

Lab Validation Report

CA ARCserve Backup

Enterprise-Class Data Protection and Deduplication™

By Brian Garrett, Tony Palmer, and Lauren Whitehouse

November 2009

Contents

Introduction.....	3
Background.....	3
CA ARCserve Backup	4
ESG Lab Validation	5
Getting Started	5
Capacity Savings over Time	8
Performance	11
Media Assure	13
Tape Integration	14
Storage Resource Reporting.....	16
ESG Lab Validation Highlights.....	17
Issues to Consider	17
The Bigger Truth	18
Appendix	19

ESG Lab Reports

The goal of ESG Lab reports is to educate IT professionals about emerging technologies and products in the storage, data management and information security industries. ESG Lab 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 objective is to go over some of the more valuable feature/functions of products, show how they can be used to solve real customer problems and identify any areas needing improvement. ESG Lab'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. This ESG Lab report was sponsored by CA, Inc.

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 from time to time. 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.

Introduction

Now, more than ever, IT is being called on to drive efficiencies, manage risk, and maintain application and data availability. Exponential data growth, compliance requirements, and service level agreements are only a few of the challenges facing organizations with data centers and remote offices requiring data protection. Compounding such an array of challenges, capital and operational expenses of these same organizations are under scrutiny. CA ARCserve Backup software is a client/server software solution providing backup/recovery, archive, high level graphical reporting on data protection and environment resources, and disaster recovery (DR). This ESG Lab Validation focuses on the key functionality in CA ARCserve backup driving greater efficiency: embedded data deduplication, backup and recovery performance, Media Assure data verification, and tape integration.

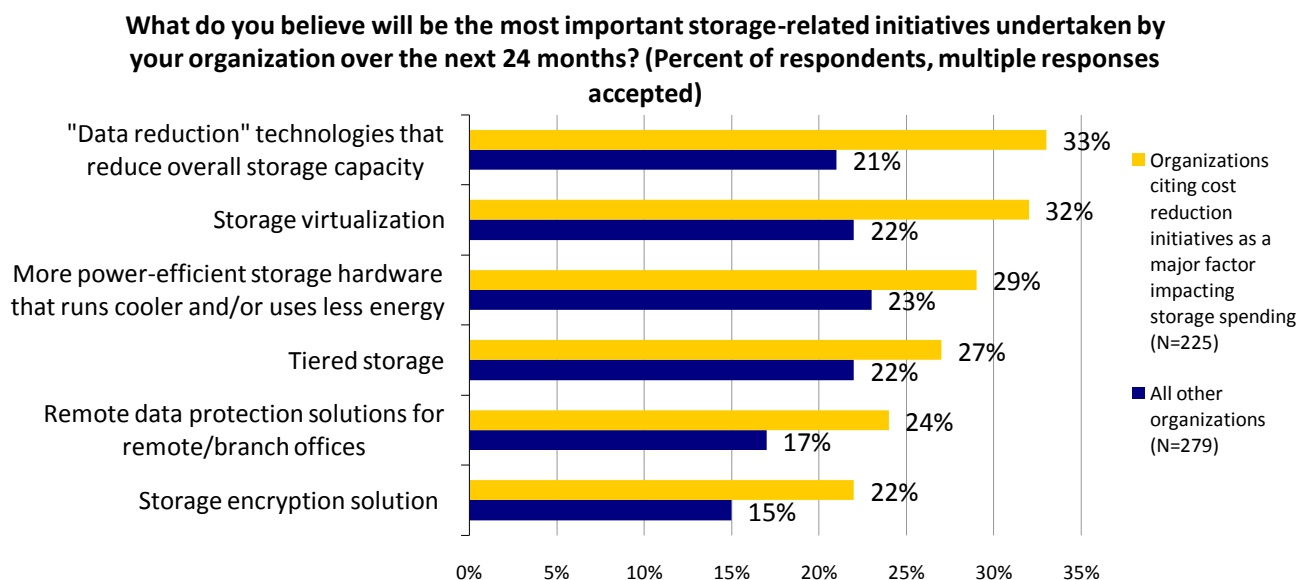
Background

Relentless information growth is necessitating greater investments in IT infrastructure. Data protection processes, such as backup, compound capacity growth issues since multiple copies of primary data are made for operational and disaster recovery. In fact, ESG research respondents cited data protection as *the* application that will be most responsible for storage growth over the next 24 months.¹

Data growth has created a challenge for IT organizations: constraining IT's ability to make copies of data for operational and disaster recovery (DR) purposes within a prescribed window of time. Inserting disk into the backup data path to accelerate backup performance and improve reliability addresses backup window issues; ESG research shows that nearly 70% of organizations back up to disk—either as an intermediary stop on the way to tape or as the final resting place for data.² Justifying investments in disk for backup is, however, difficult in a tough economy.

With the majority of IT budgets flat or down, cost reduction initiatives are viewed by more than half of ESG survey respondents as the most important factor in deciding what, if any, IT investments are to be made in 2009. This may be one of the reasons why cost-conscious organizations cited data reduction technologies as a top storage-related initiative over the next 24 months (see Figure 1).³ Optimizing backup capacity extends savings beyond the storage footprint, creating efficiencies in power, cooling, and data center floor space requirements. Additional savings can also be realized as organizations often end up managing less infrastructure.

Figure 1. Considerations in Deduplication Purchase Decisions



Source: Enterprise Strategy Group, 2008.

¹ Source: ESG Research Report, *Enterprise Storage Systems Survey*, November 2008.

² Source: ESG Research Report, *Data Protection Market Trends*, January 2008.

³ Source: ESG Research Report, *Enterprise Storage Systems Survey*, November 2008.

Deduplication identifies and eliminates redundancy, minimizing bandwidth and storage capacity requirements. In the backup process, deduplication ensures that only unique data is stored. Initially, data is backed up to the storage device and all subsequently written data is examined for redundancy, with new unique data being written to storage. When duplicate data is found, a pointer linked to the original unique piece of data is stored. Pointers consume significantly less space than multiple instances of the whole item.

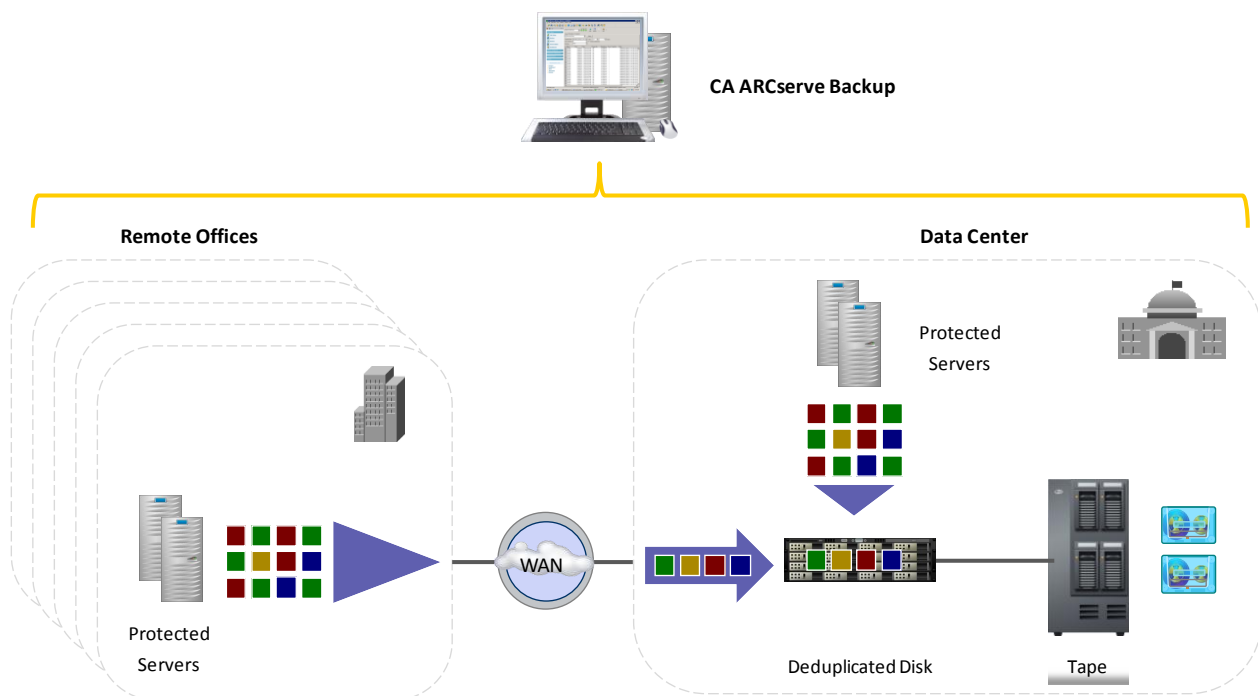
There are several factors to consider in the selection of deduplication technology: ESG research found that the cost of the solution was the most frequently cited factor (although the savings realized in data reduction can easily compensate for the investment costs). The survey data suggests that integration with existing backup processes, ease of deployment, and ease of use, as well as the impact on backup/recovery performance were important considerations—more so than technical nuances, such as deduplication approach (inline or post-process), or the deduplication ratio.⁴

CA ARCserve Backup

CA ARCserve Backup 12.5 is a client/server backup/recovery solution designed for a broad array of both physical and virtual server and application environments. With built-in features such as storage resource management (SRM), backup reporting, data deduplication, and integration with CA XOSoft replication, CA ARCserve Backup delivers cost- and management-efficient operational and disaster recovery for local and remote sites.

