

Technical Review

ExaGrid with Commvault: Maximum Deduplication Savings with Easy Management

Date: January 2022 **Author:** Kerry Dolan, Sr. IT Validation Analyst; and Craig Ledo, IT Validation Analyst

Abstract

This report documents ESG's validation of ExaGrid testing that demonstrated both the capacity savings and ease of use available from a combined ExaGrid and Commvault backup solution.

The Challenges

Storing and protecting continually growing amounts of data increase stress on both infrastructure and IT staff. According to recent ESG research, more than half (53%) of survey respondents reported having more than 500 TB of backup data (with 11% reporting having 10 PB or more), and 67% reported that backup data volumes were growing more than 20% annually (with 32% reporting that they were growing more than 50% annually).¹ Protecting this data is a critical assignment in today's world in which organizations depend on highly available applications and data, but backing up that data strains IT budgets.

In addition, while technology innovations are transforming IT in positive ways, organizations continue to struggle with IT complexity. According to recent ESG research, 56% of survey respondents reported that their IT environments are more complex than two years ago.² Further complicating IT is the trend toward having more generalists on the IT staff handling a range of tasks, rather than experts dedicated to specific areas like data protection.

Figure 1. IT and Data Protection Challenges



53%
report having
more than 500 TB of
backup data



67%
report backup
data volumes growing more
than 20% annually



56%
report that their
IT environments are more
complex than 2 years ago

Source: Enterprise Strategy Group

Together, ExaGrid and Commvault can increase deduplication to reduce storage capacity needs and costs, without adding complexity to your data protection processes.

¹ Source: ESG Research Report, [Tape's Place in an Increasingly Cloud-based IT Landscape](#), January 2021.

² Source: ESG Research Report, [2022 Technology Spending Intentions Survey](#), November 2021.

The Solution: ExaGrid Storage with Commvault Backup Software

Individually, Commvault and ExaGrid each offer data deduplication, but together they offer a highly efficient backup solution that can reduce the storage footprint by up to 300%, saving storage capacity and cost. This reduction also ensures that minimal WAN bandwidth is required for remote replication. No operational changes are needed to the Commvault configuration, making it a simple task for administrators to add an ExaGrid target to an existing Commvault environment. While a complete review of all ExaGrid and Commvault features is beyond the scope of this paper, a brief description is provided below.

ExaGrid Tiered Backup Storage is designed specifically for data protection. A key feature is its Landing Zone tier that ensures fast backups and restores. Data sent to ExaGrid goes to the Landing Zone before being deduplicated and stored in the repository; this eliminates the storage bottleneck of typical inline deduplication and minimizes the backup window. It also enables fast restores since no deduplicated data rehydration is needed.

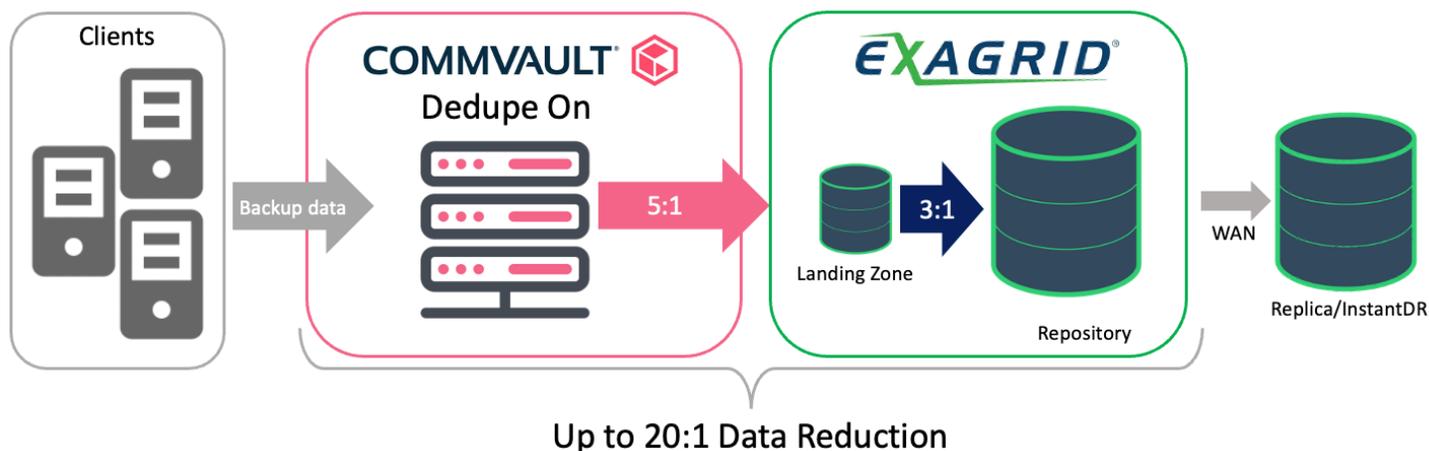
The ExaGrid repository tier provides global deduplication, long-term data retention, and replication to additional ExaGrid appliances over the WAN for disaster recovery. It is a scale-out system that supports different capacity-sized models and uses RAID6 protection with hot-swappable components. Other features include:

