

## Issue 1

- 2 CommVault paper:  
Solve Big Data  
Management  
Problems with an  
Intelligent Archive
- 3 The Trouble with  
Disparate Solutions
- 3 The Solution:  
Simpana OnePass™ for  
Converged Backup,  
Archive & Reporting
- 5 CommVault®  
Simpana® Software  
Delivers
- 6 From the Gartner  
files: Does Integrated  
Backup and Archiving  
Make Sense?
- 11 About CommVault

# Get Smart About Big Data

Integrated Backup, Archive & Reporting to Solve Big Data Management Problems

---

## Welcome Message

---

Without a doubt, we have entered the era of Big Data. According to Gartner, worldwide information volume is growing at a minimum rate of 59 percent annually. This sheer growth is fueled in part by the rapid rise of machine-generated data combined with an increasing number of devices, data sources and applications. At the same time, an overwhelming influx of unstructured and semi-structured data as well as the evolving requirements from industry regulations to preserve data is adding a layer of management complexity, cost and risk. Today's global enterprises are looking for new and better ways to manage their data, ways that are flexible and adaptable so they can also address their future needs. In doing so, companies can extract maximum value from their information to make better business decisions.

For the past 15 years, CommVault has been at the forefront of innovative breakthroughs in the backup and archiving marketplace. Our vision of modern data management, inspired by our unique Solving Forward philosophy, is transforming the way information is protected, managed and accessed. CommVault alleviates complexity and cost issues caused by multiple point products by offering a single view of all information regardless of where it resides. CommVault has evolved its Simpana



software by continually focusing on forward-looking solutions to anticipate and meet the ever-increasing data protection and information management requirements of our customers and partners.

More than half of the data in a typical enterprise does not get regularly accessed for a year or more. This unused data sitting on primary storage leads to higher backup costs, inefficient operations and non-compliance with corporate policies and external regulations. This newsletter is intended to help organizations understand how the convergence of backup and archiving simplifies data management and as a result, reduce costs, mitigate compliance risks and improve access to information.

Featuring research from

**Gartner**

In the article “A Big Idea for Big Data: Integrated Backup, Archive and Reporting to Solve Big Data Management Problems,” CommVault explores the pitfalls of taking a siloed approach in addressing explosive data growth. A companion article from Gartner, “Does Integrated Backup and Archiving Make Sense?” discusses the pros and cons of unifying these processes.

The line between backup and archive is blurring which is why we introduced the Simpna OnePass™ capability. By converging backup and archiving into a single process with Simpna OnePass™ – think of it like “back-ive” – organizations can manage Big Data more cost-effectively and holistically. For more information on how Simpna OnePass™ can simplify data management, visit [www.commvault.com/itleaders](http://www.commvault.com/itleaders).

---

## CommVault paper: Solve Big Data Management Problems with an Intelligent Archive

---

It’s no secret that “Big Data” is causing some big problems that are intensifying the demand for better retention, access, discovery and recovery of business critical information. How organizations handle the big data challenge will become a key foundation for corporate competitiveness, business intelligence and innovation, impacting every sector and every business function.

The big deal with big data starts with volume. The culprits are a growing number of devices, data sources and applications. The statistics are staggering. According to the University of California, San Diego report “How Much Information” published in April 2011, the world generated more than one zettabyte (ZB), or one million petabytes (PBs), of data in 2010. That’s over 3 terabytes of data per worker per year, or 12 gigabytes per day based on 3.18 billion people in the workforce (ILO and CIA Factbook). It takes a lot of storage and a lot of resources to manage all those zeros.

At the same time, an ever-increasing influx of content types adds another layer of management complexity, especially when determining the most efficient and reliable way to ingest, protect, organize, access and preserve all of this vital information. In particular, the need to preserve and retain data is becoming more complicated

as organizations today are being asked to retain data from a variety of sources, including emails, documents and rich media files, in appropriate business context.

This off-the-charts growth is being impacted by evolving data retention requirements and industry regulations, including HIPAA or Freedom of Information Act (FOIA) among others, which necessitate that some types of data be kept for period of a few years’ time up to indefinitely. It’s nearly impossible to predict what retention requirements will look like in the future but it’s safe to say existing mandates will evolve, new ones will appear and few will disappear.

