

# Allocating Target Storage for DeDupe Backups

Doug Soltesz  
Sr. Technical Sales Engineer  
www.phdvirtual.com  
January 2010

## I. Introduction

Source-side DeDupe backups with esXpress provide significant backup time and storage benefits, including:

- Reduced size of backed-up data.
- Less time required to generate and transport backup data to target storage device(s).
- Vastly improved restoration timeframe with point-and-click restoration of individual files or complete virtual disks without fragmentation.
- Significantly reduced network and storage requirements.

## II. Getting Started

To maximize the benefits, it's important to optimize the amount of target storage allocated to these backups. If there is not enough space available, then backups may fail, so it's better to be generous initially. At the same time, you should set trimming aggressively for the first month, and then adjust the storage/trim balance once you've established a sensible baseline.

### Setting Trim

The first step in determining target space requirements is to decide how many backups to maintain on disk at prescribed times. This is governed by your trim settings. Keep in mind that:

- DeDupe efficiency improves when additional days of backups are maintained.
- esXpress DeDupe appliance trim settings are global and apply to all backups stored on the specified DeDupe appliance.
- Despite trim settings, backups of individual VMs can be manually deleted or, alternatively, they can be archived, which prevents auto-deletion.

esXpress Auto Trimming provides efficiency by allowing you to take multiple backups of a VM each day, and then, after a set time, to delete all but the last backup for each day. Similarly, daily backups can be trimmed over time, down to only one backup per week, and finally weekly backups can be trimmed to keep only one per month.

### How Many Backups to Keep?

The total number of backups that will be kept depends on your trim settings. Here's a simple example of how you might set your backup retention:

Current Date April 30<sup>th</sup>, 2010

### Backups Kept per VM

- 2 Backups per day for one week (14 backups)
- 1 Backup per day for weeks 2 and 3 (14 backups)
- 1 Backup per week for week 4-8 (5 backups)
- 1 Backup per Month for Month 3 + 4 (2 backups)
- Keep nothing over 4 Months old (0)

On this schedule, 35 backups are being kept per VM. Note that the minimum trim setting is one week of daily backups (7 Backups per VM).

### How Much Data Will Be Backed Up?

Next, it is important to determine the amount of data that will be backed up on a nightly basis. This is the size of all of the VMDKs in your system that will be backed up, including null space. If you are already using esXpress, you can find this information in the Dedupe web interface on the **Stats** page, under the **By Date** tab.

Below is a snapshot of a sample environment.

Date	Count	VMs	VMDKs	Size_VM	All	Ratio	Nulls	Percent	Data	Ratio	Wrote	Ratio	Sent
2009-12-13	56	37	47	976,125	1,055,997	109:1	394,823	37%	661,174	68:1	20,880	2:1	9,663
2009-12-12	56	37	47	976,125	1,055,997	108:1	394,840	37%	661,157	67:1	22,257	2:1	9,818

The **Size VM** column shows the total size of the VMs being backed up daily. The **All** column shows the total size of all backups done that day. When, as in this example, the All column is greater than the Size VM column, some VMs are being backed up multiple times per day. In this case, the total size of VMs is 976,125 MB, which can be rounded up to 1 TB for simplicity in calculating the compression ratio in the following example.

### Factoring in the Compression Ratio

The final step in calculating the potential size of the backup target is to get the compression ratio. If you are already using esXpress, the current ratio can be found in the Dedupe web interface on the **Stats** page. If not, refer to the table below for some good rule of thumb compression ratios. This varies by environment, depending on how much of the data is changing as well as the number of backups kept. The more backups kept, the higher the ratio. The following screen shot shows a compression ratio of 33:1, which is slightly better than what is found typically:

Stat Value	In Megabytes	Ratio
All Blocks of All VMs	39,996,738	33:1
Null Blocks	12,879,289	32%
Real Data Blocks	27,117,449	23:1
Total Unique Blocks	2,173,041	2:1
Total Unique Blocks Compressed	1,203,589	1:1

Rule of Thumb Assumptions	
Backups Kept	Compression
7 per VM	12:1
14 per VM	20:1
21 per VM	25:1

## Sizing DeDupe Target Storage

Based on these three required statistics, the size of all VMs, the number of backups kept, and the compression ratio, the estimated DeDupe target size can be calculated using this formula:

$$\text{Estimated Target Size} = \frac{(\text{Size of all VMs}) \times (\# \text{ of Backups kept})}{(\text{Compression Ratio})}$$

### Example 1

In the example discussed earlier, the Size of all VMs is now 1TB, 35 backups of each VM will be kept, and current compression is 33:1.

$$\text{Estimated Target Size} = \frac{1 \text{ TB} \times 35 \text{ Backups}}{33:1 \text{ Compression}} = 1.06 \text{ TB}$$

Therefore, the Target Size for this example should be 1.06TB. Remember, however, that this does not account for future growth or the possibility that data may start changing more rapidly.

### Example 2

In this example, VMDKs are currently at 4TB, but are expected grow to 5TB within a year. Twenty Backups of each VM are kept at any time, and the default compression ratio of 25:1 is applied.

$$\text{Estimated Target Size} = \frac{5 \text{ TB} \times 20 \text{ Backups}}{25:1 \text{ Compression}} = 4 \text{ TB}$$

Therefore, in order to accommodate for the growth of data over the next year, the backup target should be 4TB in size. Remember that all of the allocated storage does not have to be on a single target. esXpress allows for many DeDupe Appliances; backups could be split between two appliances each with 2TB of storage space.

## Optimal Allocation of Target Storage

In summary, you can maximize the effective benefits of esXpress by allocating optimal target storage for your DeDupe backups. Adjust your trim settings to suit your needs, and then use the following formula to estimate the appropriate target size.

$$\text{Estimated Target Size} = \frac{(\text{Size of all VMs}) \times (\# \text{ of Backups kept})}{(\text{Compression Ratio})}$$

## About PHD Virtual

As the pioneer of virtual backup appliances (VBAs), PHD Virtual Technologies has been transforming data protection in virtual IT environments since 2006. Its award-winning data protection solution for virtual infrastructures, esXpress, is used today by more than 2000 enterprises worldwide to achieve unlimited dynamic growth, high availability, no single point of failure and scalable performance. PHD Virtual is committed to helping our customers and provides free, easy-to-use virtualization utilities to assist with the administration and management of virtualized environments.

### North America Headquarters

111 Howard Blvd, Suite 104B

Mount Arlington, NJ 07856

+1 (973) 288-7000

[www.phdvirtual.com](http://www.phdvirtual.com)

[www.esxpress.com](http://www.esxpress.com)