

Lab ValidationReport

VMware vSphere Data Protection

Fast, Simple, and Agentless Deduplicated Virtual Machine Backup and Recovery

By Tony Palmer, Senior Lab Analyst

February 2013



Contents

Introduction	3
Background	
vSphere Data Protection (VDP)	3
ESG Lab Validation	5
Backup and Recovery with VDP	5
VDP Data Deduplication	
Cost of Ownership	13
ESG Lab Validation Highlights	15
Issues to Consider	
The Bigger Truth	
Appendix	
Cost of Ownership Assumptions	

ESG Lab Reports

The goal of ESG Lab reports is to educate IT professionals about data center technology products for companies of all types and sizes. 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 VMware.

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

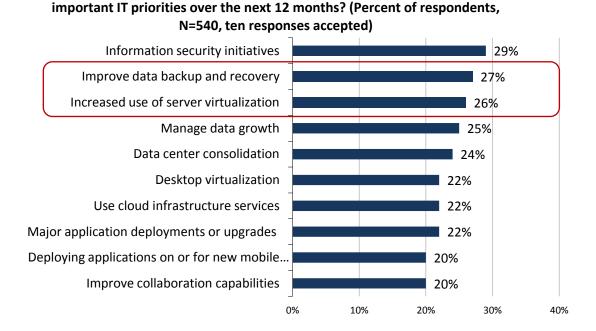
This report documents hands-on testing of <u>VMware</u> vSphere Data Protection (VDP) with a focus on the operational benefits of the simple and manageable data protection offered within the VMware management infrastructure, as well as the cost and capacity efficiency of the solution's integrated deduplication technology.

Background

IT teams are struggling (as always) to meet their backup and recovery objectives; minimize downtime and data loss; ensure that data is always recoverable; and overcome similar pervasive data protection challenges. Often, the inability to adequately protect data or services can delay or prevent deployment of critical applications or services. In a recent survey conducted by ESG, businesses report that data protection and the increased use of server virtualization continue to be high-priority IT initiatives as seen in Figure 1.¹

Which of the following would you consider to be your organization's most

Figure 1. Top Ten IT Priorities for 2013-14



Source: Enterprise Strategy Group, 2013.

Virtualization helps to increase operational agility, facilitates higher availability, and reduces equipment and operational costs through consolidation. However, many existing backup solutions have not kept pace with these new environments and often negate the resource reduction benefits realized through virtualization. Traditional backup methods with agents in each VM can often increase the load on hosts and the time required to complete backups. A backup solution that can leverage the core attributes of virtualization—including cost, agility, and availability benefits—will accelerate the virtualization of tier-1 mission- and business-critical application workloads, which must be assured of high availability and full protection.

vSphere Data Protection (VDP)

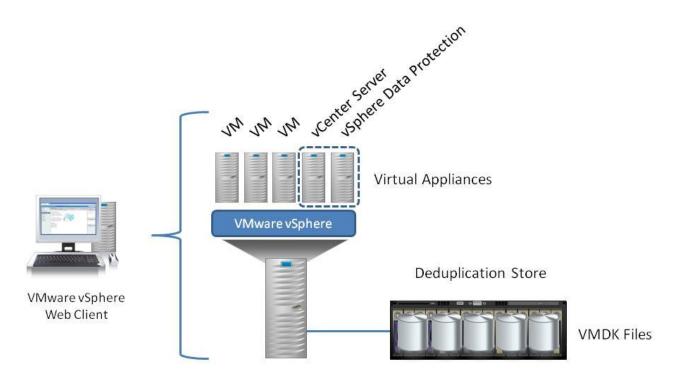
In August 2012, <u>VMware</u> introduced vSphere Data Protection (VDP) as a backup and recovery solution designed for (and in) vSphere 5.1. VDP replaced the VMware Data Recovery (VDR) product, and has been designed to protect organizations' data through fast agentless backups to disk, using deduplication to minimize use of disk space for backups.

¹ Source: ESG Research Report, <u>2013 IT Spending Intentions Survey</u>, January 2013.