Combine data growth, increasing data diversity and evolving data retention requirements, and you get a trifecta of problems that scream for an integrated solution to unlock the value of big data. This represents a clear opportunity to address these evolving dynamics with a future-proof strategy that’s supported by adaptable technology instead of shortsighted reactions to solving immediate storage needs. In doing so, backups and archives are viewed more strategically while lowering costs and governance risks.

---

Source: CommVault

## The Trouble with Disparate Solutions

In the report, “Does Integrated Backup and Archive Make Sense” published in March, 2012, Gartner states IT organizations that fail to optimize data retention and the number of data copies will suffer continuously increasing storage costs and increase their governance risk.” For too many organizations, backup and archive functions are deployed and maintained as separate “silos” within an overall data management strategy. This is not smart for a number of reasons. First, multiple and disparate hardware and software products typically manage these data silos, which lead to duplicate copies of information that must be protected and preserved. Additionally, compliance or legal pressures to find and preserve data in a silo’ed environment leads to inadequate visibility into what an organization is keeping and can therefore increase governance risk.

Compounding these problems is the fact that two distinctly different groups traditionally are responsible for data protection and preservation respectively within most corporate environments, making ongoing Big Data conversations disjointed. Backup administrators and information architects speak different languages. Gartner notes that organizations interested in archiving rarely are concerned about issues such as deduplication ratios, virtual machine support or backup modernization – all hot topics for today’s backup

administrators. Conversely, most individuals chartered with backup responsibilities aren’t interested in contextual analysis of data for retention or the need for transparent access to retained data from mobile devices. Most tools and technologies in the market today address either one or the other of these disciplines.

Enter CommVault. 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. As “Leaders” in Gartner’s 2013 “Magic Quadrant: Enterprise Backup/ Recovery Software” report and “Challengers” in Gartner’s 2012 “Quadrant for Magic Enterprise Information Archiving” report, CommVault provides industry-leading technology and superior feature-to-feature performance to address stand-alone backup and archive needs. Better yet, together CommVault Backup and Archive solutions pack a one-two punch that takes data and information management into a whole new stratosphere, literally providing benefits for every person who interacts with information in the organization.

Source: CommVault

## The Solution: Simpana OnePass™ for Converged Backup, Archive & Reporting

According to Gartner, “an improperly architected unified backup and archiving infrastructure may offer short-term improvement in operational efficiency but in the long term may fail to address user’s requirements.” To address this Gartner recommends, “When possible, seek solutions that leverage a common infrastructure.” 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. You can intelligently solve the challenge of

massive file and email growth with zero footprint archiving and no additional overhead, while managing retention and storage based on content and its inherent value to the business. Content based retention is supported by Simpana OnePass™ via the Reference Copy feature to offer policy-driven rules for deep retention. Policies retain data based on criteria such as file name, type, user/ group, keyword, Exchange classifications, tagging and more. Tiering can be applied based on content rules to any target, including tape or cloud. With these capabilities, users can design and customize

the data they want to retain and how they want to retain it. Reference Copy can be used for eDiscovery/compliance, ILM, analytics and many other use cases.

The notion of a single data store that eliminates redundancies and separate silos is compelling on many levels, including the opportunity to reduce the strain on congested IT networks, restricted hardware/software budgets and overburdened administrative teams. By moving data only once, you can minimize the impact of repetitive operations to file system and Microsoft Exchange messaging environments. A holistic approach that captures data once and then repurposes the copy for data protection and preservation is the 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 and compliance teams to obtain the most comprehensive results to an eDiscovery or audit request in the least amount of time. Having a central place to delete data also reduces both the cost and risk of inadvertently storing multiple copies. Going hand in hand with converged backup and archive are the alignment and automation of policies and best practices governing both areas to further reduce cost, risk and administrative overhead. More importantly,

