

Technical Review

Cohesity Data Cloud Data Management Platform

Date: January 2023 Author: Tony Palmer, Principal Validation Analyst; and Justin Boyer, Validation Analyst

Abstract

This Enterprise Strategy Group (ESG) Technical Review examines the Cohesity Data Cloud software-defined data management platform, focusing on how Cohesity reduces capacity, cost, and complexity by consolidating functions: collapsing workload silos, eliminating copies, and enabling efficient access to and sharing of data. ESG also examined how Cohesity delivers 96x data reduction or greater to most of their customers.

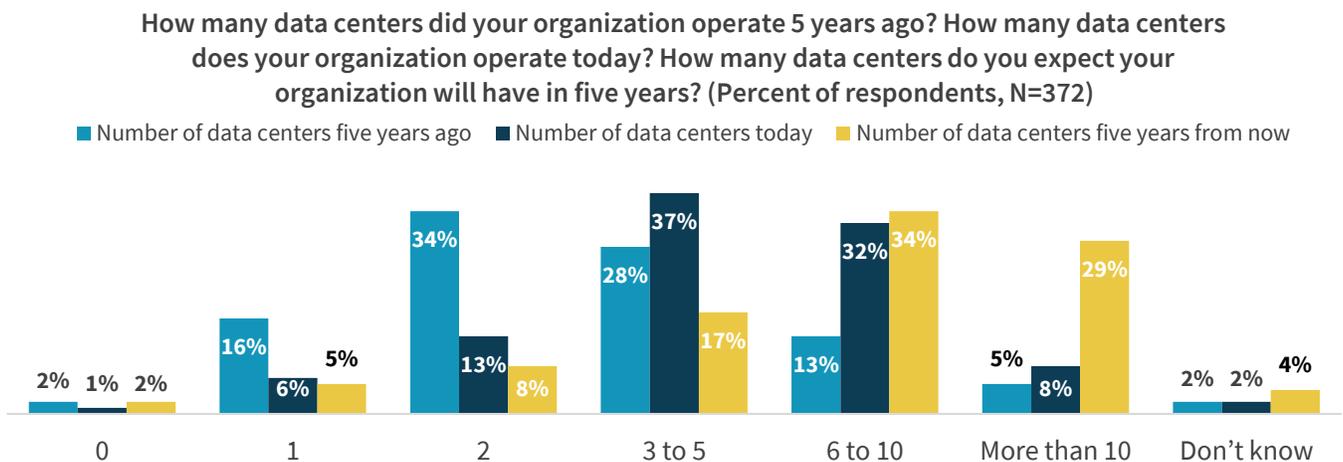
The Challenges

Organizations have made tremendous improvement in their ability to mine, organize, and create new value from their data. This ability enables companies to leverage their data to transform both IT and their business for the digital age and deliver a genuine competitive edge. Yet, as data grows and proliferates among different geographic and operational silos, including the cloud, the ability to see, access, manage, and harness the power of that data becomes considerably more challenging.

Organizations are still experiencing significant “data sprawl,” reporting that for every 1 TB of production data, they need 4 TB or more of secondary data, for both data protection (58%) and non-protection/non-production purposes (51%).¹

Digital transformation is also leading organizations to modernize and re-invest in data centers to take full advantage of the insights possible. Today, more than three-quarters (77%) of companies surveyed are operating at least three data centers, with nearly two-thirds (63%) of organizations planning to operate at least six data centers in five years (see Figure 1).² Modernization strategies for data centers can include consolidation and shifting adoption to public cloud resources. This has caused a rise in efforts to both increase interoperability across hybrid clouds, as well as invest more in data centers to achieve a cloud-like experience on-premises. This indicates that data sprawl is on an upward trend, making data reduction technologies critical.

Figure 1. Organizations Are Investing Heavily in New Data Centers Over the Next Five Years



¹ Source: Enterprise Strategy Group Complete Survey Results, [From Data Backup to Data Intelligence](#), January 2022.

² Source: Enterprise Strategy Group Research Report, [Application Infrastructure Modernization Trends Across Distributed Cloud Environments](#), March 2022.

Source: Enterprise Strategy Group, a division of TechTarget, Inc.