With VDP, VMware's small- to medium-sized business customers (fewer than 1,000 employees) are presented with a free, integrated backup solution accessible within VMware's vSphere web client. VDP is deployed as a virtual appliance in a vSphere 5.1 environment, as seen in Figure 2. The VDP appliance can utilize local, NFS, or SAN-attached storage as a deduplication store. Up to ten VDP appliances can be deployed in a single VMware vCenter environment. VDP is included with vSphere Essentials Plus and above.

Figure 2. VMware vSphere Data Protection



VDP is tightly integrated with vCenter Server, designed to allow users to quickly set up backup policies directly from the vSphere Web Client. Administrators can define different policies based on specified schedules and data retention requirements. Policies are applied to groups of virtual machines, based on business needs and data types. Administrators can also easily specify the backup window so as not to interfere with production compute requirements.

VDP is deployed as a virtual appliance with four processors (vCPUs) and 4GB of RAM. Three configurations of usable backup storage capacity are available: .5TB, 1TB, and 2TB, which consume 850GB, 1,300GB, and 3,100GB of actual storage capacity respectively. A variable-length deduplication algorithm ensures a minimum amount of disk space is used and reduces ongoing backup storage growth. Data is deduplicated across all Virtual Machines (VMs) associated with each VDP virtual appliance.

VDP reduces both the load on virtual machines and hosts as well as network bandwidth. Through tight vStorage APIs for Data Protection (VADP) integration, VDP leverages Changed Block Tracking, sending only unique changes over the network in order to minimize traffic. The process is completely transparent and automated, enabling up to eight VMs to be backed up concurrently. Since VDP resides in a dedicated virtual appliance, backup processes are offloaded from production VMs.



ESG Lab Validation

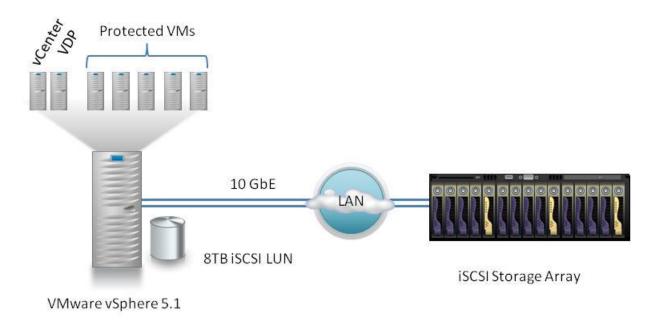
ESG Lab performed hands-on evaluation and testing of VMware vSphere Data Protection in ESG Lab facilities in Lewisville, Texas and San Mateo, California. Testing was designed to demonstrate how the integrated solution streamlines administrative effort and provides simple, reliable, cost-efficient, and agile backup and recovery in VMware vSphere environments.

Backup and Recovery with VDP

The ESG Lab test bed is summarized in Figure 3. One x86 server with dual quad-core Intel Xeon processors and 40GB of RAM was connected to an iSCSI storage array over a dedicated 1GbE LAN with 8x 2TB 7.2K SATA drives. The iSCSI array presented one 8TB LUN to the server. The physical server was running VMware vSphere 5.1, VMware vCenter, and hosting five Windows 2008 R2 virtual machines. Virtual machines were configured with approximately 20GB of data.²

A 2TB VMware vSphere Data Protection appliance was installed and used to protect all five virtual machines.

Figure 3. The ESG Lab Test Bed



ESG Lab Testing

First, ESG Lab logged into the VMware vSphere web client. Clicking on vSphere Data Protection from the home screen brings up a dialog box prompting the user to select and connect to a VDP appliance. Once connected, the VDP menu is presented, as seen in Figure 4.

² Configuration details can be found in the Appendix.

Finish Cancel



Figure 4. the VMware vSphere Web Client



Backup

The entire VDP menu consists of five tabs. The Getting Started tab allows users to quickly create backup jobs or perform full VM restores. ESG Lab clicked Create Backup Job to set up and schedule a backup job to protect all five VMs.

