified in a storage policy copy the changed will be applied to all jobs managed by the copy when the next data aging operation is run. So if you change retention from (2, 14) to (4, 30) all existing jobs that are within the (2, 14) retention period will now have a (4, 30) retention setting. If you changed retention from (4, 30) to (2, 14) all jobs within the (4, 30) period will have a (2, 14) setting and the result will be two weeks of jobs in storage aging during the next data aging operation.
Why is My Data Not Aging

The two most common types of questions handled by CommVault support are related to network and media management issues. 9 times out of 10 if its network related, it is a host name resolution or DNS issue. When it comes to media management it is almost always “Why are my tapes not recycling?” There are several reasons why tapes do not recycle:

- Mixed retention on tape
- Failing Jobs
- Unscheduled Backups
- Deconfigured Clients

Mixed Retention on Tape
If jobs with different retention are on tape, the tape will not recycle until the longest retained job exceeds retention. There are several situations where jobs with mixed retention may exist on tapes:

- Backup and archive data on the same tape with different retention settings.
- Extended retention rules have been applied to a job.
- Schedule based retention has been applied to a job.
- A job has been manually retained on the tape.
- Job dependencies.

Failing Jobs
Since retention for backups are based on both cycles and days, failing full backup jobs can cause a tape not to be recycled. Consider a policy copy using a tape library is managing 30 subclients. Some of those subclients are from a Client server that has full back up jobs consistently failing for several weeks. If the retention is configured for 2 cycles and 14 days, there will always be a minimum of 2 full backups retained on media. If new full backups are failing, the older full backups will remain on tape. That may result in all jobs on the tape being aged except for the jobs from the failing client.

There are several methods to solve this problem:

1. If you are sure that the data is not needed, you can select the tape in the Assigned media pool, right click the tape and select Delete Contents. You will then be prompted to confirm and type Erase and Reuse Media. Note that once this operation is performed, the tape is moved back into the scratch pool where data can still be recovered until the tape header is overwritten.

2. Another method is to consolidate retained jobs to a new tape so the old tape can recycle.

Unscheduled Backups
If backups have been running for a Subclient and then that Subclient was unscheduled or removed from a schedule, full backups will no longer be performed on the Subclient. Just like failing jobs, unscheduled subclients will result in previous full backups being retained based on the number of cycles configured in the storage policy copy. This will cause the tape not to recycle. If the subclient should be backed up, make sure you schedule a job or associate the subclient with a schedule policy. If the subclient no longer needs protection remove the subclient.
122 - Defining Data Retention & Destruction Policies

from the storage policy. Note that all subclients must be associated with a storage policy in order to perform data protection operations. If you delete the subclient, data previously protected for the subclient will continue to be retained until the days criteria is met. If you want to keep the subclient create a place holder storage policy that you can associate the subclient with. It is very important to note that if a subclient is not scheduled then the data managed by that subclient will not be backed up. It is strongly recommended to schedule all subclients or remove them from the data set.

**Deconfigured Clients**

When a client is deconfigured, the license is released but the client will remain in the CommCell environment. This allows data from the client to be restored. The default behavior in this situation is to continue to retain data based on both cycles and days criteria. Since the client is deconfigured, no backups will be performed moving forward so existing data will remain in storage indefinitely.

**There are a few ways to approach this situation:**

1. If the license for the client is temporarily being released, Simpana 9 allows the license to be re-allocated to the client and backup jobs can then be run. In this situation, when backups recommence new full backup jobs being run will allow older cycles to age.

2. If the client is decommissioned and will no longer be used you can wait until you know the data is no longer needed and then delete the client. The next data aging operation will age all jobs for that client.

3. Another option is if you will no longer be protecting the client but you need to maintain the client in the CommCell environment. This is usually due to compliance requirements and CommCell reporting to reflect that the client existed but is no longer being used. There is an option in Control Panel → Media Management applet → Data Aging tab called Ignore Cycles Retention On De-Configured Client. Changing this setting to 1 will make it true and data will be aged for deconfigured clients based on the days retention criteria only.

**Job Dependencies**

In order for a secondary copy to successfully be created, source data for that copy must exist. The source data is determined by the Specify Source setting in the Copy Policy tab. The default source location will be the primary copy. If the source copy is required for an auxiliary copy, the source job will not age until the secondary auxiliary copy completes successfully.

**There are several methods to solve this issue:**

1. Of course the best solution is to make sure auxiliary copy jobs are scheduled for the secondary copy.

2. If you are not planning on making copies for the secondary copy, in the General tab deselect the Active option. This will make the secondary copy inactive and data will age from the source copy tape. You can later reactivate the secondary copy.

3. If you don’t want to copy specific jobs, choose the Prevent Copy or Do not Copy option in the jobs list of the storage policy copy.
Checking Tape Status

Viewing Contents of a Tape
You can view the contents of a tape by going to the library and expanding Media by Groups, going to Assigned Media, right click the tape and select View Contents. Any jobs that have exceeded retention will appear but be greyed out. Any jobs currently being retained will be in regular type.