The Cohesity Data Cloud Data Management Platform

Cohesity Data Cloud is a modern data management platform designed to provide a comprehensive range of data services for data protection, data isolation, file and object storage, and disaster recovery. Cohesity Data Cloud is available as self-managed software and SaaS. Cohesity uses its hyperscale architecture to keep pace with organizations' businesses and data growth across data centers, edge locations, and public cloud environments. Cohesity's analysis of call-home data from their global customer base shows that the overwhelming majority of their customers achieve 96x data reduction or greater. Cohesity's variable-length deduplication engine runs globally across the cluster and can execute inline or post-process depending on the workload and data attributes. Cohesity's patented SnapTree snapshotting technology limits the number of hops to retrieve data blocks to two to enable the platform to offer virtually unlimited immutable snapshots without impact to performance.

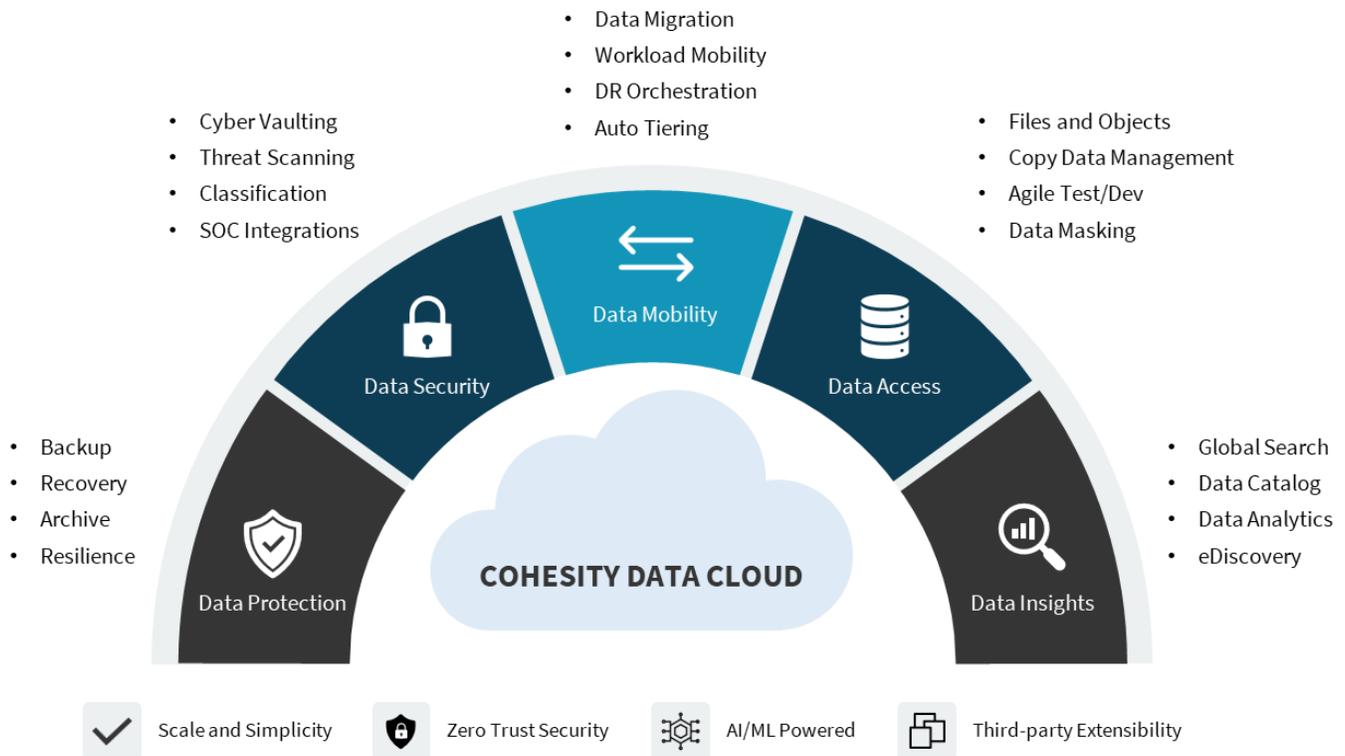
Cohesity offers multiple data services that run discreetly on the Data Cloud platform.

