Simplifying eDiscovery & Compliance in the Big Data Era

A Holistic Approach to Information Management Means Big Benefits in Cost Reduction and Risk Management for Legal and Compliance Teams

Emily Wojcik
Product Marketing Manager
# Table of Contents

- The Big Data Challenge ..................................................................................... 3
- Crossing Big Data’s Backup and Archive Chasm ............................................. 4
- Taking a United Front on Data Convergence .................................................. 4
- Top 10 Questions ............................................................................................. 6
- CommVault Simpana®
  - OnePass™ ...................................................................................................... 7
- About CommVault ............................................................................................ 8
The Big Data Challenge

A relentless explosion of Big Data continues to ignite pervasive and persistent problems as organizations grapple with how best to retain, access, discover and ultimately delete content in compliance with evolving regulations. Big Data plagues many stakeholders, from IT to Legal. While IT departments grapple with how to support complex Big Data environments, legal teams are tasked with making accommodations for Big Data in the already expensive eDiscovery process.

The big deal with Big Data starts with the sheer volume, which is being generated by a growing number of devices, data sources and applications. IDC\(^1\) predicts more than 4,200 petabytes of new storage must be deployed this year alone to keep pace with data growth. Growth is being impacted by evolving data retention requirements, and industry regulations, including HIPAA, Freedom of Information Act (FoIA) among others, which necessitate that some types of data be kept for anywhere from a few years to indefinitely.

According to IDC\(^2\), the world generated more than one zettabyte (ZB), or one million petabytes (PBs), of data in 2010. By 2014, the growth is predicted to reach 72 ZBs a year, fueled in part by the rapid rise of machine-generated data. Increased usage of programmed trading and financial transaction systems, intelligent meters and other smart instrumentation along with a plethora of call-detail records (CDRs) now available on smartphones and tablets are driving the mammoth increase in machine-generated data.

At the same time, an ever-increasing influx of unstructured data (e.g., images, audio or video files) as well as semi-structured data (e.g., emails, logs, etc.) add yet another layer of management complexity, especially when determining the most efficient and reliable way to ingest, protect, organize, access, preserve and defensibly delete all this vital information. In particular, the need to preserve and retain data is becoming more complicated as organizations increasingly are being asked to retain data from a variety of sources, including emails, documents and rich media files, in appropriate business context.

All this data can be a huge asset, but without a modern management strategy, it can also be a huge liability. With these new paradigm shifts comes an increase in both the amount and complexity of data involved in discovery efforts. In sifting through voluminous Big Data to find responsive information, organizations can spend millions of dollars to isolate relevant Electronically Stored Information (ESI) and even more to review it. Simply put, the Big Data problem brings new meaning to the phrase, “looking for a needle in a haystack.”

Clearly, exponential data growth, diversity of data types and never-ending demands for optimized retention and discovery will create the perfect storm unless companies steer toward a more holistic approach to managing Big Data. In doing so, they can begin to view data backups and archives more strategically while leveraging integrated solutions for lowering storage costs and compliance risks. In addition to keeping pace with today’s accelerated data growth, it is increasingly critical that

---

organizations consider the future too. More precisely, technology that meets the demands of the business with a flexible and adaptable strategy that best reflects the needs of the business as it evolves. As a result, companies can then extract maximum value from all their crucial information in ways that produce valuable business benefits without the limits of technology lock-in.

**Crossing Big Data’s Backup and Archive Chasm**

For too many organizations, backup and archive functions are deployed and maintained as separate “silos” within an overall information management strategy. This is not smart for a number of reasons. Multiple, disparate hardware and software products typically manage these data silos, which leads to duplicate copies of information that must be protected and preserved. Additionally, legal pressure to find and preserve data typically causes yet more silos or a worst-case scenario—indefinitely extended retention of information assets because of inadequate visibility into what an organization is keeping.

The vast variety of data sources—encompassing audio, video and social media content now living on desktops, laptops and mobile devices—further exacerbates the administrative headaches of dealing with all this disjointed data, both from a cost and risk perspective. Compounding the problem is the fact that two distinctly different groups are traditionally responsible for data protection and preservation respectively within most corporate environments.

In most organizations, storage and backup administrators oversee data protection and therefore are heavily focused on the impact Big Data has on backup windows, recovery SLAs and infrastructure costs. While information management buyers are fixated on how Big Data affects data retention, discovery and information governance policies, they often operate without regard to the operational impact of these policies.

As a result, a chasm exists between these two critical constituents in ongoing Big Data conversations. According to Gartner³, backup complements archive and vice versa—yet backup administrators and information architects traditionally haven’t spoken the same language, and most tools and technologies address either one or the other of these disciplines.