Data Retention Forecast and Compliance Report
The Data Retention Forecast and Compliance Report will display all jobs for a selected storage policy down to the subclient level. It will provide the estimated aging date and reason for not aging. The reason for not aging will appear as a hyperlink which links to CommVault documentation providing an explanation of the reason.
Media Capacity Planning for Retention Goals

In order to meet retention goals there must be enough available media to hold protected data. The first thing to consider is that since Simpana software retains and ages data based on full cycles the amount of media available must exceed the amount of data and the length you want to hold it.

Example: If you want to retain data for 2 cycles and 14 days there must be enough media to retain three cycles. The reason is that the third cycle must complete successfully before the first cycle will age.

*Media capacity planning to keep 2 weeks (cycles) of data on hand requires 3 weeks of storage space. The 3rd cycle must complete before the first cycle will age.*

Retention (2,14)

First cycle can age

2 cycles and 14 days criteria met

Deduplication’s Impact on Media Capacity Planning

Data deduplication greatly reduces the amount of data moved over a network and stored on disk. It works by removing duplicate blocks during data protection operations and referencing those blocks in indexes and the deduplication database. Over time as more full backups are run, the deduplication ratio will improve. This means the number of cycles that can be retained on disk can be significantly increased. The actual amount of disk usage should only incrementally increase since only changed blocks are being stored.

Retention policies for data recovery would traditionally only keep a few weeks of data on disk for fast recovery of user data. If older data needed to be recovered, tape recovery would be needed. Often these tapes were kept off-site requiring tape recalls that would make RTO longer. For data recovery policies the use of deduplication can greatly reduce the RTO for user data by retaining the data for considerably longer periods on disk.

Another advantage of deduplication is point in time backups to tape like weekly or monthly would not guarantee user data could be recovered. If a user created and lost an object within the time period that point in time copy would not have that object. Deduplication allows for not only all objects, but all versions of the object from the point each data protection operation was performed.
**Deduplication to SILO storage**

Simpana’s SILO feature allows deduplicated data to be copied to tape. This works by using disk management functions to copy deduplicated data to tape without rehydrating it. This can greatly increase how long data can be retained for as well as greatly reduce the amount of tapes required to store data.

One of the negative aspects of SILO, and one in which there has been much negative spin by CommVault competitors is the increase length of restore operations. The increased restore times are due to the fact that data being deduplicated over time could result in data blocks being spread across multiple SILO jobs. Before data can be recovered from a tape SILO it must first be staged to disk. All jobs required to recover the required data would be staged to disk and then recovered. For people who look negatively at the SILO feature, they are just not getting the concept of what SILO is. This feature is not for disaster recovery or even data recovery of something backed up two weeks ago. This feature is for data that may be one, two or five years old. You could use deduplication and SILO to indefinitely retain all data within an environment. It is normal policy that as data gets older the possibility that it would need to be recovered decreasing and as such the RTO for that data increases. With deduplication to disk, depending on disk capacity you might be able to keep a year’s worth of data on hand for quick recovery thereby having a short RTO. For data beyond the year time period it is placed in tape SILO where the RTO would be longer.

**SILO versus Auxiliary Copy**

The first thing to understand is that a SILO and an auxiliary copy are two completely different things. An auxiliary copy is a job that runs to generate a copy based on secondary copy settings. For deduplicated data on disk, an auxiliary copy will rehydrate the data. The data will be retained based on the secondary copy’s retention settings. A SILO copy is not based on another storage policy copy’s configuration; it is based on the deduplicated copy that SILO has been enabled for. This means that the retention for a tape SILO is the same as the retention for the deduplicated policy copy. The data will not be rehydrated when it is copied to tape.

Both auxiliary copy and SILO have their specific use cases. Auxiliary copy should be used for short term retention or point in time copies. The data will be rehydrated during the auxiliary copy so restore operations will be considerably faster. Since the data is not being retained for long periods of time, tapes will be recycled for reuse. Point in time copies such as quarter end compliance copies need to be performed through auxiliary copies. SILO does not provide capability of doing point in time operations since the deduplicated data is based on all jobs being retained not just a specific job. In this case the auxiliary copy will rehydrate data for just the point in time that was defined in the secondary copy configuration.

SILO can be used as an extension of disk storage. Older deduplicated data will be moved to tape SILO and newer data will reside on disk. This allows a deduplicated policy copy to have extremely long retention settings without risking the potential of running out of disk space. SILO can also be used when all data is needed to be archived for a time period. This can be accomplished by sealing a deduplication store. The deduplication setting *Create new store every n months* starting from a specific date can allow data to be sealed and placed into SILO based on a time period such as fiscal quarters. When a store is sealed a new store will be created. This will have a negative effect on deduplication ratios since all blocks being written to disk will be written into a new store.
Data Destruction Policies

Data retention and destruction policies are used to preserve and destroy data based on its useful lifecycle. There are many methods for implementing data lifecycle policies. The most common from a data protection aspect is retention. The problem with this approach is that it still requires the media available to destroy the data. This may not be an issue for disks always attached to Media Agents but it can be a problem for tape media. The following section is designed to explain key CommVault features that can assist in implementing destruction policies.

Data or Information?