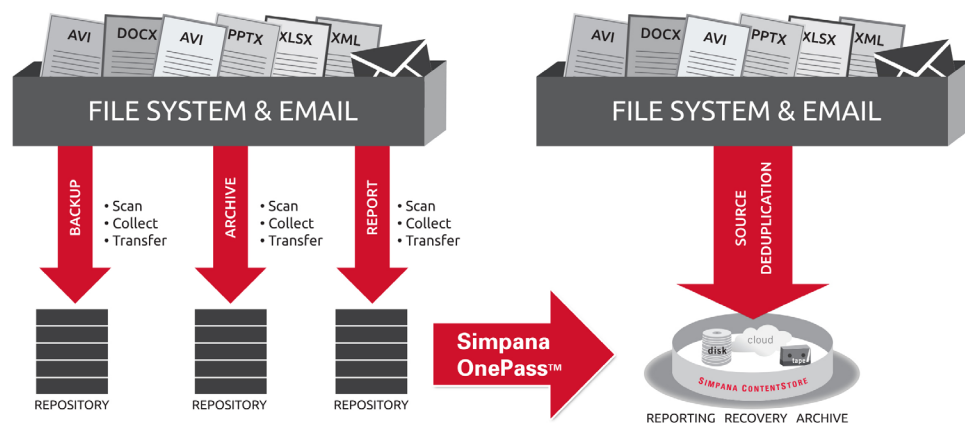
particularly in governance scenarios, companies can maintain a balance between capturing too much data or not enough as both scenarios pose potentially serious business ramifications. Armed with appropriate insight and tools, it's possible to verify whether all data sources have been collected across the enterprise, especially at the network's edge where remote laptops and mobile devices often are left unaccounted for and vulnerable.

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 also can 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. 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.

Figure 1: Separate policies and processes cause time, load and production impact vs. the Simpiana OnePass™ feature with single collection for reporting, backup & archive for both files and email.

Source: CommVault

**FIGURE 1** Comparing Three Traditional Data Management Workflows to Simpiana® OnePass™ Feature



Source: CommVault

## CommVault® Simpana® Software Delivers

In the world of Big Data, any opportunity to simplify and centralize the management of information to gain a more granular understanding of data is a step in the right direction.

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. Forward-thinking companies, on the other hand, who embrace a converged approach for managing both backups and archives, will be able to take full advantage of a future-proof solution that elevates overall data protection while providing appropriate access to business-critical information as it ages.

CommVault comes through on its promise to meet the data protection and retention challenges in this era of rapidly evolving data and information management. 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, you will be able to:

### Automate Data Management and Lower Costs

- Eliminate "point" solutions with CommVault's revolutionary data management convergence
- Reduce infrastructure, administrative and overall support costs
- Slash the combined operations time and impact typically required to backup, archive and report on a system by more than 50% to maximize productivity

### Simplify Management and Improve Operations

- 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 improve server performance

- Increase productivity at the user and administrative levels with optimized system performance and transparent end-user access

### Meet Discovery Challenges and Reduce Risk

- Streamline preservation, legal hold actions and overall discovery for legal teams with a single query data repository and a central deletion point
- Eliminate data redundancy during the review process to lower third party costs
- Enable cost-effective long-term retention to meet information governance standards

CommVault's Simpana® software eliminates the frustrations caused by data sprawl and poorly integrated legacy products, instead it provides the benefits of a truly converged solution build on a single platform that enables you to analyze, replicate, protect, archive and search data effectively and efficiently from any storage tier. Simpana software's granular view of your data provides the clarity and transparency that's often lacking in today's chaotic data management state, enabling you to lower costs, reduce risk and operational complexity. Maximum value can be extracted from all business critical information in ways that produce tangible business benefits. CommVault's holistic Solving Forward® philosophy offers unrivaled advantages over competitive options and ensures that your organization is well poised to address Big Data challenges today and into the future.

### About the Author:

Emily Wojcik is a product marketing manager for the Information Management business at CommVault. Emily has specialized in Information Management technologies for more than 12 years and has developed significant experience in the areas of archiving, compliance, eDiscovery, enterprise search and retention lifecycle management. Emily graduated with honors from Michigan State University with a Bachelor of Science degree in Communications. Emily is a member of ARMA, EDRM and ACEDS.

## Research from Gartner

## Does Integrated Backup and Archiving Make Sense?

**Gartner Foundational**

This research is reviewed periodically for accuracy. It was last reviewed on 20 June 2013.

Backup and archiving have long been thought of as complementary, yet few organizations have effectively implemented these technologies together. This research discusses the pros and cons of unifying these activities.

**Impacts**

- IT organizations that fail to optimize data retention and the number of data copies will suffer continuously increasing storage costs and increase their governance risk.
- An improperly architected unified backup and archiving infrastructure may offer a short-term improvement in operational efficiency but in the long term may fail to address users' requirements.
- Information architects who develop data life cycle policies that include backup and archiving will better support legal, compliance and user objectives.

