White
Paper

Building a Strategic Archive with CommVault Simpana Software

By Jason Buffington

February 2013

This ESG White Paper was commissioned by CommVault and is distributed under license from ESG.

Contents

Introduction

ESG has long argued that it is not a question of *if* organizations are going to archive; it's *how*. For years, organizations have reacted passively to digital information retention requirements by electing to put in place minimal resources to preserve information for compliance, legal, business reference, or system optimization purposes. Most companies have dealt with the archiving market's evolving dynamics by addressing an immediate need rather than by building any type of long-term strategy.

For example, many companies have had to deal with growing e-discovery demands that make it imperative to retain select archived data online for easy retrieval and export. The short-term resolution is to store the information on faster (yet more expensive) media. Deploying such a strategy does address the short term challenge. But over time, putting all information on costly storage is likely to be very expensive.

It is hard to fault IT departments and their business customers for simply addressing archive-related challenges as they come up. After all, it is far too complicated to predict what retention issues will occur in the future. The concern with constantly executing archive environment "fire drills," though, is that they run counter to the logic of an overall information retention process.

By nature, archiving involves *long-term* information retention. Shortsighted technology decisions usually end up costing a company more in the long run by forcing:

- Disruptive data migrations
- Unplanned purchases of additional systems
- Increased risk because business users cannot properly address legal and compliance needs

Today, when a long-term archive strategy supported by adaptable technology solutions isn't in place, potential costs rise even further: The already-flawed situation is exacerbated by explosive information growth (frequently called "big data" because the explosion is driven by higher content volumes and larger file sizes), increasing demand for end-user data access, tight budgets, and other factors.

Companies can continue to try to keep up by making tactical-level archiving process and technology decisions. Or they can embrace an archive strategy that balances solutions for today's pressing issues and with flexibility to address future retention requirements.

For example, by applying a more strategic mindset to the e-discovery situation referenced above, a company might shift its archived data to disk—choosing a platform that supports heterogeneous storage solutions, a private cloud environment, and public cloud environments. Doing so would give a company more control over its archive storage costs: The strategy and the underlying technology would enable the IT organization to pick what storage it uses for archived data and introduce cloud options for data that must be kept for extremely long periods of time.

The same type of analysis is suited to many archive solution capabilities in the marketplace today. The capability in question may not solve an immediate problem, but having a strategy that centers on both adaptability and flexibility will be extremely valuable in a few short years.

Of course, changing one's purchasing behavior relative to archiving is entirely dependent on the appropriate solutions being available. This paper discusses the reasons ESG believes CommVault Simpana software, a data management platform that delivers backup, archive, search, and analytics capabilities, could be a viable cornerstone of an organization's information retention strategy. ESG specifically examines Simpana archiving capabilities that organizations may not believe they need now but, given current archive market trends, will be extremely useful to them in the near future.

Getting More Familiar with the Archive Market

In an organization, many constituents—IT, legal, compliance, records management, knowledge worker representatives, etc.—usually get involved in information retention process and technology deployments. One factor playing into tactical archive decision-making is a lack of baseline archive market knowledge across those groups. Many people know what "has to be kept," but some don't know how or respect why it would be accessed. In other words, few know what actually drives retention in the first place.

Improving organizational understanding of the archive market, especially in regard to the trends that have affected and could affect it, will help enhance the perspective of cross-functional teams responsible for archiving technology decisions and implementations.

What We Know

Data Growth—It Is a Given

There is a reason why the IT market is enamored with the term "big data." The industry has rarely seen today's combination of increasing manual and machine-generated data and increasingly larger file/message/database sizes. For a variety of reasons, a good portion of this data needs to be archived.

As a result of primary data growth and market drivers discussed below, ESG estimates that organizations will archive more than 700 exabytes of data between 2012 and 2015 (see Figure 1).