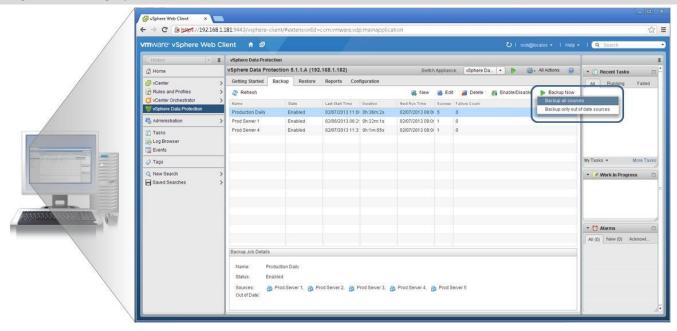


As shown in Figure 5, the Create Backup Job wizard presented a tree view of the vSphere environment. ESG Lab selected VMs to be protected by simply clicking on the checkboxes next to each virtual machine name. Next, the wizard prompted for the backup schedule (users can select daily, weekly, or monthly and specify the day/date of execution), the retention policy, and a name for the job. Configuring the backup took less than two minutes.



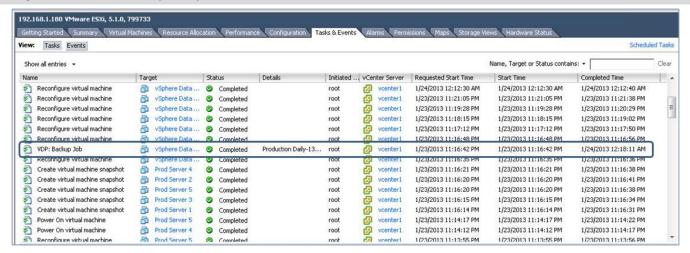
Once the backup job was created, ESG Lab executed the first full backup by simply selecting Backup Now, then clicking Backup all sources, as shown in Figure 6.

Figure 6. Backing Up all VMs With One Click



The virtual machines being backed up contained a mix of data types and file sizes, including one VM with more than two million small files in a complex directory structure.³ The first full backup of all five VMs executed in parallel and completed in just over an hour.

Figure 7. First Full Backup Completed



Next, ESG Lab copied in new data and deleted existing data to simulate a 1% daily change rate, and performed another backup. The second backup took four minutes and 36 seconds, as it only had a small amount of new data to transfer, but it created a full restore point of all five servers. This procedure was repeated ten times, to simulate ten daily backups with incremental data included in each.

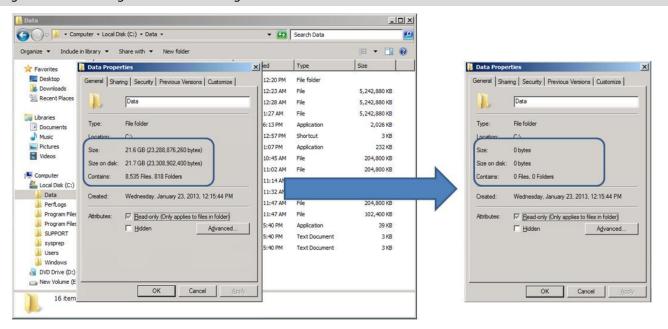
³ Data set details can be found in the Appendix.



Image Level Recovery

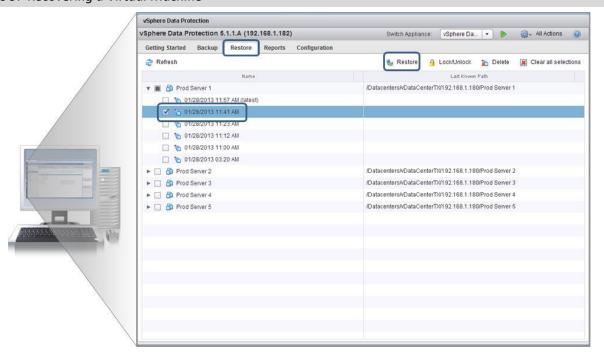
Recovery was tested next, starting with a full restore of a single virtual machine, Prod Server 1. First, ESG Lab permanently deleted all files and folders in the C:\Data directory, totaling 21.6GB, then powered off the virtual machine, as shown in Figure 8.