- **Cohesity DataProtect:** A software-defined backup and recovery solution, DataProtect eliminates multiple legacy point products, including backup software, master and media services, target storage, and cloud gateways—simplifying data protection and providing comparable functionality on a single scale-out solution.
- **Cohesity SmartFiles:** Designed from the start with security, manageability, and scale top of mind, SmartFiles optimizes cost and efficiency for unstructured data management with modern software-defined file and object services for the hybrid cloud. SmartFiles consolidates data silos target and securely manages unstructured content and application data, including digital libraries, archives, rich media, video surveillance, big data, and backup data sets.
- **Cohesity DataHawk:** DataHawk provides multiple cloud service offerings to repel and recover from cyber incidents. It leverages AI/ML to detect anomalies that could indicate an emergent attack, utilizes threat intelligence to ensure recovery data is free of malware, and provides data classification to identify the exposure of sensitive and private information when an attack occurs. It also provides point-and-click data isolation and integrates with numerous SIEM, SOAR, and vulnerability scanning solutions.
- **Cohesity FortKnox:** FortKnox is a SaaS data isolation and recovery solution designed to enable a modern air gap strategy for the cloud era and improve cyber resiliency. FortKnox balances organizations' security and agility priorities with an immutable copy of data in a Cohesity-managed cloud vault via a virtual air gap. This can give organizations relying on FortKnox an additional layer of security against ransomware and other cybersecurity threats through physical, network, and operational isolation.
- **Cohesity SiteContinuity:** This disaster recovery offering is engineered to automate failover and failback orchestration for mission-critical workloads. SiteContinuity can be self-managed by customers for site-to-site disaster recovery (DR), or it can be procured from Cohesity as a Disaster Recovery as a Service (DRaaS) solution using AWS infrastructure.
- **Cohesity Marketplace:** An ecosystem of apps and integration solutions—both Cohesity-native and third-party—that run directly on Cohesity's Data Cloud platform to allow data processing in place. By unifying all integrations and featuring them in the Marketplace, Cohesity enables customers to easily find solutions for their evolving data management and protection needs.

Cohesity Data Cloud scales simply by adding hyperconverged nodes to the cluster, each of which brings CPU and storage capacity, expanding the global filesystem. Data Cloud includes near-unlimited snapshots that can be created on demand; instant restores at scale; global deduplication across the cluster; a simple UI for end-to-end workflows; policy-based automation; linear scalability; granular VM-, file-, and object-level recovery; AES-256 encryption; and integration with tape

libraries and the cloud for long-term archiving. Once data is backed up on Cohesity, it can be used for other use cases such as spinning up zero-cost clones for test/dev and analytics.

Figure 2. Cohesity Data Management Platform



Source: Enterprise Strategy Group, a division of TechTarget, Inc.

Data is intelligently tiered between HDD and flash, and the cluster maintains two or more copies for availability. Cohesity has multiple data redundancy options that administrators can select based on their needs. Data replicates automatically within the Cohesity cluster and is optimized for fault tolerance. Upgrades, node addition or removal, and other maintenance tasks are non-disruptive.

Cohesity Data Cloud is a SaaS-based application that provides a single view and global management of all organizational data and workloads wherever they reside—on-premises, in the public cloud, or at the edge. Cohesity Data Cloud utilizes machine learning algorithms to proactively assess IT needs and automate infrastructure resources.

Case Studies

Enterprise Strategy Group recently interviewed several Cohesity customers to quantify the economic value of Cohesity’s solutions.³ One customer mentioned that their storage costs dropped by more than 70% after implementing Cohesity SmartFiles. Another customer appreciated the visibility and insight into the data that Cohesity provided, allowing them to take action to contain data sprawl. Another cited the ability to recover from ransomware attacks in less than two hours instead of two to three weeks as Cohesity’s biggest benefit. Altogether, customers interviewed cited cost savings, improved operational efficiency, and improved business agility/reduced risk as benefits they saw from using Cohesity.

³ Source: Enterprise Strategy Group Economic Validation, [Benefits of Managing Unstructured Data with Cohesity SmartFiles](#), February 2022.