**Recommendations**

- Use backup for operational recovery and deploy archiving for discovery and long-term record retention.
- Understand how applications, users and critical business processes will need access to data during its production and post-production phases in order to protect it properly, and to provide appropriate access as it ages.
- Look to leverage a common hardware and/or software infrastructure for backup and archiving to reduce storage costs and provide management efficiencies, but be aware that requirements for recovery point objective (RPO), recovery time objective (RTO) and access to data should dictate the optimal infrastructure that is deployed.

- Include backup and archiving policies and processes in the scope of information governance.
- Develop storage access and retention policies that incorporate the service requirements of users addressing both long-term access and immediate physical and logical recovery.

**Strategic Planning Assumption(s)**

By 2015, only 15% of organizations will attempt to converge backup and archiving policies and processes, up from 5% today.

**Analysis**

For years, organizations have been trying to improve the efficiency, cost and effectiveness of their backup processes. While strides have been made across all market segments, the problem of the "shrinking backup window" continues to be ever-present. Archiving has been marketed as one solution to this problem – if you archive old data and remove it from primary storage, it won't be included in the backup stream, or even deduplicated and, hence, will result in shorter backup windows. Similarly, strategies that promote removing old backups from the backup system and replacing them with archives have been marketed as a way to simplify restores and improve performance.

Certainly, archiving can reduce the amount of storage an organization must procure and manage. Archiving has also been deployed for its ability to support e-discovery and compliance, frequently with the intent to stop using backups for long-term retention and discovery activities better suited to a contextually-aware product like archiving; a practice that Gartner endorses.

Backup complements archive and archive complements backup – theoretically. In reality, what Gartner continues to see is that these technologies are deployed in silos in most environments. Storage and backup administrators are responsible for data protection, and slough off archiving as "something the lawyers and businesses need to figure out." Application administrators, archivists and information

architects realize that their retention and information governance policies should extend to backup (backup generates copies of data which need to be managed, after all), but they usually don't reach out to their data protection peers. Different administrators, different buyers and different technology value propositions all contribute to a fragmented data life cycle implementation. Gartner believes that organizations must get control of their data by implementing complementary backup and archiving policies for retention, access, recovery and discovery. For example, a simple retention policy may stipulate: "Backups should be retained for 90 days, and anything older than that is an archive. Archives should be retained for seven years." A simple access policy may stipulate: "Backup to disk should meet RTO objectives of two hours and email archives should enable immediate transparent access to archived mail for one year." Organizations that consider the nuances of these policies will be well positioned for improved cost and risk reduction in 2012 and beyond.

Despite good information and governance benefits that can be realized by managing backup and archiving holistically, most organizations today find this very difficult to do. Backup administrators and information architect/archivists haven't traditionally "spoken the same language" – and most tools and technologies on the market today address either one or the other of these disciplines. Gartner rarely speaks to organizations interested in archiving that are concerned about deduplication ratios, virtual machine support or backup modernization. Conversely, it is rare to find a backup administrator who is interested in contextual analysis of data for retention, transparent access to retained data from mobile devices and advanced search. Nevertheless, the cost reduction and risk management benefits of looking at backup and archiving holistically as described here can be significant. But a word of caution: These efforts should be balanced with an organization's readiness to tackle this – it may be more costly to establish policies and procedures for converged archiving and backup for purposes of information governance than to continue with these as separate disciplines.

**FIGURE 1** Impacts and Top Recommendations for Integrated Backup and Archiving

Impacts	Top Recommendations
IT organizations that fail to optimize data retention and the number of data copies will suffer continuously increasing storage costs and increase their risk.	<ul style="list-style-type: none"> <li>• Establish the use of backup for recovery and archive for preservation and/or discovery</li> <li>• Minimize the number of copies to decrease storage costs and simplify management</li> <li>• Align retention policies for backups with those for archiving</li> </ul>
An improperly architected unified backup and archiving infrastructure may offer short-term improvement in operational efficiency but in the long term may fail to address users requirements.	<ul style="list-style-type: none"> <li>• When possible, seek solutions that leverage common infrastructure</li> <li>• Align storage repositories with requirements to access data</li> <li>• Ensure that optimized solution can offer legal hold and integrate with tape for longer term retention</li> </ul>
Information architects who develop data life cycle policies that include backup and archiving will better support legal, compliance and user objectives.	<ul style="list-style-type: none"> <li>• Replace the use of backups for long-term retention with archives</li> <li>• Consider security: rogue copies, unmanaged copies lead to increased risk</li> <li>• Align backup and archiving policies for better responsiveness to end-users</li> </ul>

Source: Gartner (March 2012)