Figure 8. Recovering a Full VM-Deleting Data



Full VM recoveries are performed from the VDP Restore tab, seen in Figure 9. The administrator browses to the desired restore point, clicks the checkbox to select it, then clicks Restore.

Figure 9. Recovering a Virtual Machine



The 30GB VM was fully restored and booted up in less than three minutes. ESG Lab verified that the VM was restored fully and completely, with all files and folders completely recovered. VDP uses VMware Changed Block



Tracking (CBT) to optimize backups and restores. In this case the full restore transferred only the changed blocks and did not need to restore thousands of individual deleted files, enabling the backup to complete in just three minutes.

To highlight the difference between VDP and conventional backups, ESG Lab performed exactly the same file deletion and full restore using a conventional backup utility without CBT integration. The virtual machine's hard drive contained 33.5GB of data in 33,462 files. The full restore of the virtual machine completed in 18 minutes and 49 seconds, more than six times longer than VDP.

File Level Recovery

Finally, ESG Lab examined file level recovery (FLR) using the VDP Restore Client. The restore client can be accessed from any machine that is being backed up by VDP using a web browser. The Restore client is accessed in one of two modes: Basic or Advanced. Basic mode is accessed using the administrator credentials for the local machine and allows users to mount backups that were made from the machine they log in from. Files may only be restored to the same machine. Advanced mode requires both local machine administrator credentials as well as vCenter administrator credentials and allows administrators to mount and browse any backups that are contained in vSphere Data Protection.

After deleting a single file from Prod Server 1, ESG Lab logged into the VDP Restore Client in Basic mode, mounted the restore point, browsed to the correct folder, selected the file to restore, and clicked Restore Selected Files, as seen in Figure 10. The VDP Restore Client then prompted for the destination folder. ESG Lab selected the same location, but VDP allows for items to be restored to different folders or drives as needed.

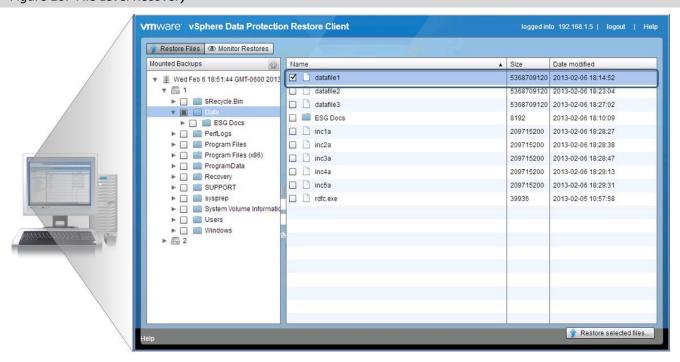


Figure 10. File Level Recovery

The 200MB file was restored to the running VM with no further intervention. ESG also performed file level restores from other protected VMs in advanced mode, and the procedure was exactly the same and just as easy.



Why This Matters

Simplifying operations is important as organizations expand virtual deployments with a goal of saving both time and money, while ensuring optimal data availability and business productivity. In addition, server virtualization has accelerated the pace of business by enabling such capabilities as nearly instant infrastructure provisioning with greater flexibility and availability. Realizing the benefits, organizations are working to virtualize more production and mission-critical applications. At the same time, organizations have little tolerance for business interruption and downtime. When surveyed by ESG, 53% of respondents said they can tolerate less than one hour of downtime for tier-1 data without significant business impact.⁴

ESG Lab confirmed that VDP is a robust and easy-to-use solution that provides tightly integrated data protection for VMware vSphere environments. VDP's simple, intuitive user interface is built specifically for virtualization and the solution is delivered as a virtual appliance, making implementation and management easy. It is designed to enable effortless backup and recovery, with no agents to install or manage and multiple deployment options to enable optimal use of existing resources. As a result, organizations can save money by requiring fewer dedicated backup components (including no agents in VMs) than traditional backup solutions.