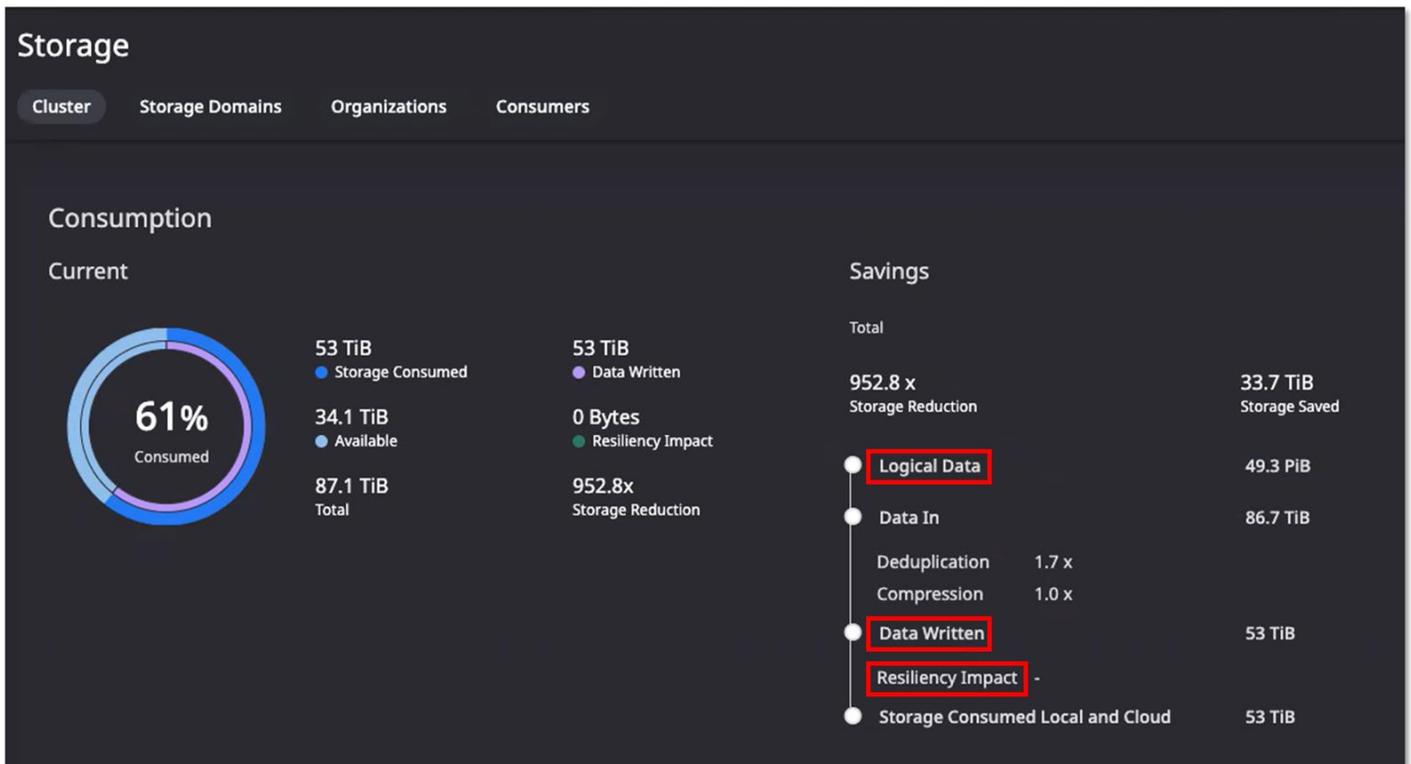
Cohesity Delivers Up to 96x or Greater Data Reduction to Optimize Savings

Enterprise Strategy Group (ESG) analyzed data reduction statistics from Cohesity’s global installed base. Our goal was to determine the extent of data reduction the typical Cohesity customer could expect. This review gave ESG an opportunity to see what real customers were experiencing with Cohesity.

Figure 3 shows the Storage Dashboard Cohesity customers use to view storage statistics. These statistics can be used to measure the average data reduction performed by Cohesity for their customers. Logical Data refers to the combined total of data in the objects managed and protected by Cohesity. Data Written refers to the data written to the Cohesity cluster after data deduplication and compression. The Resiliency Impact is the space required to honor resiliency settings in addition to the data written. Cohesity charges customers strictly on the Data Written.

ESG verified the data reduction calculations of real-world Cohesity customers. For this calculation, the resiliency impact is ignored since it represents the space Cohesity uses to ensure resiliency and the customer is not charged for it. Therefore, the calculation used to determine total data reduction for a particular customer takes the logical data (the totality of customer data being protected) and divides it by the data written (the data after compression and deduplication).