In its latest release, CA supports both VMware and Microsoft virtual server environments with full integration with VMware Consolidated Backup (VCB) in VI3 and vStorage API for data protection in vSphere 4. CA ARCserve Backup supports virtual machine-level backup and single-pass file-level recovery, mixed full image-level and incremental or differential backup, and automatic discovery of virtual machines.

Figure 2. CA ARCserve Backup



CA ARCserve Backup's integrated SRM, business intelligence dashboard, and reporting capabilities are designed to make IT operations more efficient in managing the overall data protection environment and to help reduce operational issues. Its ability to verify the recoverability of backed up data—including deduplicated backups—with its Media Assure feature can save valuable time in recovery scenarios.

⁴ Source: ESG Research Report, *Data Protection Market Trends*, January 2008.

Deduplication is another integrated feature in CA ARCserve Backup 12.5. CA ARCserve Backup uses block-level, variable-length deduplication at the media server—confining the deduplication processing load to the point in the backup data path that can be optimized for it. Data is selectively and intelligently deduplicated: CA ARCserve Backup skips deduplication processing for data it knows it's already seen and deduplication policy settings can be enabled or disabled for individual backup sets.

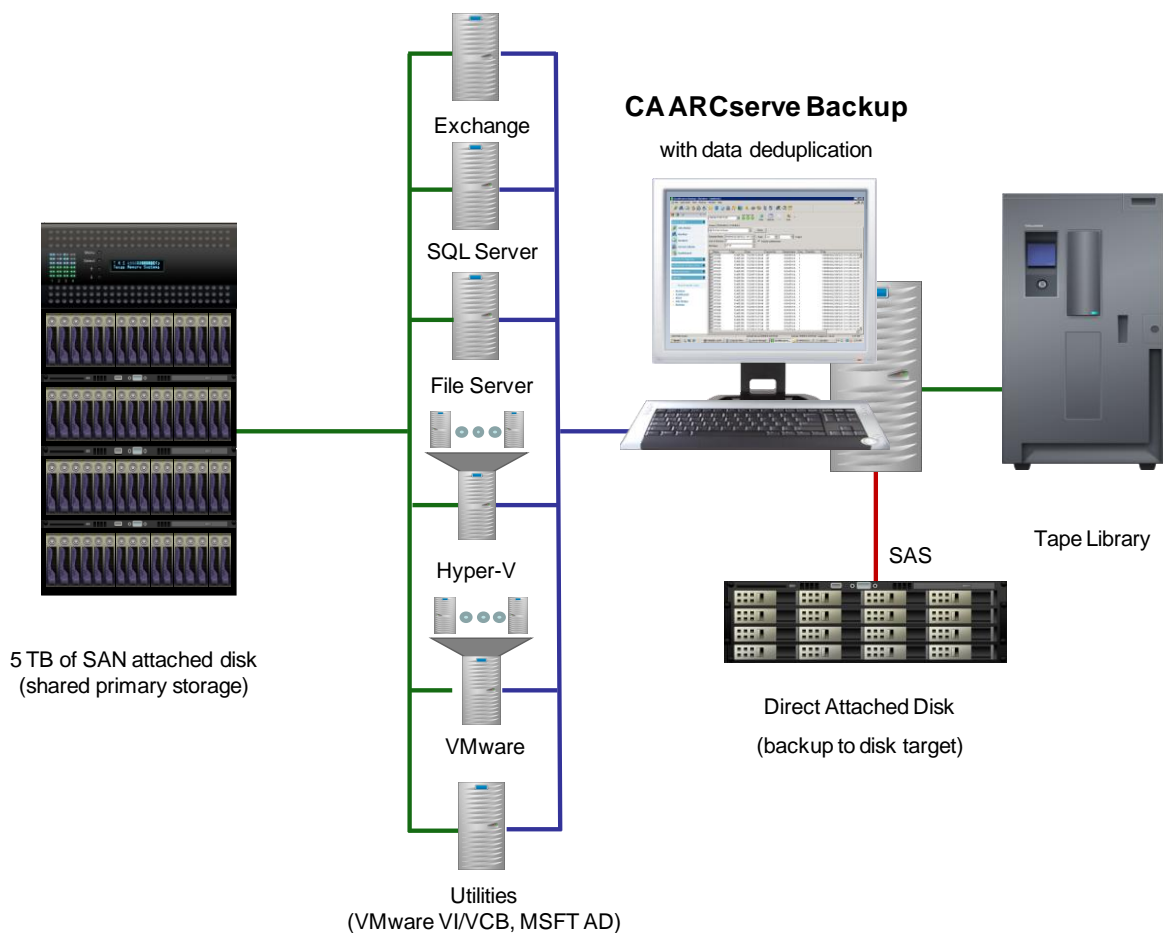
ESG Lab Validation

ESG Lab performed hands-on evaluation and testing of CA ARCserve Backup version 12.5 at CA, Inc.'s Islandia, NY facility. Testing was designed to validate CA ARCserve Backup's embedded data deduplication, backup and recovery performance, Media Assure data verification, tape integration, and how these features can be leveraged to drive greater efficiency in data protection.

Getting Started

ESG Lab testing was conducted on the test bed illustrated in Figure 3. Several servers were configured in various roles as would be commonly found in IT environments and pre-populated with data stored on a shared 5 TB SAN attached disk array. A CA ARCserve Backup server was configured with direct attached SAS disk and a Fibre Channel attached tape library for use as backup targets.