- Adaptive deduplication, enabling deduplication and replication during the backup window.
- A non-network-facing tier plus delayed deletes and immutable data objects allowing for ransomware recovery.
- Data encryption at rest and during WAN replication.
- Integration with Commvault dedupe on.
- Support for heterogeneous backup application environments.

Commvault has long been a leading provider of backup and data protection solutions. Commvault's single platform provides enterprise-class protection and recovery of on-premises and cloud-based physical and virtual files, applications, and databases. Features include:

- High performance backups.
- Fast, granular recovery.
- End-to-end encryption.
- Built-in ransomware protection.
- Flexible copy data management that enables multiple uses of backup data.
- Global deduplication.

Figure 2. ExaGrid and Commvault Deduplication



Source: Enterprise Strategy Group

ESG Tested

ESG validated two key features of the joint ExaGrid/Commvault solution: 1) the two tiers of deduplication that the joint solution delivers, and 2) the ease of integrating and managing ExaGrid with Commvault.

ExaGrid + Commvault = Enhanced Deduplication

ESG first reviewed ExaGrid testing that demonstrated how a combined ExaGrid/Commvault solution can increase data deduplication rates, saving customers money on storage capacity. Commvault deduplicates backups inline and sends them to the ExaGrid appliance. The deduplicated data first hits the ExaGrid landing zone and is then further deduplicated before being stored on the ExaGrid non-network facing repository.

It should be noted that this testing was designed to demonstrate the additional deduplication on the ExaGrid target. It emulates a real-world scenario but does not use the large amount of data that customers typically would be backing up; these large data volumes by nature increase the deduplication ratio.