Data is what we backup. Disks, folders, databases are just data. Information is what is useful from a business perspective. Simpana administrators look at backing up the home folders disk. The user views the information within their home folders to be productive. The approach for implementing data destruction policies can be based on both data and information.

Data Destruction

When a job exceeds its retention the job is marked as aged. If tape media is in the library, old jobs are overwritten with new jobs. That means until such point that data is overwritten it is recoverable in several ways. The jobs can be browsed and data can be recovered through the CommCell console. The tape can be read using the Media Explorer tool. 3rd party tape tools can read the data, though with processes such as compression and multiplexing, serious forensic knowledge would be required to produce the data. CommVault has several methods to greatly reduce the potential of the data being accessible once the data lifecycle is exceeded.

Erase Media

Erase media operation will physically mount a tape and overwrite the OML header. This will make the data unrecoverable through the CommCell console, Media Explorer, or using the Simpana 9 Catalog feature. Tapes can individually be erased by selecting the tape in the scratch pool and selecting Erase Media. You can also mark all tapes for a storage policy copy to be erased by enabling the Mark Media to be Erased after Recycling option in the Media tab of the policy copy.

In this case data on the tape is still theoretically accessible through 3rd party forensic tools since the data is not physically being destroyed. In some cases this may not be a concern. If this is a concern, CommVault recommends using data encryption. This will make it extremely hard but theoretically not impossible to recover data. It is very important to state at this point that the only true was to make data unrecoverable is to physically destroy the media. No encryption, degaussing, or erasing method will ever guarantee 100% chance that data cannot be recovered.
For media to be erased, an erase media operation must be run. Erase media operations can be run on demand or scheduled. To run or schedule an erase media operation, right click on the library managing the media and select Erase Spare Media.
Information Destruction

Erase Data

The erase data feature provides functionality to selectively destroy information within a job. This can be implemented in two different ways:

- A Simpana administrator can mark data unrecoverable
- A user can delete an archive stub file which will mark data unrecoverable in protected storage.

The erase data feature logically marks the data as unrecoverable in the CommServe database. The information is not physically removed from media. Just like the erase media feature, CommVault recommends encrypting jobs to greatly reduce the risk of someone using 3rd party tools to recover the data.

Erase data can also be used in an archiving environment. Normally if a user deletes a stub of a file that has been archived, the stub will be deleted but the data will still be retained in CommVault protected storage. If Erase Data is enabled for the storage policy and the Subclient, you can apply the erase data policy to deleted stubs. When an archiving job is run it will scan for stub files. If any stub files have been deleted, the erase data feature will mark those files unrecoverable in protected storage.

Is this Right for Me?

Erase media is a CommCell level license. Once it is applied you need to enable Erase Data in the General tab of the storage policy. Once this license is applied the Media Explorer tool and the Catalog option cannot be used to recover backup data for the storage policy. This is because random binaries are written to the OML header in the Media Password location. The password will not be able to be used and you will always get a decryption error when using these tools. The Restore by Job feature will also be disabled for the storage policy.

When the erase data license is implemented it will only be effective when writing jobs to new or recycled media. The license cannot be retroactively applied to jobs already in storage. If the license is removed, it will only be effective when writing to new or recycled media. All jobs written to media for the storage policy when the license was being used cannot be recovered through Media Explorer or the Catalog feature.

If the erase data feature is something that would be of value to your organization then it is worth the risks previously described. If you are not sure, then it is recommended that you do not use it. If you have capacity based licensing arrangements with Simpana software, check to see if this license is installed. If it is, you may want to request that the license is removed from the CommCell environment licensing.

Information Destruction with File System One Pass (Turbo) Agent

A new feature within Simpana v9 software is the introduction of the One Pass Turbo Agent. This feature works by incorporating traditional backup and file archiving in a single operation. File objects will be backed up normally and based on archiving rules if any files qualify for archiving, they will be stubbed after the backup phase completes. At this point the stubbed files will be retained based on the existence of the stub in production storage. If a user deletes a stub, the corresponding stub in CommVault storage will also be deleted. At this point the actual file that was backed up will remain in CommVault protected storage until either the backup days retention is exceeded or the subclient setting Prune archived items value is exceeded. For this process to work, data must be backed up to disk storage and synthetic full backups must be used. In this case the life and destruction of the data will be based on the user, not a policy.

**Records Management for Data Destruction**

An advanced and powerful feature in the Simpana product suite that can be used to implement data destruction policies is Content Indexing. With Content indexing retention and destruction policies can be applied to all granular objects protected by CommVault. The objects can be indexed and searched based on contents to determine its relevance. Content Director policies can be designed to automate the search process and relevant data can be moved into legal hold policies with specific retention and destruction policies defined. With this method; files, document, and Email messages can be indexed and searched for key words. Relevant terms such as: ‘Top Secret’, ‘Confidential’, ‘CEO’, ‘Insider Trading’, etc… can be used along with specific owners and data types. Information can be preserved, destroyed, analyzed, or moved to 3rd party tools for further management or analysis. The concept of records management will be discussed in more detail in the Compliance, Records Management, and eDiscovery chapter of this book.