Figure 3. The ESG Lab Test Bed

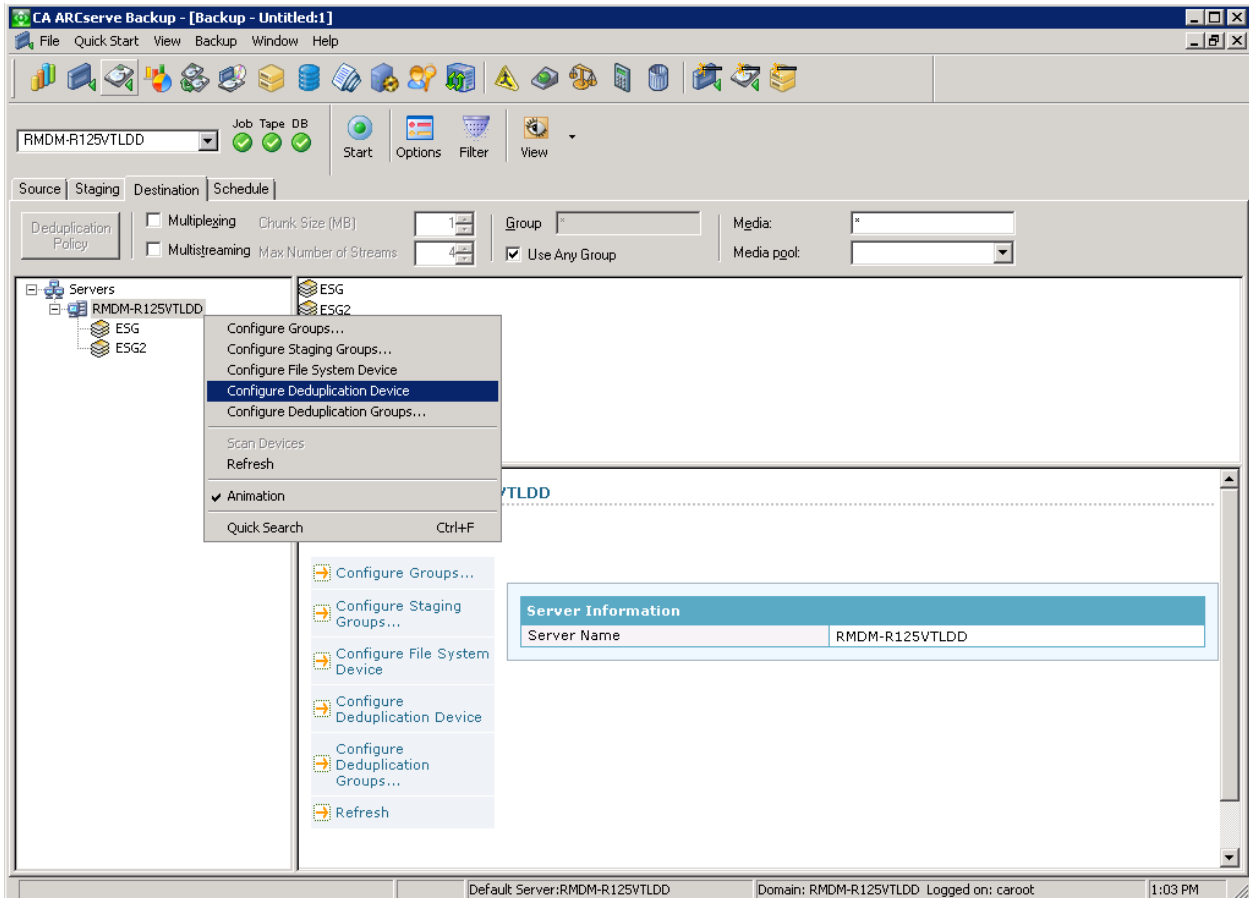


ESG Lab Testing

ESG Lab began by examining the CA ARCserve Backup management console. The user interface was intuitive and easy to use. A dashboard view provides an overview of the environment with the ability to drill down as needed by

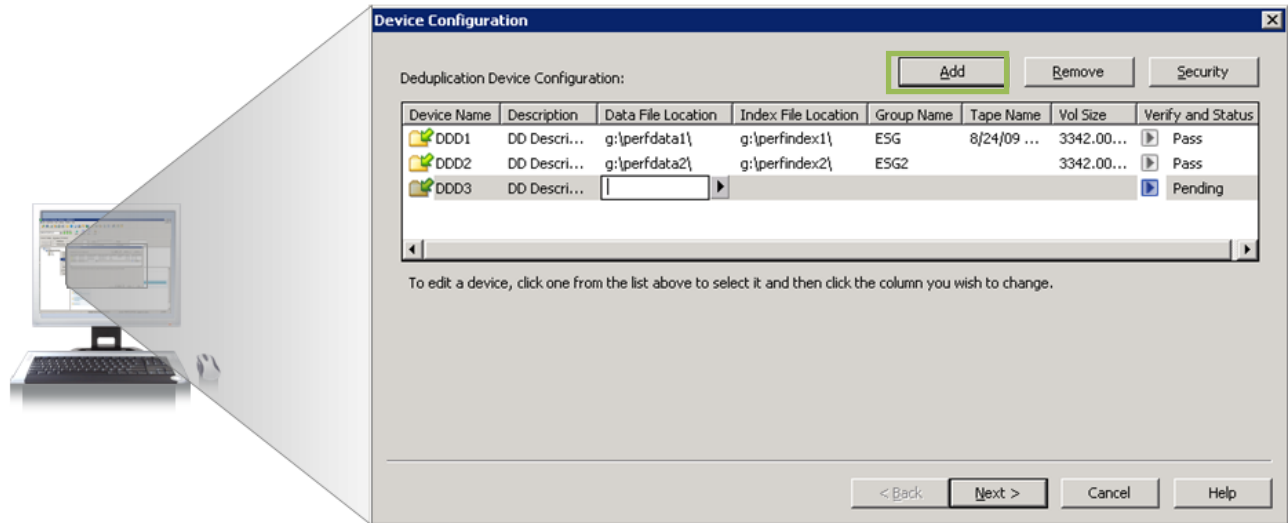
clicking on links. First, backup targets were created. When selecting a destination for a backup, a menu tree is presented, showing all servers in the CA ARCserve Backup environment and providing context sensitive, right-click management. As seen in Figure 4, ESG Lab right-clicked the server name and selected 'Configure Deduplication Device.'

Figure 4. Creating a Deduplication Device



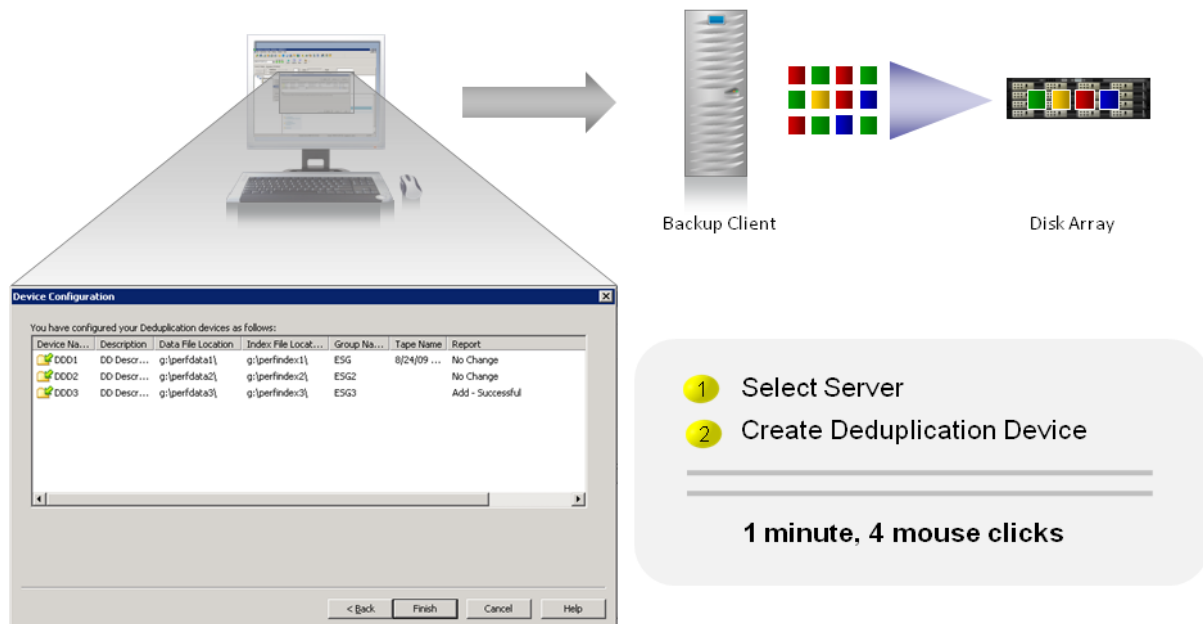
After clicking the 'Add' button in the Device Configuration dialog, the system prompted for a location to place the deduplication data and index files for this device. ESG Lab typed in g:\perfdata3\ and g:\perfindex3\ and then clicked 'Next.'

Figure 5. Specifying the Data File Location



After a few seconds, the wizard confirmed that the addition was successful, ESG Lab clicked 'Finish,' and the system was ready to perform backups with data deduplication.

Figure 6. Configuring Deduplication



The entire process of creating a deduplication device took four mouse clicks and less than one minute.

Why This Matters

ESG Research found that 31% of IT professionals surveyed indicated ease of deployment as the single most important factor in purchasing a deduplication solution.⁵ This is especially important for deployments in large, complex environments where backup policies span hundreds of servers and dozens of applications—stretching resources to the limit. ESG Lab has confirmed that CA ARCserve deduplication is extremely easy to configure and manage. Deduplication can be dropped into an existing environment and reducing backup capacity in less than a minute, by running a short configuration wizard.