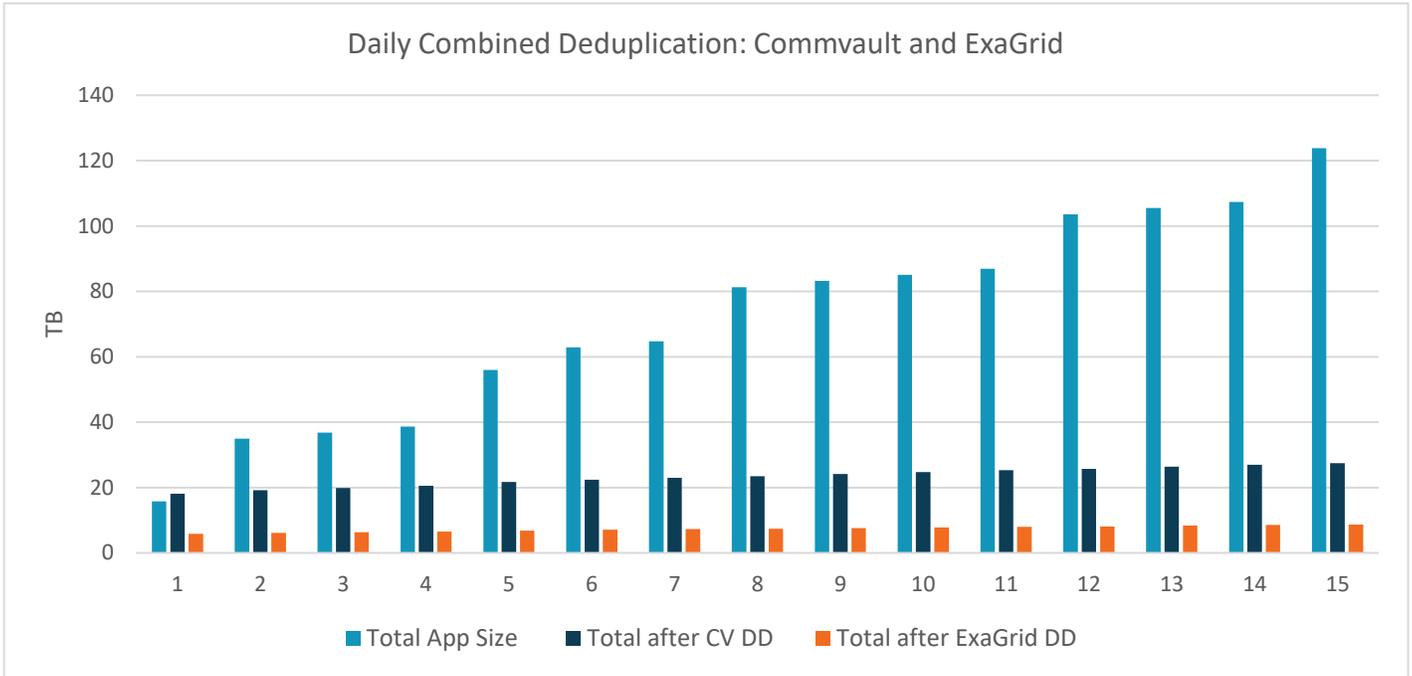
First, we reviewed the test bed setup and data-generating tool.

- *Data set.* A 15.5TB data set was created using ExaGrid's FileMod data generation tool. The data was 10 TB of file system data; five Windows Server 2019 VMware virtual machines (VMs), each with 1 TB of file system data; and a 500GB SQL Server database. File sizes ranged from 4KB to 500MB, with an average of 10MB, and were spread across more than 110K directories. Directories were assigned a seed value to ensure that successive backups included unique data.
- *Backup process.* For the file system and VM data, the testing ran five full backups with four to six incremental backups in between. All the SQL backups were full. Backups were run twice a day to simulate—but accelerate—a typical work week, with data changing 1% for file systems, and 0.025% for SQL, with each backup. Data changes included growing, shrinking, changing, renaming, deleting, reordering, etc.
- *Backup resources.* Backups were executed using Commvault CommServer version 11.24.7 and were sent to an ExaGrid EX84 appliance via the CIFS/Samba protocol. The ExaGrid Landing Zone was sized for 84TB.

The chart in Figure 3 shows the results of the 15 days of testing. Over that time, the complete application data set grew, while both Commvault and ExaGrid deduplication reduced the amount of data on disk.

- Over the course of 15 days, the 15.5TB had five full backups, plus about 25 incrementals in between the fulls, for a total data set size of about 123.76 TB.
- Of that 123.76 TB, Commvault delivered only 27.25 TB to ExaGrid, about a 4.5:1 dedupe ratio.
- ExaGrid then deduped that 27.25 TB down to 8.66 TB, an additional 3:1 ratio.
- The total combined deduplication was more than 14:1.