ESG Lab has validated that with VDP, IT administrators can perform full restores of VMs from Change Block Tracking in seconds to minutes—enabling restart of a 33.5GB VM in less than three minutes in ESG's testing, compared to nearly 19 minutes for a conventional restore utility which had to restore every file and every byte of data in the volume. This powerful feature minimizes downtime and disruption to not only business users, but also IT. File level restore was extremely easy, and can be leveraged by both application and server administrators to reduce recovery time and streamline the recovery process.

⁴Source: ESG Market Landscape Report, *Replication Technologies for Business Continuity and Disaster Recovery*, August 2011.

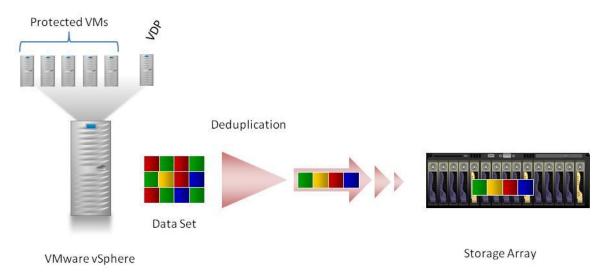


VDP Data Deduplication

Data deduplication reduces capacity requirements by ensuring that only unique data is written to disk. As backup data is stored, the VDP appliance leverages an inline, variable-length block-level data deduplication process which identifies unique blocks of data. As backup data is written, unique blocks are identified and only unique backup data is stored. In other words, when a block is ingested that has already been processed, the appliance stores a pointer to the original block instead of copying the block again.

To illustrate how this works, consider the diagram shown in Figure 11. On the left, the colored blocks represent chunks of data in a data set. As VDP is processing and sending backup data to disk, the unique chunks are identified, and duplicate blocks are replaced by pointers while the unique chunks are written to disk, reducing the amount of data stored on disk.

Figure 11. Deduplication in VDP



This represents only part of the story. The deduplication algorithm in VDP examines data looking for patterns but is not restricted to a single, fixed chunk size. Variable-length deduplication in VDP is able to find patterns of blocks on disks of varying length and deduplicate matching patterns.

Figure 12. Variable- and Fixed-Length Deduplication 12 12654399999987 6543 Variable-length deduplication 98799999126543 99999 987 12654399999987 1265439 9999987 Fixed-length deduplication 98799999126543 9879999 9126543

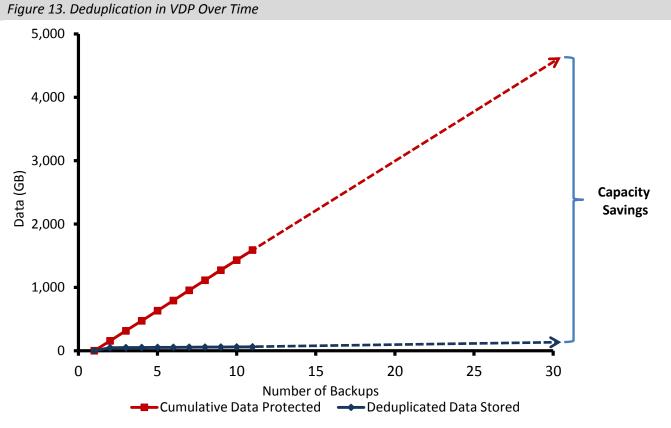


As illustrated in Figure 12, variable-length deduplication is able to find more duplicate data patterns than fixed length because some of the repeating patterns are smaller than the fixed-length deduplication chunk size and are effectively "hidden" inside the larger fixed-size chunks that are being examined.

ESG Lab Testing

ESG Lab ran multiple full backups using VDP and another backup utility, which uses fixed-length deduplication. The backups were run against two identical groups of five virtual machines using a combination of real-world file data and randomly generated data sets. The five machines in each group contained a total of 156.1GB of data at the start of testing, including approximately 8GB of OS data and 22GB of user and application data per VM.