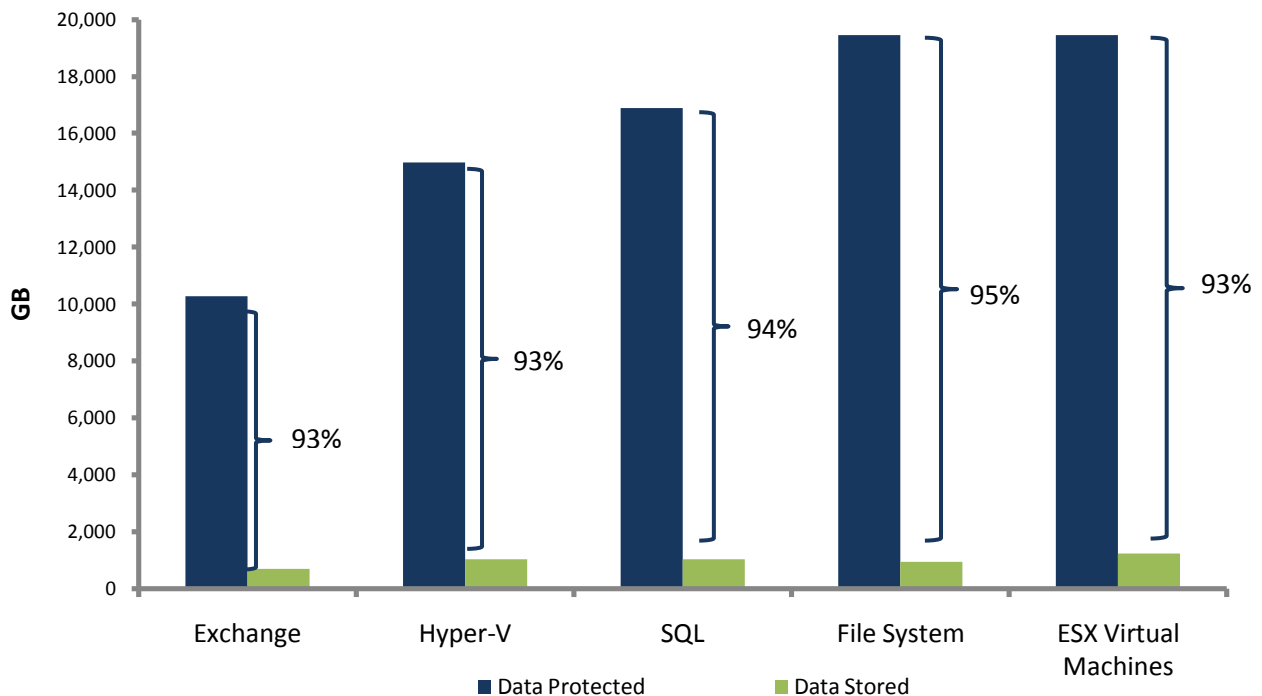
Capacity Savings over Time

CA ARCserve Backup deduplication reduces the capacity required to store backed up data. Data that has been backed up to a deduplication device is scanned and CA ARCserve Backup examines the backed-up data, recognizing and replacing redundant blocks with pointers to the original blocks. Because deduplication is integrated with the backup application, CA ARCserve Backup checks for files that have been backed up before and not changed. These files are not processed by the deduplication device at all, reducing processing overhead.

ESG Lab Testing

To evaluate deduplication in CA ARCserve Backup, ESG Lab audited multiple full and differential backups of several different data sets and validated the results.⁶ The datasets were partially harvested from a corporate data center—and partially generated using application specific tools. Data reduction was measured for the backups performed and projected out to 30 days using a five day per week daily full backup schedule. Figure 7 shows the data reduction achieved after 30 days broken out by data type.

Figure 7. Capacity Savings Over 30 Days by Data Type



⁵ Source: ESG Research Report, *Data Protection Market Trends*, 2008.

⁶ Full configuration details for the Backup clients and server configurations can be found in the Appendix.

Overall data reduction for full backups of the complete environment can be seen in Figure 8. The blue line traces the amount of capacity protected over the course of 30 days using daily full backups on a five day per week schedule. The solid blue line represents actual backups; the dotted line represents the ESG Lab projection. The yellow line represents the data stored on disk after deduplication. The overall data reduction after 30 days of full backups on a five day per week schedule was 95%.

Figure 8. Capacity Savings Over 30 Days Overall

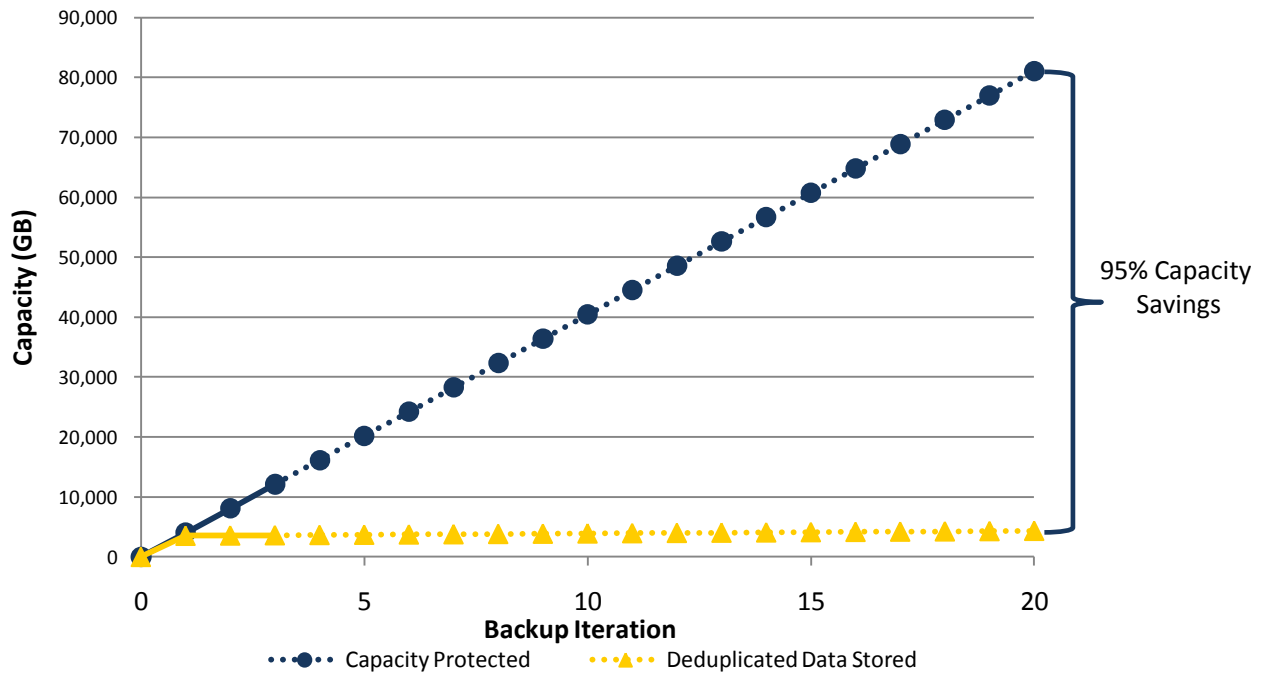


Table 1: Capacity Reduction Detail

Backup Iteration	Data Protected (GB)	Capacity Consumed: Deduplicated on Disk	Cumulative Deduplication Savings
1 st	4,037	3,519	13%
2 nd	8,074	3,562	56%
5 th	20,186	3,690	82%
10 th	40,436	3,905	90%
20 th	81,053	4,337	95%

What the Numbers Mean

- The dataset used for the first backup was manufactured to be more unique than data on real servers. Deduplication savings of 40% for first backups have been observed using real data sets on real servers.
- The total amount of data that CA ARCserve Backup would have stored on disk over the course of 30 days without using deduplication was calculated to be 81.05 TB.
- CA ARCserve Backup deduplication reduced the required disk capacity over 30 days to only 4.34 TB. This enables organizations to keep backups on disk for much longer retention periods, reducing the need to go back to tape for recoveries.

- The amount of capacity reduction that can be achieved with any deduplication solution will vary according to the backup policy in effect, the number of backups retained on disk, and the type of data being stored. In this scenario, capacity was reduced by 95% over 30 days.

Organizations have been doing their own file-level deduplication in the form of weekly full and daily incremental or differential backups for a very long time. These types of backup policies save space by simply not backing up files that have not changed since the last backup. The cost to users is much longer recovery times. To restore in a weekly full, daily differential scenario requires restoring from the weekly full, and in most cases, the latest daily differential backup to capture the latest saved versions of data. ESG Lab examined backups performed on a weekly full, daily differential schedule to determine additional capacity savings that could be achieved using deduplication.