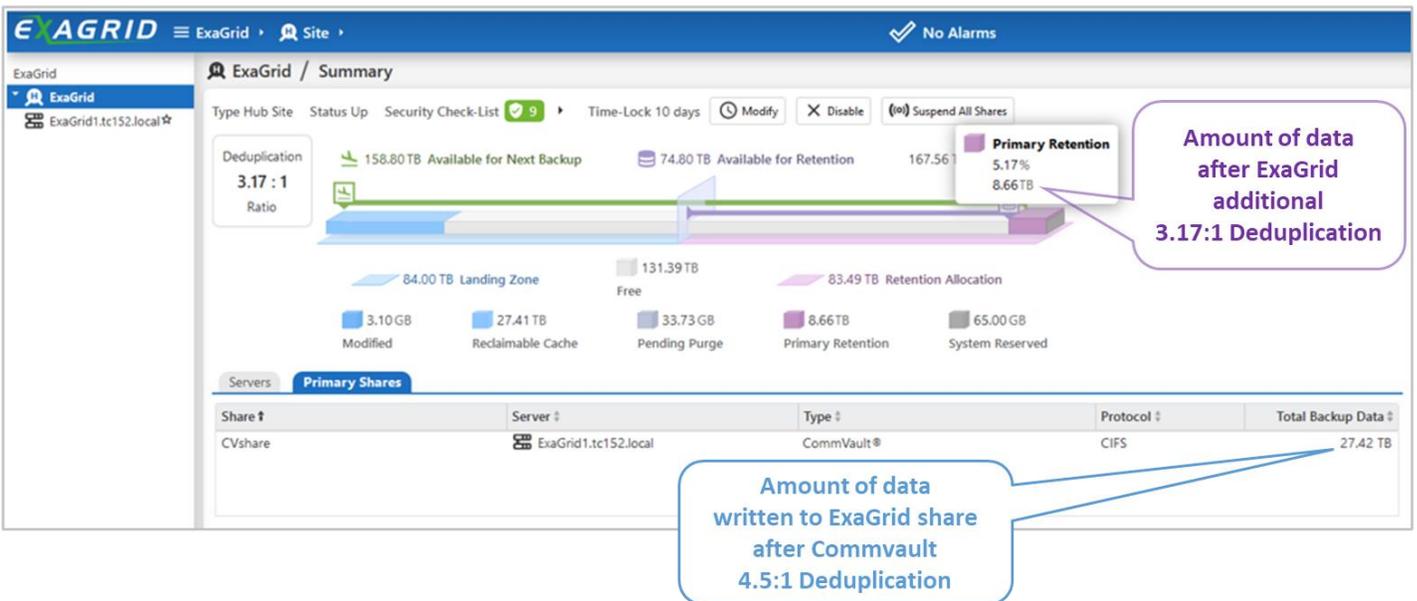
Figure 3. Data Reduction by Commvault and ExaGrid



Source: Enterprise Strategy Group

Next, we logged into the ExaGrid appliance to view the final day details from the ExaGrid GUI.³ Figure 4 shows the details of total deduplication.

Figure 4. Total Combined Deduplication



Source: Enterprise Strategy Group

³ Note that the screenshot in Figure 4 was taken just as the final backups were completing. As a result, the total data stored on ExaGrid shows as 8.63TB, while the actual final data amount was 8.66 TB.

Adaptive Deduplication

ExaGrid has a feature called adaptive deduplication that enables ExaGrid to start deduplication only when bandwidth usage for other processes is low. This enables ExaGrid to promptly get the back-end, additional deduplication accomplished without slowing down any backup or restore processes. ExaGrid also offers graphs showing write rates, modified bytes, deduplication rates, and replication queue over time, making it easy for administrators to understand how data is being managed. Administrators can zoom in to view specific dates or zoom out for an overview. Figure 5 shows that ExaGrid adaptive deduplication in purple increased as the write rate of the Commvault backup in green began to slow.

Figure 5. ExaGrid Adaptive Deduplication



Source: Enterprise Strategy Group

Why This Matters

Storing backup data is critical for ensuring maximum business productivity, but continual data growth strains budgets. Organizations can save money by deduplicating backup data to reduce the amount of storage they need.

ESG validated that the combined ExaGrid/Commvault solution reduced 123 TB of data down to 8.66 TB, reducing storage capacity requirements by 14:1. It should be noted that the ExaGrid test setup was conservative, and the methodology delivered deterministically random data. Customers in real-world production environments are likely to see even higher deduplication rates with this joint solution.

Ease of Deploying and Managing Commvault and ExaGrid

Next, we viewed the ease of deploying and managing backups using Commvault and ExaGrid.

