

Backup Storage Sizing Guide for VMware vSphere and Citrix XenServer

Frank Musacchia
www.phdvirtual.com
November 2010

Introduction

PHD provides the best backup optimization for LAN/WAN environments. Leveraging our TrueDedupe™ Technology we can provide the most highly efficient LAN/WAN based backup solution with the minimum impact to your network. Deduplicating and compressing backup data on the source side at the host ensures that only the minimum amount of unique data needed for backup is transmitted across the wire, unlike other solutions which can have a dramatic impact on the network or require more bandwidth.

Source-side Deduplication with PHD Virtual provides significant backup time and storage benefits, including:

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

Getting Started

It is important to optimize the amount of Backup Storage allocated to storing your backup data. This will keep costs to a minimum by decreasing the amount storage space necessary to perform backups. There is a balance between keeping large quantities of data, and keeping enough data that is required by the business to meet your compliance and storage requirements.

Retention

The first step in determining Backup Storage space requirements is to decide how many backups will be maintained for each virtual machine that is being backed up. This will be governed by your retention settings, as configured in the VBA Configuration settings. Keep in mind that:

- Dedupe efficiency improves when additional days of backups are maintained.
- PHD Retention settings are associated with the VBA's configuration and apply to all backups stored on the designated Backup Storage target as configured in the VBA Configuration settings.
- Despite the Retention settings, backups of individual VMs can be manually deleted.
- VMs can also be archived, which will prevent deletion even during the auto deletion process.

Retention Settings

By default, PHD Virtual Backup will keep all backups for each VM. By using the Retention options you can select how many backups you want to keep in order to meet your individual compliance and storage requirements. When enabled, a job runs and performs the retention processing each hour and also after every backup job completes.

Retention Options

Set the retention options below in the VBA Configuration menu, under the "Retention Tab" to keep the preferred number of backups necessary to meet your compliance and storage requirements.

- **Keep All** - Retain all backups for all VMs, this is the default setting.
- **Typical** - These settings are preset recommendations and are configured as follows:

Retention Setting	Typical
Recent Backups to Keep	5
And Keep the most recent backup from each of the last:	
Days	7
Weeks	4
Months	12
Years	5

- **Custom** - Define the values for each retention setting. Retain the last "X" recent backups as well as the most recent backups from each of the:
 - Last "X" days
 - Last "X" weeks
 - Last "X" months
 - Last "X" year

Retention Setting	Custom
Recent Backups to Keep	"X"
And Keep the most recent backup from each of the last:	
Days	"X"
Weeks	"X"
Months	"X"
Years	"X"

How Much Data Will Be Backed Up?

It is important to determine the amount of data that will be backed up on a nightly basis. This will be determined by the total size of all of the VMDKs associated with a VBA that will be backed up. This should also be inclusive of the null space contained within those VMs. However, it excludes the storage space in thin provisioned virtual disks that has not yet been allocated to the VM.

Factoring DeDuplication and Compression

It is important to take into consideration the deduplication and compression ratio when sizing the Backup Storage target. PHD Virtual will deduplicate blocks of data across all VMs on the same Backup Storage target. Deduplication will vary by environment based on several factors:

- Whether VMs are being deployed from Templates/Clones
- Change Rate of the data
- The number of backups being kept for each VM

Refer to the table below for an estimation of deduplication and compression ratios.

Backups Kept	Deduplication and Compression
7 per VM	12:1
14 per VM	20:1
21 per VM	25:1
28 per VM	32:1
35 per VM	39:1

Sizing Backup Storage Target

The following three required statistics are used to calculate the Backup Storage size:

1. The total size of all VMs
2. The total number of backups kept
3. The deduplication and compression ratio

The estimated Backup Storage size can be calculated using the following formula:

$$\frac{(\text{Size of all VMs}) \times (\# \text{ of Backups Kept}) \times}{(\text{Deduplication and Compression Ratio})} \approx (\text{Backup Storage Size})$$

Example #1 - Typical Setting

In this example:

- The sum of the VMDKs equals 1TB
- Assumption: Performing daily backups of all VMs.
- The deduplication and compression ratio of 35:1 is applied (approximation)
- The number of backups kept are approximately thirty three (33), based on the table below.

Retention Setting	Typical
Recent Backups to Keep	5
And Keep the most recent backup from each of the last:	
Days	7
Weeks	4
Months	12
Years	5

$$\frac{1 \text{ TB} \times 33 \text{ Backups}}{35 : 1} \approx 1 \text{ TB}$$

Example #2 - Custom Setting

In this example:

- The sum of the VMDKs equals 4TB but are expected to grow to 5TB within a year.
- Assumption: Performing daily backups of all VMs.
- The deduplication and compression ratio of 25:1 is applied (approximation)
- The number of backups kept are approximately twenty (20), based on the table below.

Retention Setting	Custom
Recent Backups to Keep	3
And Keep the most recent backup from each of the last:	
Days	7
Weeks	4
Months	6
Years	0

$$\frac{5 \text{ TB} \times 20 \text{ Backups}}{25 : 1} \approx 4 \text{ TB}$$

Example #3 - Custom Setting

In this example:

- The sum of the VMDKs equal 4TB but are expected to grow to 5TB within a year.
- Assumption: Performing daily backups of all VMs.
- The deduplication and compression ratio of 15:1 is applied (approximation)
- The number of backups kept are approximately ten (10), based on the table below.

Retention Setting	Custom
Recent Backups to Keep	10
And Keep the most recent backup from each of the last:	
Days	0
Weeks	0
Months	0
Years	0

$$\frac{5 \text{ TB} \times 10 \text{ Backups}}{15 : 1} \approx 3.3 \text{ TB}$$

Therefore, the Target Size for this example should be approximately 3.3TB. Remember, however, this does not account for the possibility that the data change rate may fluctuate over time.

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, PHD Virtual Backup (formerly, 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.

For more information contact a PHD Representative today!

Call: 1.866.710.1882 Email: sales@phdvirtual.com