Figure 9. Capacity Savings using Weekly Full, Daily Differential Backups over 60 days.

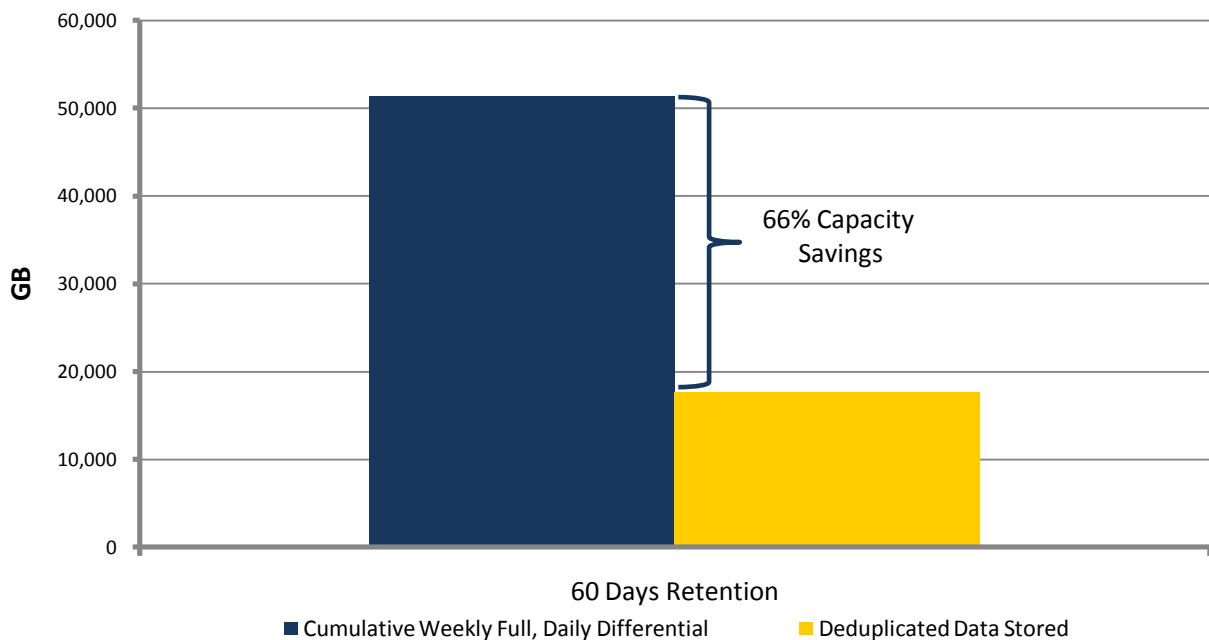


Figure 9 shows clearly that even with all duplicate files excluded using differential backups; CA ARCserve Backup deduplication was still able to achieve a 66% capacity savings over weekly full, daily differential backups alone.

Why This Matters

ESG research indicates⁷ that cost is the leading obstacle to disk-based backup deployments. Data reduction technologies change the economics of backup to disk by reducing the cost of the storage required to retain backups on disk.

ESG Lab has validated that CA ARCserve Backup can be used to significantly reduce disk capacity while enabling administrators to apply deduplication selectively to address data type and retention needs. CA ARCserve administrators can effectively provide high performance backup services, plus fast and reliable restores, using greatly reduced disk capacity. This lowers the cost per GB for backup data and enables companies to retain data exponentially longer for recovery purposes while minimizing the impact of deduplication on backup windows and recovery SLAs.

⁷ Source: ESG Research Report, *Data Protection Market Trends*, January 2008.

Performance

A tape drive can only perform one backup at a time. To get more than one backup job running at the same time in a tape environment, more tape drives need to be added and run in parallel. A backup and recovery solution incorporating many random access disk drives can run many backup jobs simultaneously. Much like the difference between a DVD and a VHS tape, the random access of disk also provides improved performance when locating individual files to be restored.

ESG Lab Testing

ESG Lab performed a series of backups using both raw file system devices and deduplication devices to examine the relative performance of backups and restores with and without deduplication. Full backups were performed against data sets commonly found in production IT environments. The data sets were partially harvested from a corporate data center and partially generated using application specific tools. Figure 10 shows the relative performance of deduplicated and raw file system backups.

Figure 10. Backup and Restore Performance, by Data Type

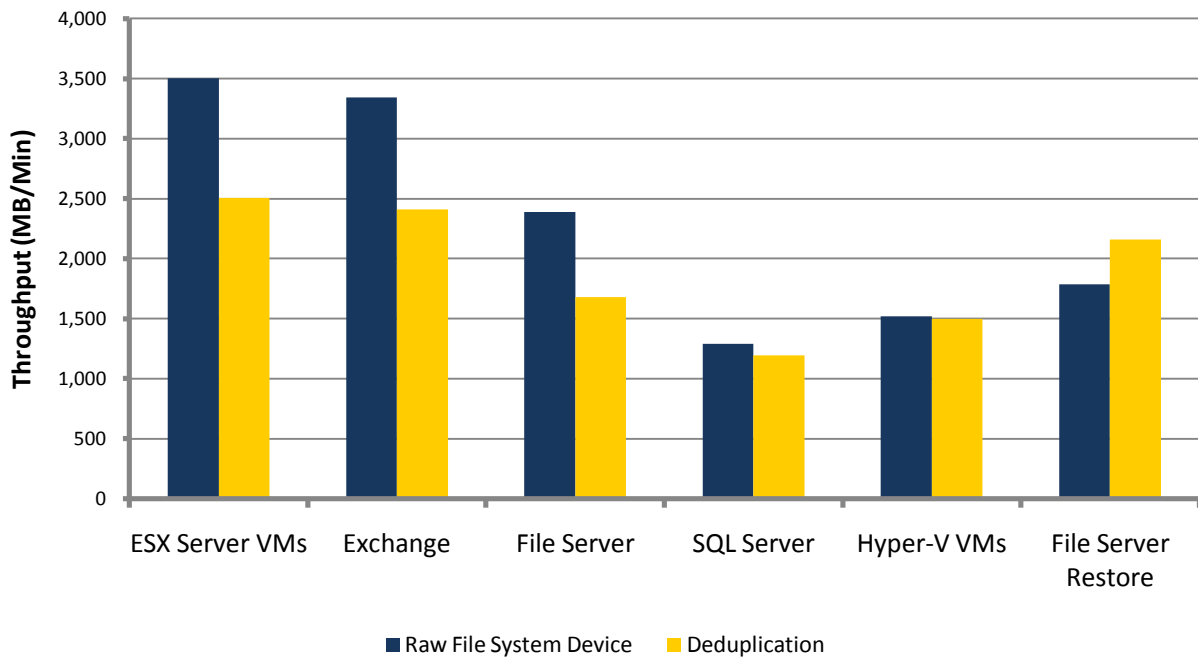
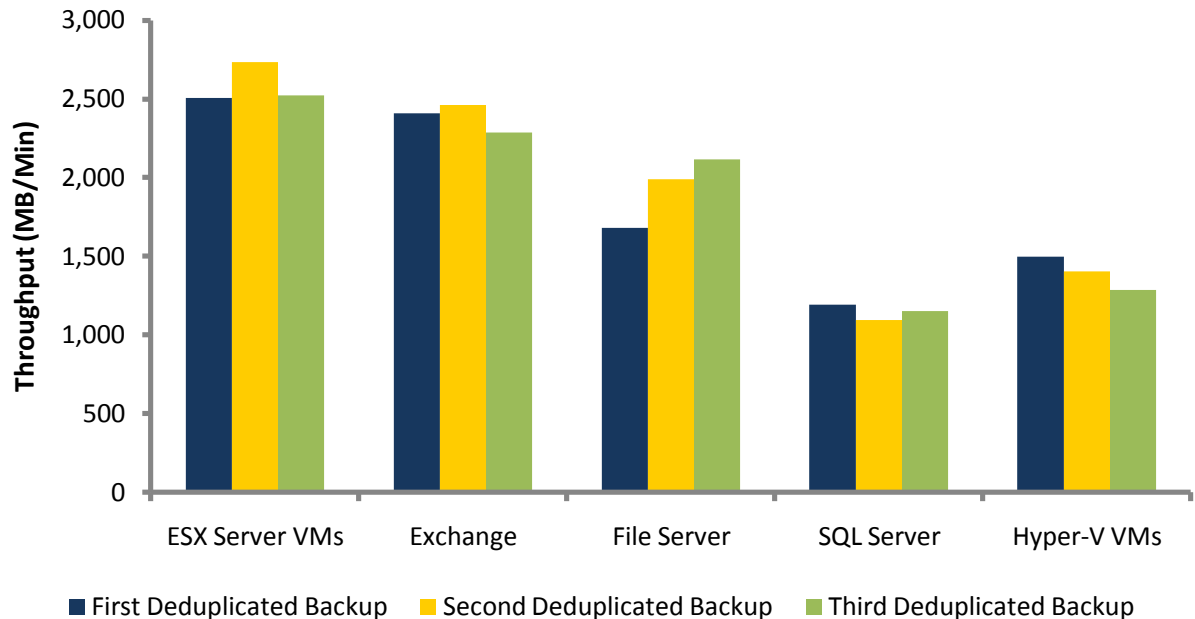