A daily change rate of 1% (approximately 200MB of new/changed data per day) was simulated by copying 300MB of new file data unique to each VM, then deleting 100MB of existing data from each VM prior to each full backup. After 11 iterations, 55 backups had been performed and 1,748GB of data had been ingested. ESG Lab confirmed that at this point VDP had stored only 31GB of data. Fixed-length deduplication had reduced the data as well, but had used 115.6GB of capacity to protect the same 1.7TB.



In 13, ESG Lab projects out the cumulative effects of variable-length deduplication for a 156GB data set over 30 days. Running daily full backups, the total data protected would grow to 4.77TB while the disk space required would only reach 45.2GB, representing a reduction rate of 99.1%.



Table 1 details the data captured during the 11 backup iterations performed by ESG Lab.

Table 1. Deduplication Efficiency Over Time

Number of Backups (Five VMs)	Cumulative Data Protected (GB)	Deduplicated Data Stored (Variable-Length)		Deduplicated Data Stored (Fixed-Length)	
1	156.1	23.5	85.0%	84.2	46.1%
2	313.7	24.2	92.3%	87.3	72.2%
3	472.1	25.0	94.7%	90.5	80.8%
4	631.5	25.7	95.9%	93.6	85.2%
5	791.9	26.5	96.7%	96.8	87.8%
6	953.4	27.2	97.1%	99.9	89.5%
7	1,112.3	28.0	97.5%	103.0	90.7%
8	1,271.2	28.7	97.7%	106.2	91.6%
9	1,430.1	29.5	97.9%	109.3	92.4%
10	1,589.0	30.2	98.1%	112.5	92.9%
11	1747.9	31.0	98.2%	115.6	93.4%

What the Numbers Mean

- VDP variable-length deduplication was able to store 1.7TB of restore points for five virtual machines in just 31GB of disk space.
- A single 2TB VDP appliance would be able to protect more than 100TB of user backups with similar characteristics and change rate.
- Fixed-length deduplication required nearly four times the disk space to store the same data sets with the same daily full backup policy.

Why This Matters

ESG research confirms that data growth and data protection continue to challenge IT organizations; respondents to ESG's annual IT spending surveys for the past three years have listed managing data growth and improving backup and recovery as two of their top five IT priorities. The costs of storing and protecting ever-expanding data sets can stress capital and operational budgets.

ESG Lab has validated that VDP variable-length deduplication can be used to reduce disk capacity required to store backups by up to 99% depending on data type, change rate, backup policies in use, and retention period. It is also clear that some data sets are more easily reduced than others. When larger data sets are examined, the potential savings are astounding—customers can effectively retain 100TB of backup data for quick and reliable restores using only 2TB of disk capacity. This results in lowering the cost per GB for backup data and enables companies to retain data exponentially longer for recovery purposes.

Cost of Ownership

ESG Lab modeled and analyzed the total cost of backup storage ownership for customers in the medium-sized business segment. The costs were modeled using midrange systems that could provide up to 100TB of block-based storage capacity to act as a backup repository.

The ESG Lab analysis was quantitative in that it examined the cost of acquisition (hardware and software), support, management (including manpower), and power and cooling. ESG calculated the cost of acquiring and maintaining



sufficient storage to house user backups for environments ranging from 1.6TB to 16TB of data to be protected, with a retention policy of 60 days.⁵

ESG Lab calculated that an organization with 1.6TB of primary data, running daily full backups for 60 days would require approximately 100TB of non-deduplicated capacity to store all backups. Figure 14 shows the relative cost of ownership for non-deduplicated storage versus VDP variable-length deduplication as well as the relative cost of storage for a fixed-length deduplication solution.

Figure 14. Relative Cost of Ownership for 100TB of Retained Backups.