Adding ExaGrid to an existing Commvault environment is as simple as adding a new Commvault Library and Storage Policy for ExaGrid and selecting them for use. This is a quick and easy task with which Commvault administrators are familiar. ESG viewed a demo of the one-time initial deployment, which is also simple, involving a few tasks on the ExaGrid side and then the Commvault side. Then, we viewed a demo of the ExaGrid share creation process.

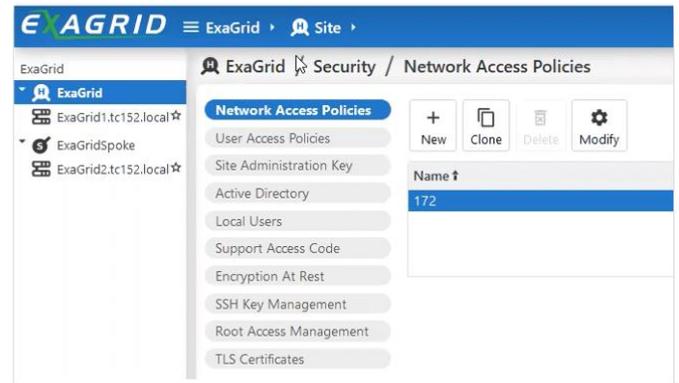
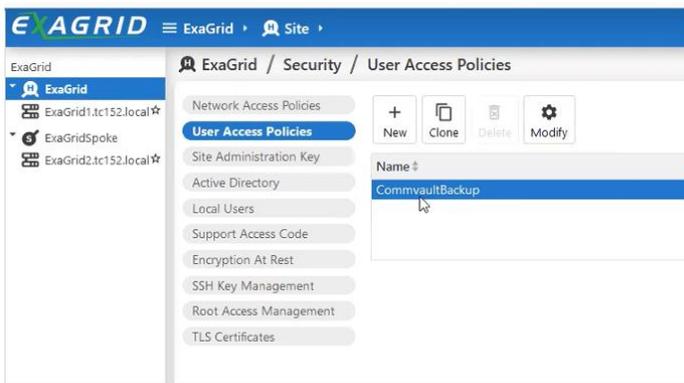
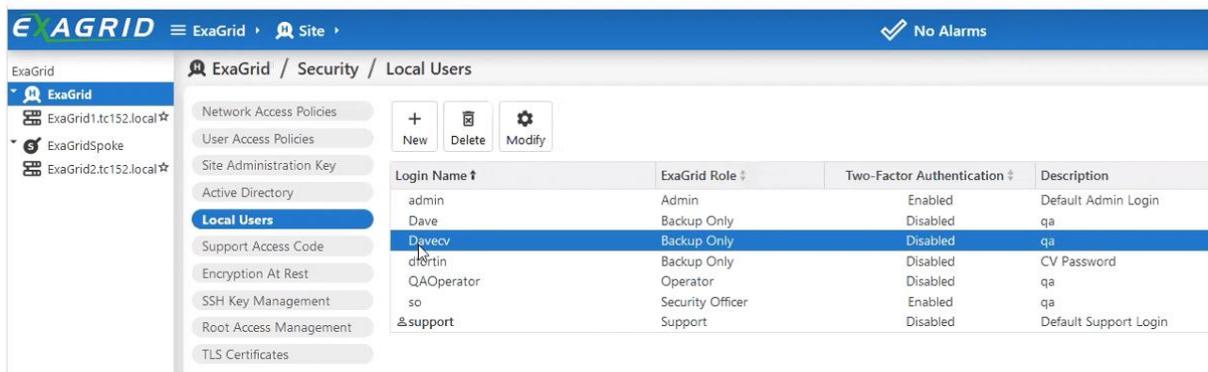
Initial Deployment

ExaGrid Tasks

After logging into the ExaGrid EX84, we added users and created access policies before creating an ExaGrid share for backups.