Table 2: Backup and Restore Performance Detail

Data Type	Data Set Size (GB)	Raw Throughput (MB/Minute)	Deduplicated Throughput (MB/Minute)
ESX Server VM Backup	973	3,501	2,504
Exchange Backup	554	3,341	2,408
File Server Backup	972	2,388	1,681
SQL Server Backup	837	1,289	1,193
Hyper-V VM Backup	749	1,517	1,497
Restore			
File Server Restore	5	1,786	2,159

Figure 11. Backup Performance by Data Type Over Multiple Backups



ESG Lab observed multiple deduplicated backups for each data type. Figure 11 shows that deduplication did not have a significant impact on backup performance over multiple backups. CA ARCserve Backup does not process files for deduplication that have not changed since the last backup, which reduces processing overhead and can increase performance over time.

What the Numbers Mean

- The impact of deduplication in our tests using extremely unique manufactured data on a single first backup ranged from a negligible 1% to at most 28%; real world data sets have been observed to have an impact of 30-50% on performance of the first backup.
- Restore performance was 21% faster for a deduplicated data set, indicating that reconstituting data does not have a negative impact on restores.
- ESG Lab observed negligible differences in backup performance between first, second, and third deduplicated backups. The impact of deduplication should lessen over time since CA ARCserve Backup will not need to process files for deduplication that have not changed since the last backup, reducing processing overhead.

Why This Matters

For years, backup administrators have been struggling to complete nightly backups before business resumes in the morning. ESG research confirmed that IT professionals rank performance as one of the top three criteria when evaluating deduplication solutions.

As backup windows continue to shrink, IT managers are increasingly adopting backup to disk and virtual tape technologies to get nightly backups done quicker. ESG Lab has confirmed that CA ARCserve can provide data deduplication and reconstitution of real-world data sets with minimal impact on performance.

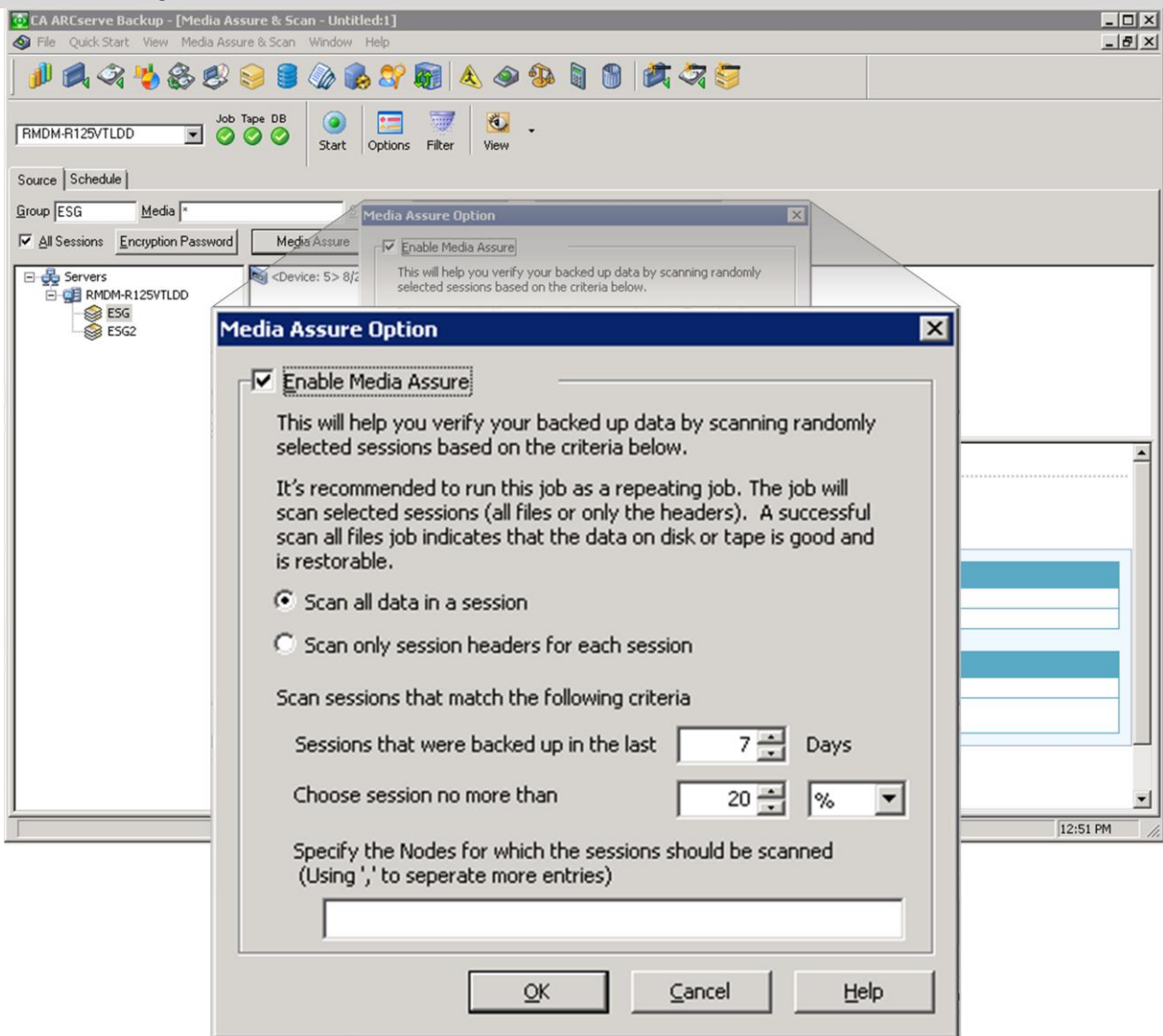
Media Assure

CA ARCserve 12.5 introduces Media Assure, which enables storage administrators to verify that backup data can be read and restored successfully. Media Assure is a policy-based feature that differs from the Scan utility in previous versions of CA ARCserve. The Scan utility was limited to searching session header data for the purpose of determining which devices have which data files. Media Assure examines the actual blocks of data that are written to the backup device. A successful Media Assure job verifies that the backup data is readable and can be successfully restored. An additional benefit of Media Assure is that it randomly selects sessions to scan. This is a significant advantage with deduplication, where all sessions are dependent on each other; a random scan will give the same results as a full image scan.

ESG Lab Testing

ESG Lab used CA ARCserve Manager to configure Media Assure for a media group containing a small test backup.

Figure 12. Enabling Media Assure



As seen in Figure 12, 'Scan all data in a session' was selected. After clicking OK, users are given the option to run Media Assure immediately or to wait for the next scheduled runtime. ESG Lab selected 'Run Now.' Media Assure ran against the backup data and returned a result of 'Scan Operation Successful,' indicating that the backup session tested was readable and able to be restored.