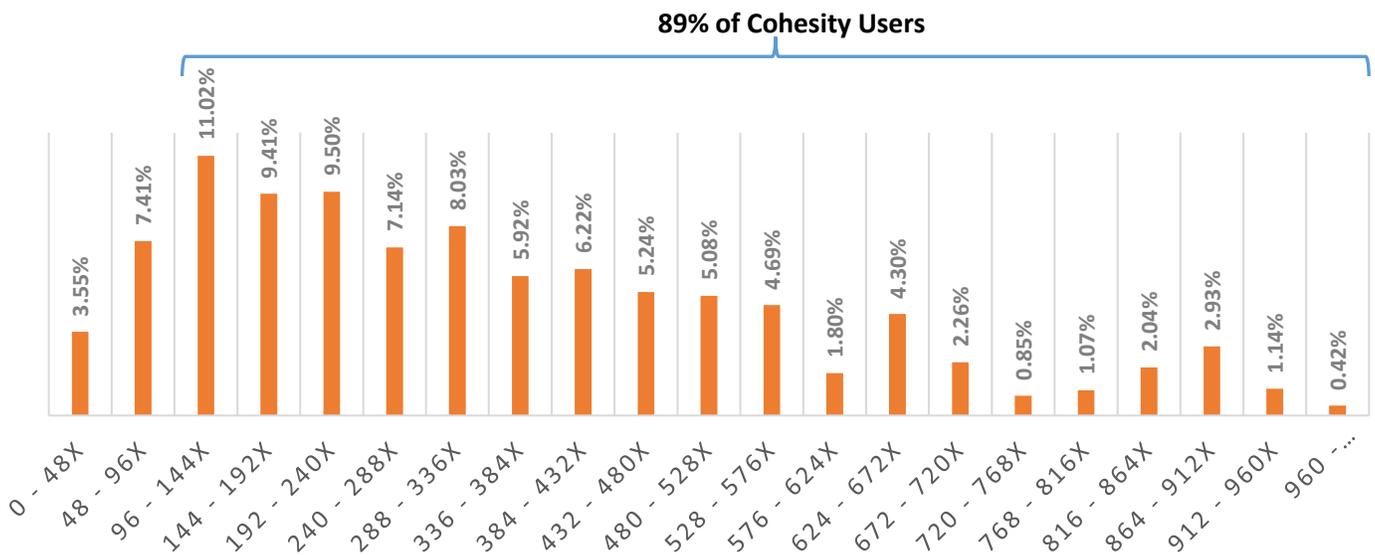
Figure 3. Cohesity Real World Data Reduction Screenshot



Source: Enterprise Strategy Group, a division of TechTarget, Inc.

Figure 4 shows the real-world data reduction based on a sampling of over 3,000 Cohesity customers. The X-axis shows the ranges of data reduction the customers experienced. Above each range, the chart displays the percentage of customers whose data reduction calculation fell within that range. ESG found that 89% of customers achieved data reduction of 96x or higher, with many achieving much higher reduction rates.

Figure 4. Cohesity Real World Data Reduction Chart



Source: Enterprise Strategy Group, a division of TechTarget, Inc.