- From the **Security/Local Users** tab, we added a user called *DaveCV* and assigned backup-only privileges.
- Next, we clicked on **Security/User Access Policies** and created a new policy called *CommvaultBackup*.
 - With a click of the **Modify** button, we added *DaveCV* to that policy; users can be added singly or in groups. (While it is not required to assign users to shares, it is a good security practice.)
- Next, we created a **Network Access Policy** with an open 172 IP address. These policies can assign specific hosts, IP addresses, and ranges of IP addresses.

Figure 6. Creating ExaGrid Access



Source: Enterprise Strategy Group

After these policies were set up, we created a share on the ExaGrid EX84.⁴ From the **Shares and Replicas** tab we clicked **+New Share**, named it *CVshare2*, and selected *Commvault* as the type; the tight integration between Commvault and ExaGrid ensures that data sent to Commvault shares is optimized by ExaGrid features. Next, we selected the CIFS/SMB protocol and the Network Access Policy and *CommvaultBackup* User Access Policy created previously. A Delete Monitor is also available, which will alert the administrator in case a selected percentage of the share is deleted within 24 hours as a protection against ransomware or other unauthorized data deletion.

Next, we added an ExaGrid EX40000E appliance to serve as a remote *ExaGridSpoke* and initiated replication from the first appliance. The Commvault backup data was then backed up to the ExaGrid Landing Zone, deduplicated and stored on the ExaGrid share, and then replicated to the remote site. We also enabled ExaGrid InstantDR on the remote server so that data could be exposed using the same User Access Policy.

Figure 7. Creating ExaGrid Primary and InstantDR Shares

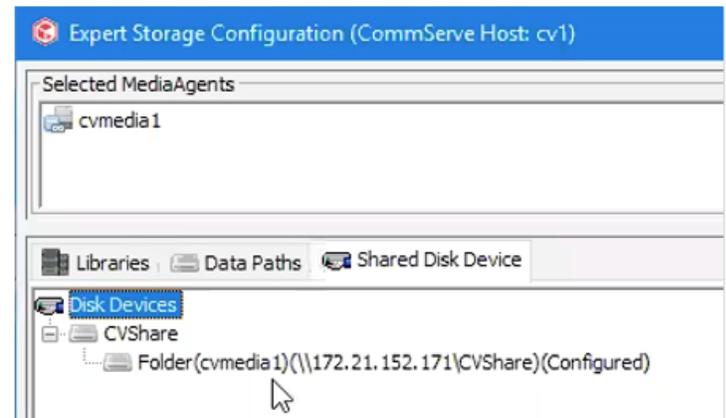
Source: Enterprise Strategy Group

⁴ Note that the share creation screenshots were complete after the deduplication testing. The new share we created was *CVshare2*, while the actual share used in the testing was *CVshare*.

Commvault Tasks

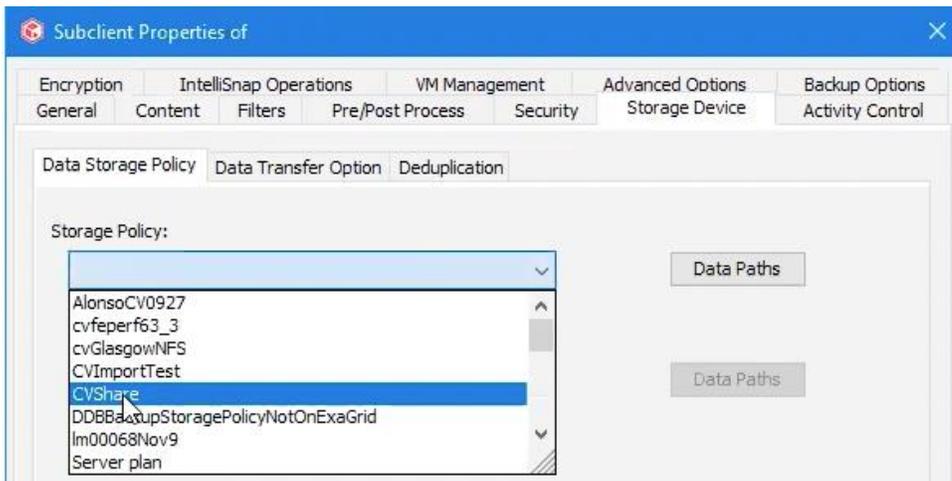
On the Commvault side, we created a Library and Storage Policy that together tell Commvault how to access the ExaGrid share.