Why This Matters

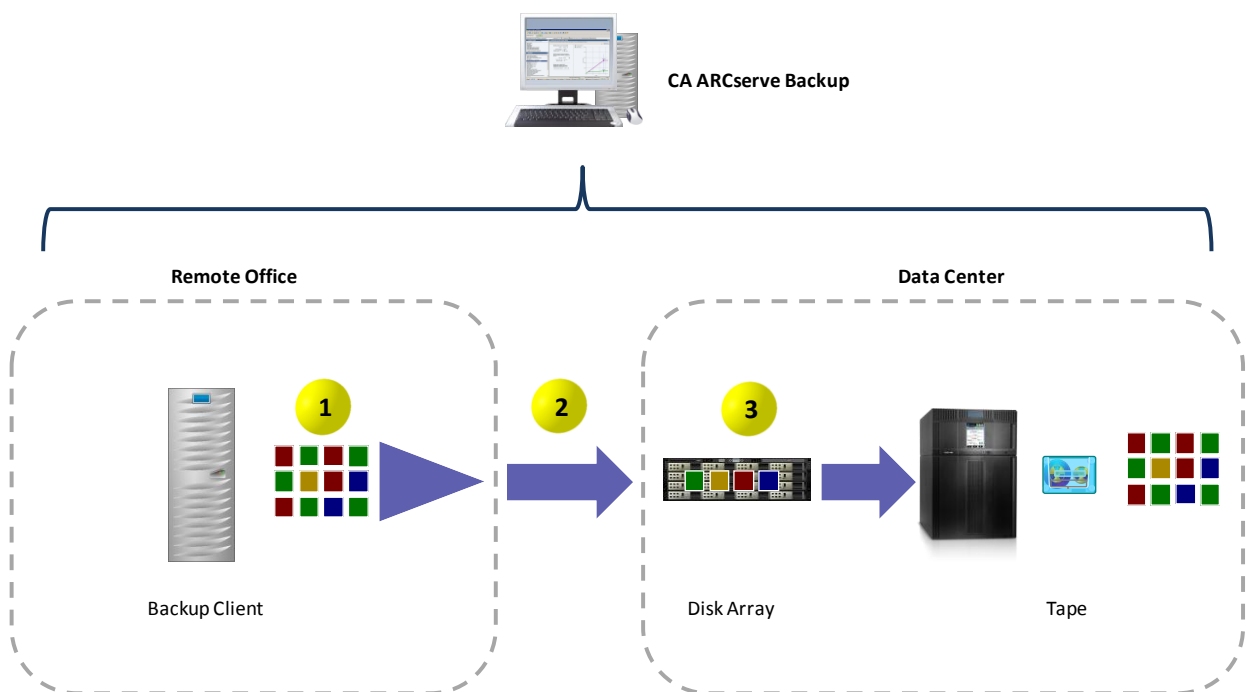
Nearly half of IT professionals surveyed by ESG reported reliability as a challenge with their current data protection processes and technologies.⁸ As a matter of fact, an ESG survey of small to medium-sized businesses indicated that nearly one out of every six (16%) of attempted restores fail. Unreadable media often isn't discovered until a restore is attempted. Missing a restore can be extremely costly in both lost productivity and hard dollars. As a result, backup administrators and IT managers frequently run time-consuming test restore jobs in an attempt to validate that backed up data can be read.

ESG Lab has validated that CA ARCserve Media Assure can be easily configured to regularly examine backup data and verify that all blocks of data in a particular backup set can be read successfully. This policy-based feature assures administrators and managers that backed-up data is ready to restore when it is needed without manual intervention.

Tape Integration

In most backup environments, physical tape is still used for long term and offsite archiving. Organizations that have completely embraced backup-to-disk technology still frequently need to export backups to physical tapes for long term retention or archiving. CA ARCserve Backup provides a single pane of glass to manage data movement and a single catalogue to track and manage all copies of backups. This includes the ability to define separate retention policies for capacity efficient disk-based backups in multiple locations and exportable backups on tape, as seen in Figure 13.

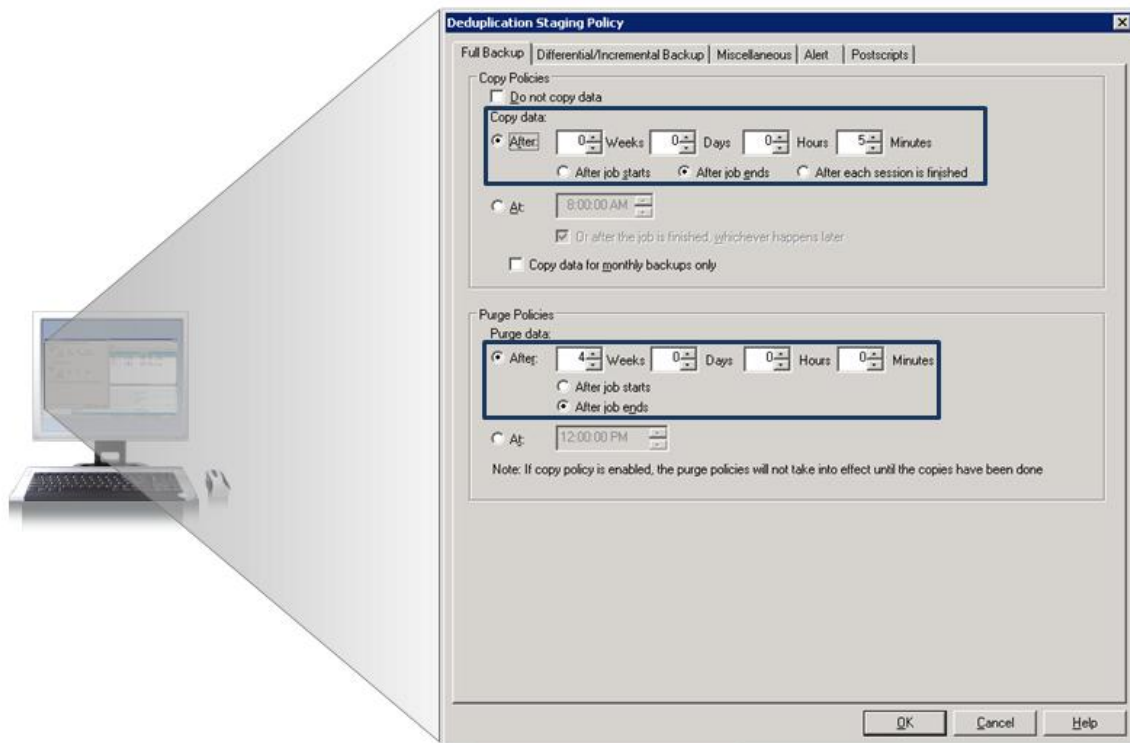
Figure 13. Disk to Deduplicated Disk to Tape



⁸ Source: ESG Research Report, *Data Protection Market Trends*, January 2008.

Figure 14 shows the deduplication policies dialog box, used to configure the automatic copying of backups from disk to tape and the automatic purging of old backups from disk. Administrators can set independent copy and purge policies for full as well as incremental/differential backups.

Figure 14. Configuring Deduplication Staging Policies (Disk to Deduplication to Tape)



Why This Matters

ESG research shows that 86% of organizations surveyed still incorporate physical tape in their backup policies.⁹ Tape copies are made for DR, archival, and compliance requirements. CA ARCserve integration with physical tape enables IT managers to meet their physical tape creation requirements with existing legacy tape systems and to significantly downsize their tape infrastructure as they move an increasing amount of operational backup and restore operations onto deduplicated disk. Hard cost savings are realized from reduced expenditures on tape drives, tape libraries, and media servers. Soft cost savings flow from increased reliability, decreased physical tape management, and reduced offsite handling requirements, all of which free up dedicated backup administrators for other productive tasks.