While backup and archive serve different purposes, the functionality is similar: both processes make a copy of original data either for recovery or preservation. With that said, Gartner, among others, predicts that being able to look at backup and archive holistically promises significant cost reduction and risk management benefits⁴. The convergence of backup and archive is an emerging concept that’s gaining traction as organizations seek solutions to reduce the number of copies created for backup and archiving while more closely aligning data access policies for both.

**Taking a United Front on Data Convergence**

Given a shortage of budget and resources, though not driven by a lack of executive-level interest in good information management or data governance, the mantra seems to be “do more with less”.

---

⁴Id.
One way to accomplish this is the unification of backup and archive, but it requires cross functional teaming ensuring that the needs of the business are met for every stakeholder. This starts with developing a better understanding of how applications, users and critical business processes need to access data throughout its lifecycle. This effort requires cooperation and collaboration among business and IT stakeholders responsible for both recovery and discovery. This collective group should examine all the different policies and practices used to move, copy, catalog and access data for backups, retention, recovery, discovery and disposition. As part of this process, many of the hurdles thwarting streamlined access to individual and corporate data across the enterprise will be uncovered while at the same time, areas will be identified where limited visibility into vital information assets has created undue exposure to compliance and information governance risks.

Another typical outcome of the initial review process is the eye-opening realization that multiple copies of data reside everywhere—on physical and virtual servers, in the cloud, in backup repositories, in legal and IT archives as well as on employees’ desktops and mobile devices scattered throughout the company. While the number of redundant data copies can be reduced effectively and efficiently through deduplication, the biggest benefits come from consolidating data in a single data store that leverages a common hardware and/or software infrastructure for backup and archive.

The notion of a single data repository that eliminates redundancies and separate silos is compelling on many levels. A holistic approach that captures data once and then repurposes it for data protection and preservation is key to getting the right data into the hands of the right people so they can turn it into something more meaningful and actionable for the business.

Moreover, the ability to leverage a single-query data repository enables legal teams to obtain the most comprehensive results to an eDiscovery request in the least amount of time. Having a single collection ensures that all data sources are accounted for in a discovery effort, ensuring all case critical data has been collected, preserved and is ready for review and export to outside counsel if needed. Also, a central place to delete data also reduces both the cost and risk of inadvertently storing multiple copies. Understanding large data pools well enough to extract and collect relevant subsets for both reactive and proactive eDiscovery can prove to be a huge cost and risk reduction exercise.

An additional benefit of a converged data protection and retention strategy is centralized reporting that enables business and IT leaders to make more informed decisions with their data while bolstering analytical skills. Organizations can also extend their view into the business with embedded intelligence and analytical tools that provide granular insights into the ever-evolving role data can, and should play, in driving business direction.

Most important, companies can maintain a balance between capturing too much data or not enough as both scenarios pose potentially serious business risks. Armed with appropriate insight and tools, it’s possible to verify whether all data sources have been collected across the enterprise, especially at the network’s edge where remote laptops and mobile devices are often left unaccounted and vulnerable. With robust reporting and predictive tools, it’s much easier to forecast, analyze and budget properly for the ongoing onslaught of Big Data without compromising
data integrity, security, accessibility or accountability. Reporting can be used as a tool in the eDiscovery process to effectively defend methodologies of a data collection and preservation effort of an organization responding to litigation, regulatory request or an internal investigation.

The benefits of deploying an integrated information management strategy resonate throughout all levels of an organization, including outside of IT. Specifically, the corporate legal team will be able to preserve content with one mouse click on search results marked for “Legal Hold.” Another benefit will be reduced litigation risks and costs while gaining new insight into evidence earlier in the discovery process by eliminating duplicate copies of data, ensuring critical case data is captured and reducing the volume of data to be reviewed with a targeted approach. Those tasked with compliance will find easier conformity to industry regulations with improved records organization, synchronized retention, simplified disposition and upgraded ease of supervision, monitoring and auditing. For end-users, there will be better co-worker collaboration and sharing of passive content enterprise-wide.

In the world of Big Data, any opportunity to reduce the Tsunami-like flow of information is a step in the right direction. In the near future, however, organizations that are still using backups for long-term retention will encounter insurmountable challenges in accessing massive amounts of information for discovery and compliance purposes. They will have additional risk exposure when gathering large amounts of information from disparate silos, not having the comfort of a single platform to conduct enterprise searches for case critical data. Forward-thinking companies, on the other hand, which have embraced a unified approach for managing both backups and archives, will be able to take full advantage of a future-proof solution that elevates overall information management while providing appropriate access to business-critical information as it ages.