Figure 1. Total Worldwide Digital Archive Capacity, 2010-2015

Total Worldwide Digital Archive Capacity, All Content Types, 2010-2015 (Petabytes)
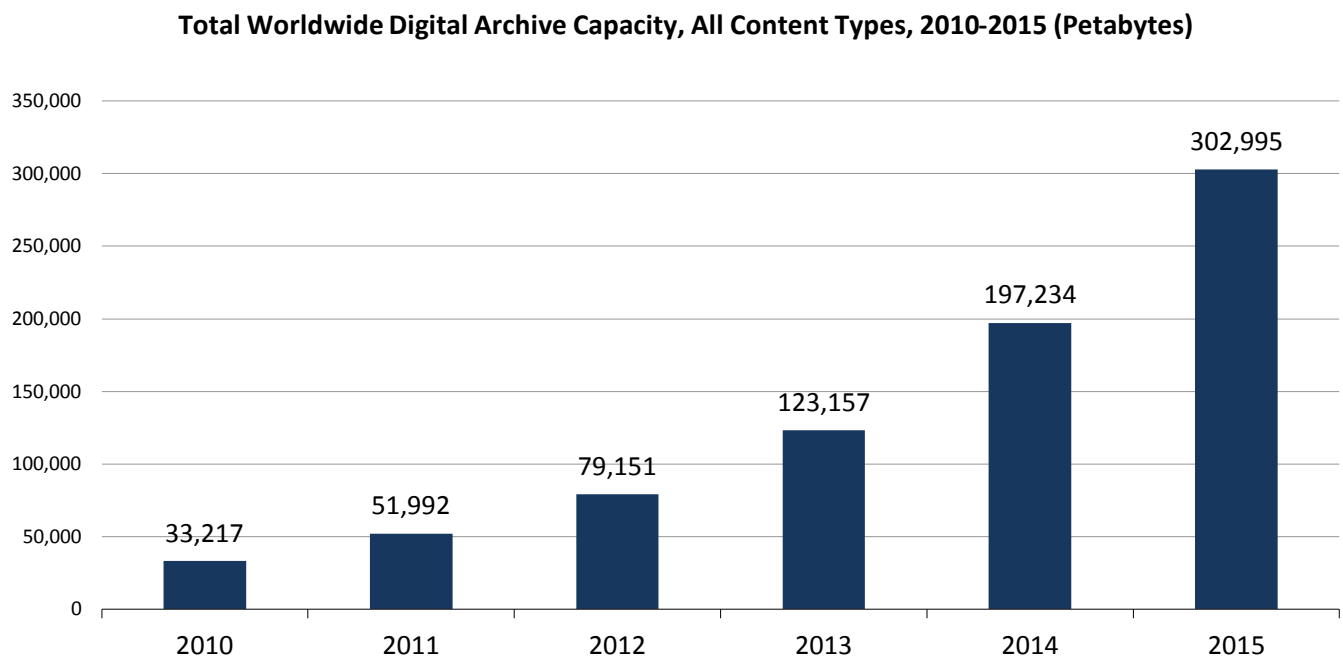
Source: Enterprise Strategy Group, 2013.

Staple Market Drivers

There is no foreseeable reprieve in reasons companies must or should archive. In the "must" category, electronic records management to satisfy compliance and corporate governance mandates won't subside unless the business world suddenly reverts to creating relevant documents on paper or if global governments and industry regulatory bodies relax specific rules requiring data retention.

In the "should" category, e-discovery continues to force companies to centralize critical data sources and place a subset of corporate information on indefinite retention until matters are resolved or disposition is legally acceptable. Most matters now involve electronically stored information, and it is unfeasible for companies to manually print out, preserve, and review all relevant digital data. Also, rampant data growth is stressing primary application environments and slowing response times. Shifting data from a primary environment while keeping it accessible is an effective way to balance application response times, data accessibility, and IT cost-control efforts.

Resources Are Limited, Not Infinite

Everyone talks about flat budgets and headcount freezes. We have to look at how those issues pertain specifically to archiving. Some companies have stretched backup environments too far to support meaningful archiving. Others have deployed separate, purpose-built archive solutions for every content type they need to archive. Still others are using a combination of backup and archiving solutions.