First, we created the Commvault Library. From the Commvault CommCell Console, we clicked on **Libraries/Storage/Expert Storage Configuration** and added the previously created *cvmedia1* media server to the **CVShare** library to access the share. Next, we clicked **Shared Disk Device** and added the path and a base folder for backups to the ExaGrid share.



Next, we created the Commvault Storage Policy. We scrolled through the Commvault directory of hosts and clusters to find the file data, VMware VMs, and SQL database that we wanted to backup. From the **Storage Policies** tab, we selected *CVShare*, clicked on **Properties**, and added the 24 subclients containing that data. Next, we added retention time of 90 days and enabled creation of a media server deduplication database to track application size and data reduction. From the **Properties** menu, we could add content as well as filters, pre- and post-process tasks, security, and other properties. From this menu, we clicked **Storage Device/Data Storage Policy** and selected *CVShare*. Choosing this storage policy, which contains the library that we created to write to the ExaGrid share, is what linked our selected backup data set to the target. For optimal results, we turned off compression and

enabled deduplication on the Commvault media server. We disabled encryption in Commvault since ExaGrid does disk-level encryption. Finally, we created a schedule to execute backups every 12 hours.



Why This Matters

As IT complexity increases, organizations are looking carefully at new technology solutions. Optimal technology innovations consider the need for simplicity and ease of use. Complexity breeds inefficiency and cost, while simplicity ensures faster time to value.

ESG validated how easy and fast it was to set up an initial ExaGrid/Commvault deployment to securely back up data, replicate it, and make it available for instant restore. The ExaGrid tasks were simple to complete using the intuitive GUI, and the Commvault tasks are familiar to Commvault administrators.

The Bigger Truth

Increasing efficiency never goes out of style for obvious reasons: saving money is a key business objective. According to ESG's latest Technology Spending Intentions Survey, becoming more operationally efficient remains the most-cited objective for organizations' digital transformation efforts, as it has been for the past three years.⁵

It is well known that both ExaGrid and Commvault offer backup solutions that are easy to use and can help you to store data more efficiently with deduplication. What may be less well known is that a combined ExaGrid and Commvault solution can achieve even greater deduplication: up to 20:1 in many cases. A 20x reduction in backup storage costs can be a significant boost to any budget.

ESG validated that:

- The combined ExaGrid/Commvault solution reduced a 123TB backup data set that included 15 days of file data, VMware VMs, and SQL database backups down to 8.66TB, a 14:1 reduction, in a realistic yet intentionally conservative test environment.
- Deploying and managing the combined solution was simple, intuitive, and fast.

We also reviewed an ExaGrid TCO calculator that shows significant savings using Commvault and ExaGrid over Commvault with straight disk.

Any solution should be tested and planned with your specific needs and objectives in mind. However, if you are looking to spend less on backup storage without adding complexity to your data protection scheme, ESG recommends looking closely at the combined ExaGrid and Commvault solution.

All trademark names are property of their respective companies. Information contained in this publication has been obtained by sources The Enterprise Strategy Group (ESG) considers to be reliable but is not warranted by ESG. This publication may contain opinions of ESG, which are subject to change. This publication is copyrighted by The Enterprise Strategy Group, 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 The Enterprise Strategy Group, 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 ESG Client Relations at 508.482.0188.

The goal of ESG Validation reports is to educate IT professionals about information technology solutions for companies of all types and sizes. ESG Validation reports are not meant to replace the evaluation process that should be conducted before making purchasing decisions, but rather to provide insight into these emerging technologies. Our objectives are to explore some of the more valuable features and functions of IT solutions, show how they can be used to solve real customer problems, and identify any areas needing improvement. The ESG Validation Team's expert third-party perspective is based on our own hands-on testing as well as on interviews with customers who use these products in production environments.

⁵ Source: ESG Research Report: [2022 Technology Spending Intentions Survey](#), November 2021.