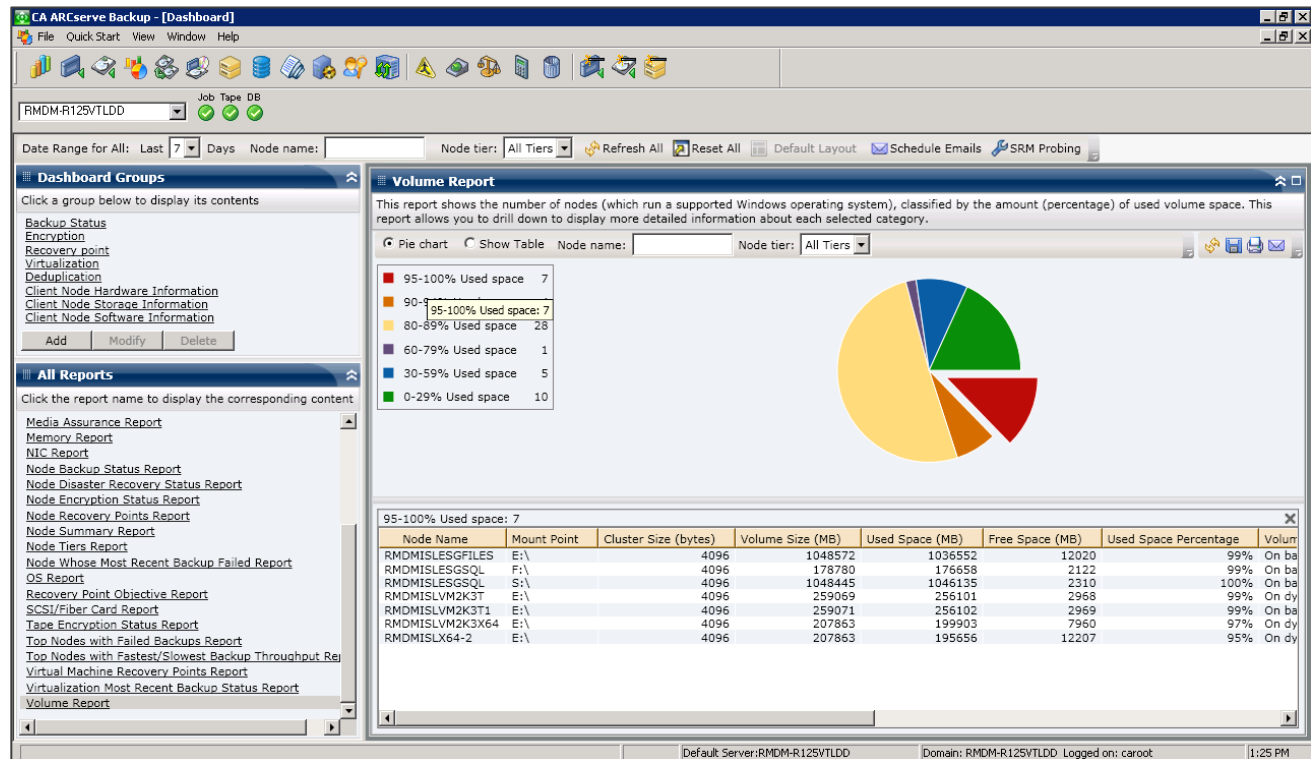
ESG Lab has confirmed that CA ARCserve Backup provides the benefits of a centrally managed disk to disk to tape approach that can be extended across sites and across storage technologies. In this scenario, CA ARCserve Backup reduces the cost of disk capacity and WAN bandwidth while managing all data and metadata centrally, through one pane of glass. ARCserve initiates all data movement, provides the ability to define separate retention policies for disk and tape copies in multiple locations, and maintains a single catalogue to manage it all. For users with multiple sites and the need for multiple tiers of backup data, the combination provides a significant set of benefits.

⁹ Source: ESG Research Report, *Data Protection Market Trends*, January 2008.

Storage Resource Reporting

CA ARCserve Backup 12.5 incorporates high level graphical reporting on data protection and environment resources for the backup environment. The CA ARCserve Backup Dashboard offers graphical reports, including near-real-time and historical information on storage hardware and software in the backup environment. As seen in Figure 15, the Volume Report shows nodes classified by percentage of used volume space. System administrators can drill down to display more detailed information about any selected category.

Figure 15. Storage Resource Management with CA ARCserve Backup



Why This Matters

More than half of the IT professionals surveyed by ESG reported 'keeping pace with the capacity of data to protect' as a challenge with their current data protection processes and technologies.¹⁰ As a matter of fact, 14% of those surveyed named this as their primary challenge.

ESG Lab examined the CA ARCserve Backup Dashboard and found that data protection and environment resource reports were easy to generate and provided extremely useful information, including the amount of storage space consumed and available, memory and CPU utilization, and load distribution for performance analysis, as well as deduplication status.

¹⁰ Source: ESG Research Report, *Data Protection Market Trends*, January 2008.

ESG Lab Validation Highlights

- ☑ CA ARCserve Backup deduplication was extremely easy to configure and manage. Deduplication was implemented and reducing backup capacity in less than three minutes thanks to a short configuration wizard.
- ☑ CA ARCserve Backup deduplication reduced disk capacity required to store backups by 95% over 30 days.
- ☑ The impact of deduplication on a single first backup ranged from 1% to 28%, depending on data type and uniqueness of data. The data for these backups was manufactured to be very unique; real world data sets have been observed to have an impact of 30-50% on performance of the first backup.
- ☑ Restore performance was 21% faster for a deduplicated data set than for a raw file system.
- ☑ CA ARCserve Media Assure was easily configured to regularly examine backup data and verify that all blocks of data in a particular backup set could be read successfully without user intervention.
- ☑ CA ARCserve Backup provided a centrally managed disk to disk to tape solution that was easy to set up and manage across storage technologies.

Issues to Consider

- ☑ While the cost of deduplication software functionality is built into CA ARCserve backup 12.5 at no additional cost, based on the amount of data to be protected, users may need to deploy more media server horsepower for deduplication.

The Bigger Truth

Organizations are challenged to keep pace with the volume of data maintained for recovery purposes. The capacity of data to be backed up makes it difficult to complete backup processing within prescribed windows. To accelerate backup and recovery, IT organizations have deployed disk in the backup data path—improving both performance and reliability of backup and recovery processes. Because recovery from disk is more rapid than recovery from physical tape, leveraging disk in backup/recovery improves recovery time objectives (RTOs), which are the time between an interruption in service and the time when the system is again operational. The retention of multiple backup cycles on disk can boost capacity requirements significantly, increasing hardware and management costs.

During this independent validation, ESG Lab confirmed that CA ARCserve Backup with deduplication provides flexible policy-based data deduplication technology, delivering dramatic disk capacity savings while offering data protection options optimized for both the type and value of the data protected. Given the cost-consciousness of IT organizations today, capacity optimization techniques such as data deduplication are very desirable.

CA ARCserve Backup deduplication runs in the backup target and is licensed based on the amount of data users protect, not the number of media servers it runs on. The bottom line benefit for organizations of all sizes is that deduplication, storage resource management, and continuous data protection are built into the product at no extra charge. That, along with simple pricing models that are very easy to understand and manage, makes CA deduplication extremely cost competitive.

ESG Lab was able to configure deduplication with a few mouse clicks and begin running backups in minutes. Deduplication was efficient and effective, reducing capacity required by a month's worth of backups by 95% with little performance impact. Creation of physical tape was simple and automatable and Media Assure provided the ability to verify that backups are readable and recoverable. Storage resource management capabilities provided insight into the configuration and capacity utilization of CA ARCserve Backup clients.

CA ARCserve Backup delivers comprehensive and cost-efficient backup and recovery for heterogeneous physical and virtual server environments. CA incorporates many high-value features that optimize the data protection environment, including built-in SRM, reporting and management, and data deduplication—features that other vendors charge a premium for.

Appendix

Table 3. ESG Lab Test Bed

Servers	
Primary backup server	CA ARCserve v12.5 build 5854 Quad core Xeon 2.33 GHz, 8 GB RA Windows 2008, Enterprise Edition, R2 18 TB SAS attached backup to disk capacity
File server	Dual core Xeon 2.4 GHz, 4GB RAM Windows 2008, Enterprise Edition, R2 68,868 files on a 2 TB SAN attached LUN
E-mail server	Dual core Xeon 2.4 GHz, 4GB RAM Windows 2008, Enterprise Edition, R2 Exchange Server 2008, SP1 2 TB SAN attached LUN
Database server	Dual core Xeon 2.4 GHz, 4GB RAM Windows 2008, Enterprise Edition, R2 SQL Server 2008, SP1 1 TB SAN attached LUN
VMware ESX server	VMware ESX version 3.5 Quad core Xeon 2.4 GHz, 32 GB RAM Windows 2008, Enterprise Edition, R2 2 TB SAN attached LUN
VMware VCB proxy server	Dual core Xeon 2.4 GHz, 4GB RAM Windows 2008, Enterprise Edition, R2 2 TB SAN attached LUN
Microsoft Hyper-V server	Quad core Xeon 2.4 GHz, 32 GB RAM Windows 2008, R1 1 TB SAN attached LUN
Connectivity	
Gigabit Ethernet LAN	NetGear ProSafe GSM7224
Fibre Channel SAN	Brocade 3900
Backup to Disk Hardware	
Primary storage	4x4 Gbps FC disk array with 1 TB SATA drives
Secondary storage	2x3 Gbps SAS disk array with 1 TB SATA drives



Enterprise Strategy Group | **Getting to the bigger truth.**