Which one is right? It depends on the IT staff's skills and the budget. Some companies save all data. Others delete nearly everything. In both cases, they often don't know to identify and save only what is dictated by business policy. A company should strive to be more efficient in executing archiving because based on underlying market drivers, the process isn't going away and in fact will get harder due to the expected data growth.

What We Can Expect

Requirements/Drivers May Arise or Change

It is impossible to know what governments and industry regulatory bodies may do in dictating what content must be saved and for how long. But it is safe to assume that existing mandates will evolve, new ones will appear, and few are likely to disappear. e-Discovery requirements are influenced by local and national judiciary bodies as well case precedent. Any legal matter can result in a new opinion or sanction that influences how electronically stored information has to be managed.

And cloud computing, too, is dramatically altering how companies tier their infrastructures, offering an entirely new way to cost-effectively optimize IT environments. Data already stored in the cloud may later be mandated for archiving. Clearly, cloud could be a great place to store archival data.

Different Content Types Will Have to Be Retained

Too many people think archiving applies only to e-mail because that was where the focus of "electronic" records management and "electronic" evidence started. Today, though, we have to account for data repository sources such as SharePoint that we didn't have a few years ago. Cloud applications are on the horizon as well. In addition, industry-specific data—such as healthcare medical images, telco call detail records, and oil and gas-related seismic imaging data—are (or could be) subject to retention requirements. Or, a business may simply want to keep this newer data for business-reference purposes.

Archive Access Will Evolve

Just a few years ago, access to archived information still had to go through IT, which meant access delays. More recently, IT organizations worked to offer broader, faster access to compliance, legal, and other groups. Today, many employees need ready access to what's been archived.

And the process now has to work without IT's involvement. It is even better if the archived information is available through the application that was originally used to create it (a native access experience). It is easy to envision external constituents such as contractors, service providers (external law firms or auditors for instance), and suppliers who may benefit from archive access. And, just as what happened with other corporate applications, archive access has to be extended to mobile devices (which have become integral to maximizing people's productivity).

What It Means and How to Prepare

Bigger Archives, Bigger Challenges

A bigger archive creates multiple challenges, including the challenge of accurately identifying what data has to be saved and how long to keep it. With a bigger archive, it also becomes more complicated to:

- Analyze the data and determine where to store it during its archive lifecycle
- Apply and update retention policies
- Delete all copies of data when retention requirements expire
- Properly secure data to allow only authorized access
- Find relevant information in a timely manner

Archiving Will Always Be a Moving Target

There is no wrong way to solve information retention challenges. However, it is smart to admit that room for improvement exists, and such improvements, if executed properly, can have financially positive benefits.

For example, most companies still solve archive needs with backup processes and technologies. This isn't wrong. But most of these approaches make it hard to archive individual files. (You either backup an entire data set and save it, or you don't copy it all.) These approaches also make it hard to alter retention policies upon receiving a discovery request.

A better way might be to use a platform designed to analyze, archive, and subsequently manage individual files. The savings manifest in faster e-discovery response times, reduced burdens on IT (if the archiving platform is self-service vs. requiring IT involvement), and lower storage costs because only a subset of data (vs. an entire backup data set) is actually kept. Companies can spur even more improvements if they can combine common backup and archive processes (such as file scanning and data deduplication) in a single operation while still supporting the two functional use cases (recovery and retention).

Apply a Strategic Perspective to Archive Decisions

In short, organizations have to be aware that information retention is unlike other IT back-end processes due to the lengths of time involved. Companies may need to or want to save data for many, many years. Investment decisions have to be based on today's problems *and* on future readiness. Otherwise, a company will be making archive-related purchases every time the market changes or evolves which, if history is any indication, will be frequently.

CommVault Software's Viability as a Strategic Archive

Known primarily for helping companies protect their critical business data, CommVault is quickly gaining momentum in the archive space. The rapid expansion—CommVault boasts thousands of archive customers—is attributable to the unique Simpana software platform. The Simpana Archive module runs on the same technology platform as the CommVault Simpana data protection offering and utilizes extended content capture options, a sophisticated search engine, and e-discovery and compliance information management workflows to support customers' long-term information challenges. These feature sets are the foundational elements of traditional purpose-built archive solutions. Yet many do not give CommVault credit for being a visionary in this market.