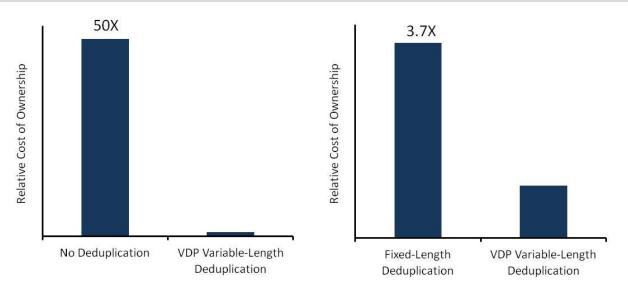


Table 2 shows the calculated costs based on cumulative protected storage.

Table 2. Cost Analysis

	No Deduplication	Fixed-length Deduplication	VDP	No Deduplication	Fixed-length Deduplication	VDP
Cumulative Protected Data Capacity	100ТВ	100TB	100TB	1,000TB	1,000TB	1,000TB
Required Capacity	100TB	7.1TB	1.9TB	1,000TB	70.8TB	19.1TB
Entry Level Storage	\$119,466	\$8,482	\$2,270	\$1,194,663	\$84,582	\$22,699
Mid Tier Storage	\$257,374	\$18,274	\$4,890	\$2,573,740	\$182,221	\$48,901
Enterprise Storage	\$481,384	\$34,178	\$9,146	\$4,813,844	\$340,820	\$91,463

What the Numbers Mean

- VDP variable-length deduplication can reduce overall storage costs by up to 98%.
- A fixed-length deduplication solution would require 3.7X more capacity and incur higher storage costs than VDP.
- VDP is included with vSphere Essentials Plus and above, with no additional software or server licensing costs.

-

⁵ For more details, please refer to the Appendix.



ESG Lab Validation Highlights

- ☑ ESG Lab found VDP easy to implement and use, providing tightly integrated data protection for VMware vSphere environments.
- ✓ VDP enabled effortless backup and recovery, with no agents to install or manage.
- ☑ ESG Lab was able to perform fast, full restores of VMs from Changed Block Tracking in seconds to minutes—enabling restart of a VM in less than three minutes in ESG's testing.
- ☑ Changed Block Tracking was able to recover a full virtual machine 6 times faster than a full VM restore.
- File level restore was extremely fast and easy as well, and can be leveraged by both application and server administrators to reduce recovery time and streamline the recovery process.
- ☑ VDP variable-length deduplication reduced disk capacity required to store backups by more than 98% in ESG Lab testing, effectively lowering the cost per GB for backup data and enabling companies to retain data on disk longer for fast recovery.
- ☑ Fixed Length deduplication required 3.7 times more storage capacity than VDP to store 1.7TB of backups.

Issues to Consider

- ☑ VDP supports up to 2TB per appliance. While up to 10 VDP appliances can be implemented in each vSphere cluster and with deduplication 2TB can be used to effectively protect relatively large data sets, higher capacity versions would be a great option for larger mid-sized organizations.
- ☑ While VDP provides excellent guest OS and vSphere support for granular backups and restores, integration with the most common applications in small and medium sized businesses such as Exchange and SQL Server, would be a valuable enhancement.



The Bigger Truth

Organizations of all sizes recognize not only the economic benefits of consolidating workloads, but also the greater flexibility, ability to quickly provision new applications, and opportunity to increase data and application availability that comes with server virtualization. In fact, in recent ESG research, increased use of server virtualization was one of the top-three IT priorities reported by respondents for the next 12-18 months.⁶

This reflects the continued expansion of virtualization into tier-1 applications and explains why backup and recovery of VMs is such a critical topic. Like other IT processes, traditional backup methods were built for the one-application-per-server paradigm. Backup solutions presumed that only a single workload at a time would be interrupted by backing up a server. That is no longer true with server virtualization and as a result, backing up multiple VMs can clog networks and interrupt operations. Trying to make physically-based backup work in the virtual domain has proven difficult and has left IT expending significant time, effort, and money without much confidence that backups have been done properly or will be recoverable. When you add these challenges to a generally low tolerance for downtime, particularly of tier-1 applications, it is clear that a fast, reliable, non-disruptive backup solution, purpose built for VMs is urgently needed.