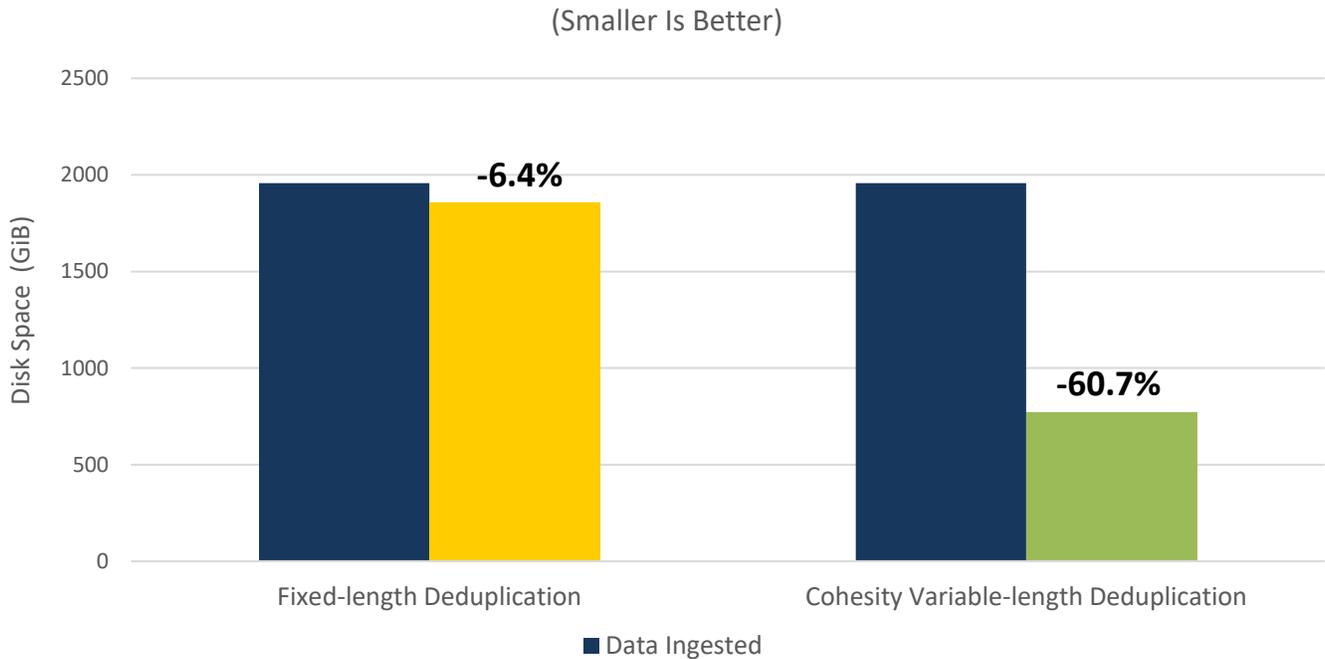
Enterprise Strategy Group Tested

Enterprise Strategy Group (ESG) previously examined data reduction and small file efficiency in a controlled environment.⁴ We compared Cohesity Data Cloud global variable-length deduplication and small file optimization to a system that utilizes fixed-length deduplication and does not deduplicate globally across views—Vendor X. We used synthetic data sets generated specifically for these tests: A 1TB unstructured data set with a duplication factor of 2:1 for the deduplication test, and a set of 1 million 1KB files for the small file efficiency test. Erasure Code (EC) redundancy was set to 2:1 in both systems. While these tests used fixed data sets in isolation, the results support the global deduplication rates observed in the field.

In the first series of tests, we copied the 1TB unstructured data set to two views—think volumes—in the Cohesity platform and two views in the alternative system. As seen in Figure 5, Cohesity’s global deduplication alone reduced the data footprint by 60.7% while the non-global deduplication system was only able to achieve 6.4% data reduction.

⁴ Source: Enterprise Strategy Group Technical Review, [Cohesity Data Management](#), October 2019.

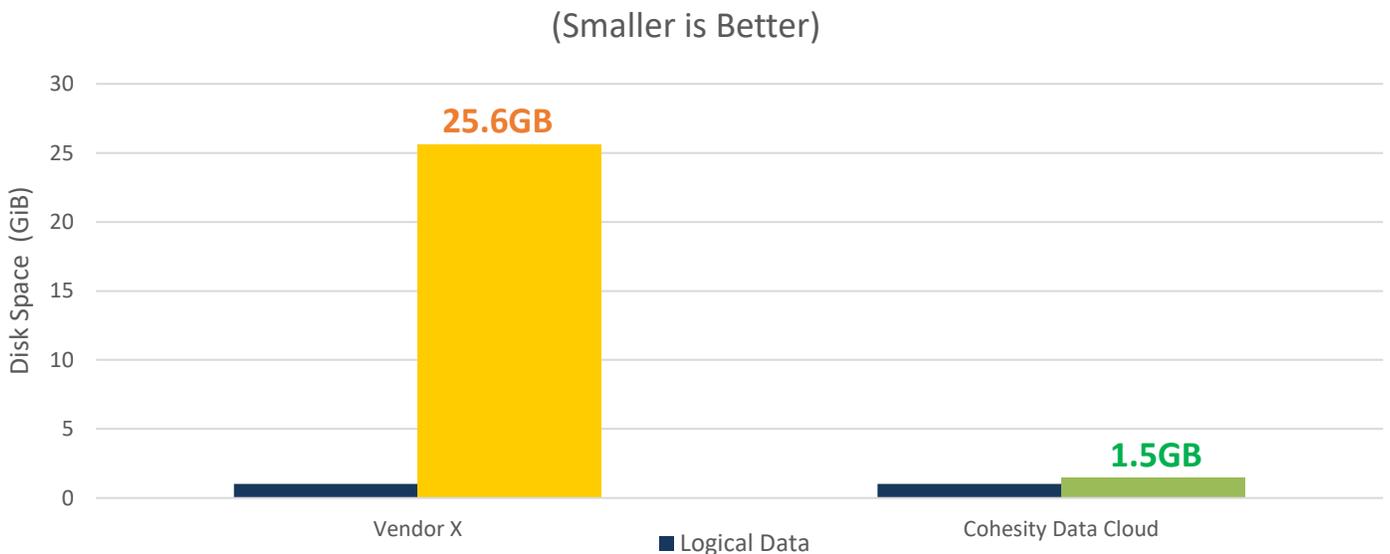
Figure 5. Cohesity Variable-length Deduplication Versus Fixed-length Deduplication