Customers using Simpana software for data management, that is, for both backup and archiving, will attest that they have actually separated these processes. They have just chosen to do so with a single technology platform, which has obvious economic and operational benefits.

This thoughtful approach to backup and archiving will be more valuable to companies as they optimize their archiving strategies. The Simpana software feature-set, which supports the cost-saving, risk mitigation, and process improvements discussed below, is becoming too hard to ignore for those that traditionally bought or upgraded archiving solutions to only solve an immediate need.

A Strong Architectural Foundation

Simpana software is built on a single platform. It provides a virtual information retention repository called Content Store, combined with an intelligent index that simultaneously supports data protection, archive, and storage infrastructure reporting operations.

Customers can achieve immediate savings by having only one solution to manage—there is no need to have separate application "silos" for archiving and backup. Instead, customers can set retention policies for backup and archive in one place. For legal purposes, a single query data repository to obtain the most comprehensive results in the least amount of time streamlines discovery. The legal department will also appreciate a central place to delete data, reducing the risk of lingering copies.

The centrality helps customers who are preparing for the future: Organizations running archiving as a derivative to backup today can make an easy transition/addition. Customers wishing to consolidate two separate processes can do that if they wish. And customers wanting the benefit of a purpose-built archive without the separate environment get what they want as well.

Because Simpana technology is a data management software-only offering, customers have the option to choose their own storage, avoiding hardware lock-in and potentially higher costs. Support includes immutable storage for those customers with unique legal and compliance requirements, lower-cost tape devices, and cloud storage. That option may not be needed (or even desirable) today, but it will be very good to have over the next five years, as companies look to reduce archive capital and operating expenditures, and as cloud offerings mature and become more central to mainstream IT.

By supporting both backup and archiving, CommVault has a unique engineering design point, especially when it comes to supporting new content types for either function. From a data-protection standpoint, all business information, no matter what application created it or where it is saved, has to be protected.

As result, the product has to find ways to identify and analyze the data so it can be managed under the Simpana platform. Many of the techniques, including snapshot management, application-aware data-management copy operations, and file system analysis techniques, can be used to bring data into the Simpana platform so that data-protection or archive policies can be applied.

Additionally, the company has architected unique archiving capture capabilities, such as e-mail journaling and SharePoint Blob Storage integration, into Simpana software. CommVault has also established partnerships, such as a relationship with Informatica, to add optimized functionality to identify database record archive candidates and help move data into the Simpana software. The result is that customers should be extremely comfortable that Simpana software will handle any current or potential content needing retention for data protection or archiving purposes.

Optimizing Data Management Functions

Information capture is just one of the data management functions that can be converged with Simpana software. Creation and enforcement of retention policies is another. Customers can establish rules that determine what must be archived and protected, the specific retention policy, and what happens when the data is deleted (automatically expired, notification before expiry, etc.).

The centralized, automatic management of retention and disposition eliminates redundant administrative efforts and provides business users, especially records managers and legal staff, with comfort that data is being properly managed and retained and that the policies can be easily audited.

Another set of efficiencies resulting from the single Simpana information management platform manifests in the storage process. Data is deduplicated globally across both data protection and archive functions, reducing the amount of information that needs to be physically stored. The capacity-reduction benefit is obvious, but it may be unappreciated in terms of what it means for deletion purposes. Once all retention policies for a file have expired,

the content can be deleted. There is no risk that another copy of that file resides elsewhere in archive or in the data protection environment.

Companies will undoubtedly need CommVault's storage resource management capabilities as both primary and archive environments get larger. Right now, IT departments do undertake some form of resource management, trying to figure out what type of data they have, where it is, and when it was last accessed. Such an exercise is extremely helpful to optimize storage. Simpana software allows customers to take the next step and archive data after the resource management analysis has been done.