**Top 10 Questions: Is Big Data Creating a Big Information Management Problem?**

1. Have your file systems become “too big” to backup?
2. Are you using backups for both recovery and long-term retention?
3. Do you collect and capture backup and archive data in separate product silos?
4. Can you accommodate internal/external requests for information relating to discovery and/or compliance in a timely manner? What legal or other business resources are being expended to accomplish the same task?
5. Are your backup and archive policies aligned and automated?
6. Do you have duplicate data that exists across backup and archive copies? What unnecessary IT spend and resources are dedicated to managing this duplicated content?
7. How are you managing data sprawl? Are you collecting data from all sources, including remote desktops, laptops and mobile devices? What is your risk exposure if you miss critical data located “at the edge”?
8. Do you have sufficient visibility into vital data to assist with critical business decisions or Early Case Assessments?
9. Can you accurately predict the impact Big Data is having on your backup, recovery and retention infrastructure requirements in the near future?
10. Is your legal department able to self-service themselves on eDiscovery requests? Are these resources being consumed for manual tasks that may otherwise be accomplished by the use of appropriate technology?

If you answered yes to any of these questions, you might already be encountering or on the brink of being faced with a big backup and archive problem.

**CommVault Simpana OnePass™**

CommVault can help. CommVault’s Simpana OnePass™ feature answers the “Big Data” challenge with the industry’s first converged process for backup, archive and reporting from a single data collection and common infrastructure. The need for cost control within archiving environments is not going to disappear any time soon. And while CommVault’s approach of converging archive, backup and reporting within a single platform helps drive down costs, boost productivity and turn retained data into business value, there is more that can be done. With that in mind, Reference Copy is a content-based retention capability that has policy-driven rules for deep retention. With this capability, users can design and customize the data they want to retain and how they want to retain it. For IT teams, this means storage is optimized because only relevant data is retained. For legal and compliance teams, risk is managed because only what has business, compliance or evidentiary value is retained, while data that does not meet these standards is defensibly deleted from a single source. All managed data is kept in the ContentStore which provides a scalable, hardware-agnostic, virtual repository combined with an intelligent index that simultaneously supports data protection, archive, and storage infrastructure reporting operations (see Figure 1).

**Comparing Three Separate Workflows to Simpana OnePass™ Feature**

[Image of diagram comparing three separate workflows to Simpana OnePass™ feature]

Figure 1: Separate policies and processes create multiple data repositories, increasing cost and risks vs. the Simpana OnePass™ feature with single collection for reporting, backup & archive.
By delivering backup, archive, search, and analytics capabilities from a single, unifying code base and platform, CommVault is a viable cornerstone of an organization’s long-term information strategy. By embracing the Simpana OnePass feature and CommVault, information management stakeholders, from IT to Legal, will be able to realize the following benefits:

**Lower Costs, Reduce Complexity and Improve Operations**
- Eliminate “point” solutions with CommVault’s revolutionary data management convergence
- Reduce infrastructure, administrative and overall support costs
- Consolidate management and policies into a single console to maximize productivity
- Reduce infrastructure demands with less data traversing the network and less impact to the file system, improving disk I/O impact
- Collapse the backup window and server performance
- Increase productivity at the user and administrative levels with optimized system performance and transparent end-user access

**Meet Discovery and Compliance Challenges to Reduce Risk**
- Streamline preservation, legal hold actions and overall discovery for legal teams with a single query data repository and a central deletion point
- Reduce Big Data volume and eliminate data redundancy during the review process to lower third party costs
- Enable cost-effective long-term retention to meet information governance standards
- Analyze data with built-in reporting to make informed decisions about capacity usage and data relevance for retention or deletion
- Improve access to business information while increasing productivity

**About CommVault**

A singular vision—a belief in a better way to address current and future data management needs—guides CommVault in the development of Singular Information Management® solutions for high-performance data protection, universal availability and simplified management of data on complex storage networks. CommVault's exclusive single-platform architecture gives companies unprecedented control over data growth, costs and risk. CommVault Simpana® software was designed to work together seamlessly from the ground up, sharing a single code and common function set, to deliver superlative backup and recovery, archive, replication, search and resource management capabilities. More companies every day join those who have discovered the unparalleled efficiency, performance, reliability, and control only CommVault can offer. Information about CommVault is available at www.commvault.com. CommVault's corporate headquarters is located in Oceanport, New Jersey, in the United States.