Source: Enterprise Strategy Group, a division of TechTarget, Inc.

Next, we looked at small file efficiency. We copied 1 million 1KB files to a view on the Cohesity data platform and a view on the Vendor X system. To isolate the effect of small file optimization, we disabled deduplication for this test and set EC redundancy again to 2:1 in both systems. Due to its 8KB block size and the requirement to store three copies of the data for redundancy, Vendor X inflated the data to 25.6GB. Cohesity was able to store the data on disk using only 1.5GB of capacity (see Figure 6). It is important to note that this test was designed to highlight the small file efficiency of the Cohesity platform and does not represent a real-world application by itself. Organizations that have a sizable number of small (sub 8KB) files should see some benefit, depending on the workload and how the files are written.

Figure 6. Small File Efficiency



Source: Enterprise Strategy Group, a division of TechTarget, Inc.

Why This Matters

Enterprise Strategy Group (ESG) research clearly shows that organizations today are still dealing with extremely fragmented data assets, but what are the ramifications? In 2019, organizations told ESG that 42% of administrators' day-to-day tasks—on average—concern managing their organization's secondary data, applications, and copies across on-premises and cloud environments.⁵ If 42% of a typical IT admin's job is managing fragmented data, then 42% of their cost is ultimately wasted in non-productive, non-profitable endeavors. Worse, 49% of respondents reported believing that MDF leads directly to overworked employees.⁶ There's every reason to believe these are still challenges today since data volume is growing and organizations are still managing multiple secondary copies of production data.

Since we know that employees seek intelligent, meaningful tasks as a condition of employment and job satisfaction, organizations are making their businesses undesirable places for the best talent to work by continuously demanding that skilled employees perform menial, mundane jobs.

Cohesity global, variable-length deduplication was able to provide nearly 60% data reduction in testing audited by ESG, compared with just 6.4% reduction for the tested non-global deduplication system. Cohesity also clearly demonstrated a significant small file efficiency advantage. Cohesity was able to store one million 1KB files on disk using only 1.5GB of capacity, whereas the alternative system tested increased the footprint of the files on disk to more than 25GB.

These factors, combined with compression, enable the majority of Cohesity customers to achieve a 96x or higher data reduction rate. This translates to easier integration into an organization's existing environment and substantial savings in disk capacity, network bandwidth, and administration.

⁵ Source: Enterprise Strategy Group Research Insight Paper, [Mass Data Fragmentation Is Quietly Killing Digital Transformation Efforts](#), May 2019.