ESG Lab has verified that VMware VDP offers an integrated solution to VMware data protection challenges that meshes perfectly with vSphere, reducing effort and costs and improving service levels for small- to medium-sized organizations. Leveraging VMware Changed Block tracking, VDP can restore full VMs in minutes and enable object-level recovery from any protected system using a universal web-based client.

ESG Lab found VDP fast and easy to implement, with backup configuration being a matter of a few clicks to select the machines to back up, then set the schedule and retention policy. ESG Lab validated that VDP variable-length deduplication transparently reduced the storage footprint of backup data by up to 98% without a negative impact on backup performance while requiring less than 27% of the space than fixed-length deduplication technology. ESG recovered a 30GB VM from deduplicated storage in less than three minutes.

ESG Lab set out to validate VDP as a simple, capacity-efficient, robust, and cost-effective virtual machine backup and recovery solution for small- to medium-sized businesses. After extensive testing, ESG Lab can state with confidence that VDP meets or exceeds all these claims.

The capabilities tested demonstrate clearly that VDP addresses the key challenges of IT managers and the needs of users. Rapid virtual machine recovery through VMware Changed Block Tracking enables VMs to be restarted in minutes, on production storage and without a performance impact. VDP was extremely easy to deploy, requires no agents, and is incredibly easy to operate and manage from the vSphere web client. By tightly integrating robust data protection directly into vSphere, VDP adeptly addresses the challenges that virtualization presents to data protection. Mid-sized organizations using or considering vSphere would be smart to take a close look at VDP for data protection and recovery with compelling advantages.

© 2013 by The Enterprise Strategy Group, Inc. All Rights Reserved.

⁶ Source: ESG Research Report, <u>2013 IT Spending Intentions Survey</u>, January 2013.



Appendix

Table 3. ESG Lab Test Bed

VMware Configuration			
VMware vSphere Essentials Plus	Version 5.1		
vCenter Server (VM)	Version 5.1		
vSphere Data protection Appliance	Version 5.1.0.56.179 Physical Capacity: 2 TB		
Server Configuration			
Dell Precision T4700	2x Quad core 2.992 GHz Xeon Processors 40GB RAM		
Virtual Machines			
Five (5) Clients 2 vCPU, 2GB RAM 40 GB virtual disk	Microsoft Windows Server 2008 R2		
Data Set			

Five virtual machines were each populated with a 22GB initial data set.

- One machine contained 5GB of office files and 15GB of files containing generated random data.
- Three machines contained 20GB of generated random data in large files (1GB to 5GB).
- One machine contained two million small files (4KB to 16KB) containing generated random data.

A net daily change rate of 1% (200MB) was simulated by copying in 1.5% (300MB) of new data unique to each VM, then deleting 100MB of existing data from each VM.

Storage Configuration			
Drobo B800i	Connectivity: iSCSI Usable Capacity: 16TB		

Cost of Ownership Assumptions

A cost-of ownership model was created by ESG Lab to estimate the total cost of backup to disk storage for environments needing to protect up to 16TB of primary data with a 60 day retention policy. Aside from hardware and software costs, as well as support costs for both, the model also took into account power/cooling expenses and management expenses. Power and cooling costs were estimated to be 9.3 cents per kWh based on the average retail cost of electricity in the U.S. as documented by the U.S. Energy Information Administration (http://www.eia.gov/electricity/state/).

Pricing data was gathered from publicly available sources and quotes were provided to ESG Lab by two resellers as of March 1, 2012. An average total cost per terabyte was used.

Management costs were calculated based on average salaries, as well as common tasks associated with the management of an IT infrastructure. It was assumed that a senior storage administrator with an hourly rate of \$55/hour is needed to manage storage solutions (http://www.salarydom.com). The tasks ESG Lab modeled were: monitor, plan, provision, expand, tier, snap setup, snap recover, DR setup, DR test, and network configuration. Two other larger tasks included in the management costs were the migration of old data to the newly deployed infrastructure and the addition of a new system to an existing infrastructure. Each task was assigned an amount of time to complete (in minutes), as well as a monthly frequency.