IT is also able to leverage the product's resource management capability—analyzing data managed by Simpana software to determine if they should shift some data to a lower-cost storage platform or delete it because the retention policy expired but customers configured the system to "not automatically delete." Having direct insight into the archive enables customers to manage it intelligently from a single console while taking advantage of a heterogeneous storage hardware environment.

Simpana OnePass represents an even greater level of data management function convergence. From a single scan across file systems and Microsoft Exchange messaging applications, customers can perform backup, archive, and reporting functions without redundant operations affecting resources. For file system data, multiple agents installed on file servers and multiple file scans supporting these processes are not needed. Extending OnePass integration from files to Exchange now doubles email protection & archive and leverages a new Outlook plug in for self-service access to improve productivity. This is an example of an ideal future-friendly feature. Email and file data is driving significant capacity growth, and scale-out systems (file systems spanning multiple physical devices) are becoming more common as a result. Moving data once, customers can eliminate redundant process and reduce the frequency in which these large quantities of data have to be analyzed in order to be properly managed.

Extensible Archive Access

Simpana software's archive capture techniques leverage unique integration points such as SharePoint Blob Storage from Microsoft. It enables archive data to be accessed from within the application in which it is created, minimizing the need for end-users to go to separate environment to initiate a retrieval. This type of access is possible via Object-Based Retention, a feature within Simpana software that facilitates intelligent stubbing. A link (stub) is left in the primary application environment, yet the data is stored centrally. Users enjoy a native access experience, and retention policies can be applied in a single location.

It is also much easier to delete content after retention requirements expire. And the operation can be executed with confidence that no other copies exist. The resulting benefit: Both IT and end-user productivity are boosted.

Companies will find that the extremely sophisticated search engine within Simpana software—supporting both auto-classification and manual tagging/classification of data—will be of great benefit to attorneys and compliance officers. Today, they have to search large volumes of data specific to a topic, and they need it to be organized in order to make critical legal/compliance decisions. It will also be very useful for employees in future years who are looking for that "needle in a haystack" without even knowing where to start their searches.

With role-based access to search, companies can set up any number of secure roles with unique permissions. Employees can search their own data across backup and archives. Legal can search all content. In the future, organizations may want to create roles for partners or outside counsel or other external constituents. Today, this may seem like a strange concept. But keep in mind that requirements are going to evolve. In addition to defining access, Simpana software supports retrievals without IT intervention, speeding-up "time to information." IT gets tapped to set up the roles, but that is far less of a burden than servicing backup/restore requests and old-style daily data retrieval requests.

Simpana software has planned support to allow access to managed data via mobile devices running on the Windows, iOS, and Android platforms. The benefit of that capability is self-explanatory. Empowering a distributed workforce is key to productivity because archive is becoming a business-reference, quasi-business-intelligence application for many. Access from anywhere is crucial to keeping knowledge workers connected to information.

The Bigger Truth

When IT application infrastructures are overloaded with data, or when records managers need to extend retention requirements to a new digital data source, or when attorneys have to quickly search and preserve data for a "make or break" case, it is very hard to think about anything but the problem at hand, which can be solved by an archive solution.

Unfortunately, a hastily made technology purchase may not adequately address the next retention fire drill. Or, when IT departments evaluate financial and operational resource allocations for the upcoming year, they may realize how expensive a tactical archive decision really is, whether it involves using an aged backup process or an non-scalable purpose-built solution. These realizations are exacerbated by the longevity of the information-retention process and the market drivers that evolve over that time frame.

It is time for organizations to start to strategically evaluate archive solutions for capabilities they need now and feature sets that are likely to address to future needs. It is hard to predict the future. But as an industry, we already do know some things about archiving. Clearly, it is wise to focus investments on platforms with value, ones that have:

- A history of supporting new content types
- A plethora of storage options including cloud
- Access capabilities that are prepared to organize large quantities of data
- An ability to reduce IT management needs

CommVault is well positioned to meet these needs and, while no technology solution is future proof, Simpana software can make customers "future ready." Even if an organization doesn't need all the capabilities Simpana software has to offer (which are far too substantial to cover in one paper), they should consider ones that may be useful and beneficial to them down the line.