**Impact: IT organizations that fail to optimize data retention and the number of data copies will suffer continuously increasing storage costs and increase their governance risk**

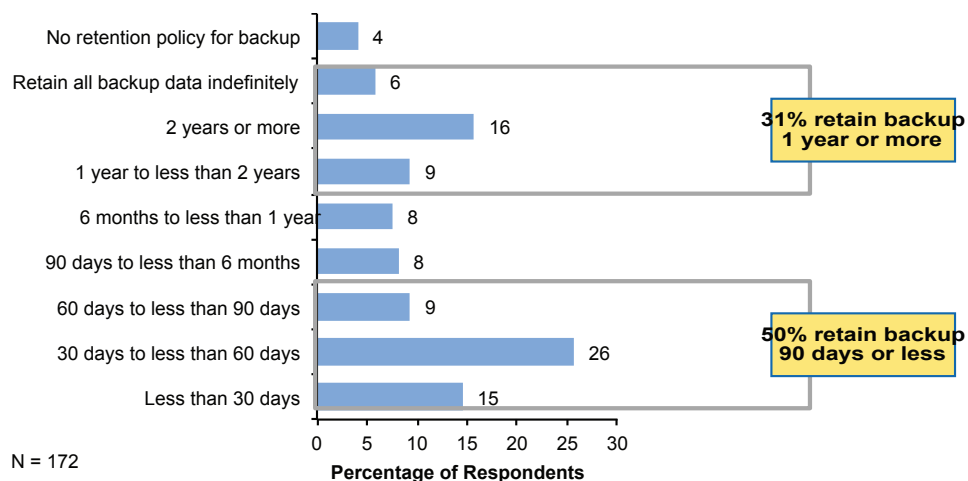
Backup practices and policies are designed to ensure recoverability of data, and are an element of a risk management strategy. Managing information for compliance and e-discovery are also part of a risk management strategy, with a focus on ensuring that the right data can be produced when required. Both of these functions employ retention policy management, that is, the original data (in the case of archiving) or a copy of the data (in the case of backup) is kept for a specified period of time.

While these disciplines serve very different purposes, the functionality inherent in backup and archiving (making a secondary copy of data and preserving it for a period of time) greatly overlaps as both processes can keep the original data in place while also making a copy of the data. Many organizations will deploy one of these strategies (nearly always backup) and call it a day. In this case, old, redundant and infrequently accessed data is being backed up over and over leading to increased storage costs. For those organizations that do employ backup and archiving, many do not align policies for retention and deletion and, consequently, data that is archived continues to be backed up, again leading to increased storage costs.

Things can be better: operational efficiency can be improved via a rationalization of retention according to desired outcome. Cost can be contained by decreasing the amount of storage that is required when the number of redundant copies of data that are captured, pulled through the network and retained on disk and/or tape (and that are later copied or replicated for disaster recovery reasons) are reduced. A well-deployed archiving strategy can save up to 60% in backup costs and reduce backup times by as much as 80% by reducing the redundant and nonactive data that is continuously captured and stored but is not likely to be accessed.

In the last three years, Gartner has found that backup retention has been decreasing as measured by responses to questions about backup retention policies. From European and North American Gartner conference survey data, Gartner found that in late 2011, large enterprises were retaining 50% of backup data for 90 days or less, up from 42% in 2010 (see Figure 2<sup>1</sup>). The largest backup retention period response is 30 to 60 days, with just over a quarter (26%) of those sampled using this policy. This trend accelerated beginning in 2009 due to global macroeconomic concerns and the need to reduce the cost of backup/recovery and consume less storage. Despite half of organizations effectively using backup for operational recovery of recent (90 days or less) data, 31% of respondents retain data for one year or more, which was also exactly the same percentage as in 2010. However, the number of organizations retaining backup data indefinitely fell from 10% to 6%.

**FIGURE 2** Backup Retention Policy



Source: Gartner (March 2012)

Using backups as an archive is not Gartner's recommended best practice. Backup is for recovery and archiving is for discovery and long-term data preservation. We reiterate our opinion that backups should be used for operational recoveries only, which typically is in the 30- to 90-day.

#### *Recommendations*

- Establish a practice whereby backup is used for operational recovery and archive is used for preservation and/or discovery.
- Minimize the number of copies created for backup and archiving to decrease storage costs and simplify the management process.
- Align retention policies for backups with those for archiving.