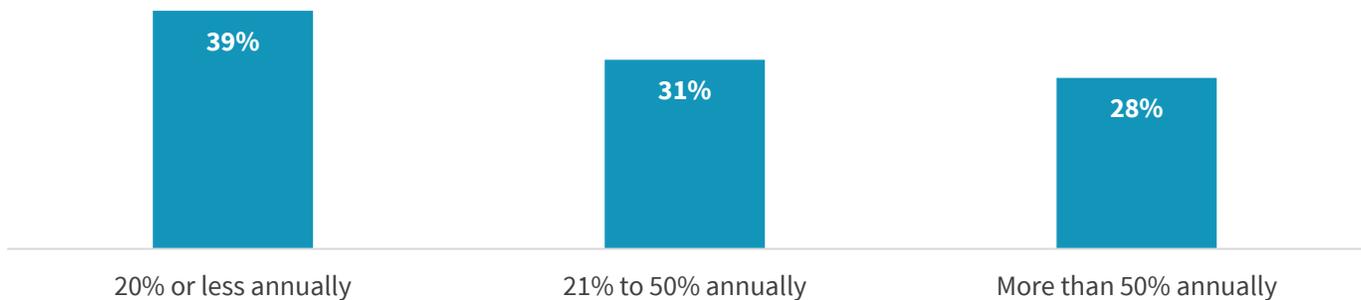
⁶ Ibid.

The Bigger Truth

Data growth is the root cause of all data fragmentation, but it is also a universal truth among digitally enabled organizations. Among organizations surveyed by Enterprise Strategy Group (ESG), growth in storage capacity is significant: 59% of organizations believe their total storage is increasing by more than 20% annually (see Figure 7). Interestingly, when asked how much secondary storage they needed per 1 TB of primary storage, the majority of organizations reported that they needed 4 TB or more, for both data protection (58%) and non/protection/non-production purposes (51%).⁷ It's easy to conclude that organizations with the biggest data management jobs to do today are going to be under the greatest data management pressure as time passes.

Figure 7. Total Annual Volume of Data Growth

At approximately what rate do you believe your organization's total volume of data is growing annually? (Percent of respondents, N=360)



Source: Enterprise Strategy Group, a division of TechTarget, Inc.

Cohesity has created a way to eliminate not only costly redundant data copies, but also the redundant infrastructure silos organizations buy and manage to act on that data. Cohesity's Data Cloud is a single, highly scalable, intelligent solution that first protects data and then uses the copy an organization already has for production and to create snapshots for secondary uses. Many organizations are still relying on legacy integrated backup appliances, deduplication target appliances, and copy data management solutions. The question organizations should be asking is: Why maintain multiple copies of data on disparate hardware and software platforms when all can be retained in a single, scalable solution that not only stores data more efficiently, but allows organizations to repurpose data for compliance and analytics use cases?

ESG validated that Cohesity Data Cloud can house all these copies of an organization's data, deduplicating across views, applications, and workloads, with a highly efficient mechanism for storing small files. The platform is expandable and flexible to meet the needs of each customer. The ability to scale incrementally and seamlessly means the environment can grow organically as needed, with predictable cost and performance. It offers greater insight into data, with simpler management.

ESG also validated that the majority of Cohesity customers see a 96x or higher data reduction rate, based on an audit of Cohesity's call home data, confirmed by interviews with Cohesity customers.

Organizations need to consider how to store, protect, access, govern, and retire data based on each data operational area—and stop just assuming “someone will do it.” ESG believes that a more intelligent plan can save incalculable time and money and lead to much better knowledge worker productivity, all of which mean a better bottom line. If your organization is ready to address the challenges of data sprawl and redundancy, a look at Cohesity's data management platform would be a smart first step.

⁷ Source: ESG Complete Survey Results, [From Data Backup to Data Intelligence](#), January 2022.

All product names, logos, brands, and trademarks are the property of their respective owners. Information contained in this publication has been obtained by sources TechTarget, Inc. considers to be reliable but is not warranted by TechTarget, Inc. This publication may contain opinions of TechTarget, Inc., which are subject to change. This publication may include forecasts, projections, and other predictive statements that represent TechTarget, Inc.'s assumptions and expectations in light of currently available information. These forecasts are based on industry trends and involve variables and uncertainties. Consequently, TechTarget, Inc. makes no warranty as to the accuracy of specific forecasts, projections or predictive statements contained herein.

This publication is copyrighted by TechTarget, Inc. Any reproduction or redistribution of this publication, in whole or in part, whether in hard-copy format, electronically, or otherwise to persons not authorized to receive it, without the express consent of TechTarget, Inc., is in violation of U.S. copyright law and will be subject to an action for civil damages and, if applicable, criminal prosecution. Should you have any questions, please contact Client Relations at cr@esg-global.com.



Enterprise Strategy Group is an integrated technology analysis, research, and strategy firm that provides market intelligence, actionable insight, and go-to-market content services to the global IT community.