**Impact: An improperly architected unified backup and archiving infrastructure may offer short-term improvement in operational efficiency but in the long term may fail to address users' requirements**

If organizations were to design data life cycle management tools from scratch, with no legacy constraints, they might design a solution that would capture data once through a single set of infrastructure and be able to repurpose that copy for operational recovery all the way through long-term record retention and e-discovery. In the last five years, consolidation has been a major goal of IT organizations and is particularly evident in the backup space. Reducing the number of backup applications, deploying deduplication to contain the redundant number of copies of data and standardizing on common policies and practices to drive down errors and reduce risk have all been common activities within the data center. With this in mind, the notion of a collapsed set of infrastructures for backup and archiving, where multiple agents, policy engines, repositories and administrative consoles give way to a unified toolset becomes attractive. Note that vendors today offer varying degrees of actual integration, but some have road maps for greater future optimization possibilities.

Infrastructure should be streamlined as much as possible. Solutions that reduce the software infrastructure and, particularly, that can manage retention holistically are desirable for cost containment and improved manageability.

#### *Recommendations*

- As dictated by RPO and RTO requirements, seek solutions that leverage the same infrastructure for backup and archiving to simplify management and to contain costs (e.g., common agent, common policy and reporting engine, common repository, etc.).
- Align storage repositories with requirements to access the data (operational recovery, preservation, discovery or audit).
- Implement appropriate policies for backup retention and archive retention and reduce the number of redundant copies of data that are captured and retained (especially if these copies are stored in multiple repositories).
- Ensure that optimized solutions can offer appropriate legal hold capabilities for both backup and archive data and that the solution can integrate with tape for longer-term retention of the archived data.

**Impact: Information architects who develop data life cycle policies that include backup and archiving will better support legal, compliance and user objectives**

More organizations are focusing on information governance as a way to ensure that their data meets internal and external compliance requirements (see Note 1). Certainly the management of backup and archiving falls into this definition.

Managing information for compliance and e-discovery ensures that you can produce the right data when you need to, that you delete unnecessary data in a timely way and, importantly, that you understand where all the copies of your data reside. Archiving is more suited than backup to these requirements by virtue of its object-level retention management, indexing, basic e-discovery and auditing capabilities. Keeping track of the number of copies of a data object by holding the "copy of record" in an archive and managing all other copies according to a well-documented retention and deletion schedule means that organizations are less likely to be out of compliance with internal and external regulations for data management. With respect to backup, it's critical that organizations understand how many copies of a backup they're keeping and for how

long in order to appropriately respond to requests for information. With information governance as a focus, organizations can focus on keeping track of backup copies, move data into an archive when appropriate and, at that point, discontinue backup of original data. As a documented policy, this lets the organization authoritatively and defensibly produce an inventory of all copies of data when required for discovery or audit.

Optimizing backup and archiving policies for purposes of information governance also means better responsiveness to end-user requests for information beyond e-discovery and audits. Users who have lost data should be able to retrieve data according to an SLA with the backup administrator, while users who need transparent access to much older data should be able to get to that data using whatever search criteria or access methodology makes sense for their application. Proper management of data as it ages means that user requirements for retrieval and/or access are consistently being met.

#### *Recommendations*

- Replace the use of backups for long-term retention with archives; unplanned redundant backup copies contribute to the explosion of data volume and little else.
- Utilize backup and archiving to address some data security issues, including the generation of rogue and unmanaged copies (such as Microsoft Exchange .PST files), and endpoint device issues such as data loss and legal hold.
- Develop aligned access policies for better responsiveness to auditors, litigation support and end-users for data recovery or contextual search; understand which tools support which requirements and eliminate redundancy.

#### **Note 1. Information Governance**

Gartner defines information governance as the specification of decision rights and an accountability framework to encourage desirable behavior in the valuation, creation, archival and deletion of information. It includes the processes, roles, standards and metrics that ensure the effective and efficient use of information in enabling an organization to achieve its goals.

#### **Evidence**

<sup>1</sup>Conference kiosk polling data was used from Gartner's 4Q11 European Data Center Summit and U.S. Data Center Conference. These conferences have a large enterprise audience and the respondents in this survey represent a broad cross section of industry verticals.

Source: Gartner Research G00231898, David Russell, Sheila Childs,  
21 March 2012

## 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 employs individually-licensable modules 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](http://www.commvault.com). CommVault's corporate headquarters is located in Oceanport, New Jersey, in the United States.

