PHD Virtual™

Virtual Backup Appliance (VBA™) Architecture Sizing Guide

Frank Musacchia
www.phdvirtual.com
May 2, 2011

Introduction

Delivering the highest performance and most scalable cross-platform backup solutions on the market, PHD Virtual Technologies has been transforming data protection for virtual IT environments since 2006. Our mission is to preserve the agility, flexibility and cost savings that drive your IT virtualization goals.

PHD Virtual solutions are purpose-built for virtualization and leverage the award winning PHD VBA™ (Virtual Backup Appliance) Architecture to provide virtualized backup and recovery for VMware and Citrix environments. This evolutionary approach allows PHD Virtual Backup to deliver high performance data protection that seamlessly scales for large and distributed deployments. Unlike alternative solutions, PHD Virtual Backup removes the need to deploy and manage separate physical servers, additional software, scripts or agents for backup and recovery of the virtual environment.

PHD Virtual Benefits

PHD provides a completely virtualized solution leveraging a virtual backup appliance VBA™.

- Using a virtual appliance removes the need to deploy and manage a separate physical infrastructure to backup your virtual environment. The VBA™ is purpose-built for data protection and can leverage the advanced capabilities of virtualization.

PHD snaps into your virtualization platform architecture and scales seamlessly.

- Unlike other alternatives, our solution deploys directly on your virtualization platform and provides integrated single pane of glass management. Scaling up for increased demand is easily accomplished by deploying additional VBA's, without the need to add additional physical servers.

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.

PHD dramatically lowers your backup storage requirements and solution costs.

- With our TrueDedupe™ Technology, customers regularly experience a 25:1 dedupe ratio reducing ongoing backup storage needs by 96%. Since our solution is delivered as an integrated and optimized virtual appliance, management overhead is significantly reduced and there is no need to purchase and manage additional hardware, software or agents.

Overview

This paper will outline at a high level the design considerations necessary for deploying PHD Virtual software. It is imperative to understand the variables that will have an impact on the amount of Virtual Backup Appliances (VBAsTM) necessary to be deployed to meet a customer's backup requirements. This paper will describe the method and formulas for calculating the quantity of VBAs that will need to be deployed in a customer's virtual environment.

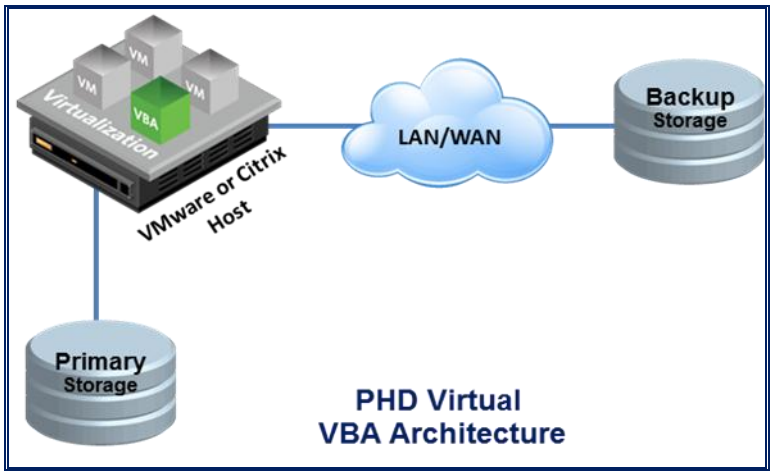
PHD Virtual Architecture

VBA™ Architecture

A VBA™ is a Virtual Backup Appliance - a virtual machine that backs up other virtual machines

- Deploys as a small virtual appliance on either a VMware vSphere or Citrix XenServer host
- Need a minimum of one VBA per VMware Cluster or XenServer Pool.
- Performs image-based block level backup of virtual machines
- Fully integrated with VMware and Citrix API's
- Supports Change Block Tracking (CBT) for VMware
- Leverages the Virtual infrastructure such as vMotion/Live Motion, DRS, HA and Resource Pools
- Requires no additional proxy/media servers or agents
- Purpose built for VM backup and recovery

VBA™ Diagram



Design Considerations

The amount of VBAs necessary to be deployed in a virtualization environment will be determined by several factors as listed below:

- Number of VMware Clusters or XenServer Pools
- Number of Hosts per Cluster/Pool
- Total amount of provisioned disk space for all VMs
- Backup Network performance
- Backup Window timeframe

Number of Clusters/Pools

PHD VBAs will be deployed on a per Cluster/Pool basis. A minimum of one VBA is necessary in order to perform Backups/Restores. Additional VBAs may need to be installed, which can be used to increase the amount of Backup processing power, in order to meet a particular backup window.

Number of Hosts

The number of Hosts is simply the Hypervisor Host count per Cluster/Pool. This will determine the maximum number of PHD VBAs that can be deployed in a virtualized environment.

Amount of Data

In order to determine the number of VBAs to be installed, it is necessary to sum the Total amount of VMDK/VHD disk data, which is provisioned in the virtualized environment for both Thick or Thin provisioned disks.

Backup Network Performance

It is important to gauge the throughput of the backup network when performing PHD backups. The most accurate way of accomplishing this is to backup a Virtual Machine using PHD software and documenting the results of: How much data was written and how much time it took to process and write that data, which will be considered the “Calculated Throughput.”

The “Effective Throughput” can be calculated by taking the total amount of provisioned VM data and dividing by the time it takes to complete the backup. This calculation will yield a substantially larger value than the Calculated Throughput because deduplication and null data removal will be factored into this result, which is normal for backup vendors.

Backup Window

The Backup Window is the total amount of time allocated for performing backups. Usually the backup jobs will be configured to use a subset of the total backup window, to accommodate for any issues.

VBA™ Design

Determining the Amount of Data to be Backed up

The amount of data to be backed up will be determined by the total amount of provisioned disk space per Cluster/Pool. When performing a backup, PHD will need to process and write a certain amount of data to the backup storage target, depending on the deduplication and compression ratios. The amount of data that will eventually be written can usually be calculated by using the formulas below. *These formulas are to be used for planning purposes only, since the disk density and data rate of change can vary for different customers and environments.*

First Full Backup

The amount of data to be processed and written during the first Full Backup will equal approximately 15%. We generally see that only about 15% of provisioned storage has unique data after deduplication and null block removal. So the rule of thumb is only 15% of provisioned data will be written to PHD backup data store.

Full Backup = (Total disk data provisioned) x (15%)

Subsequent Backups

The amount of data to be processed and written during Subsequent Backups will equal approximately 2.5%. We see this generally as a good rule of thumb for the change rate of data in most environments. If a different change rate of data needs to be used, just insert that percentage in the equation below.

Subsequent Backups = (Total disk data provisioned) x (2.5%)

VBA sizing factors

To determine the amount of PHD VBAs that need to be deployed, several factors need to be considered such as:

1. Amount of data to be backed up
2. Calculated Throughput per VBA
3. Backup Window

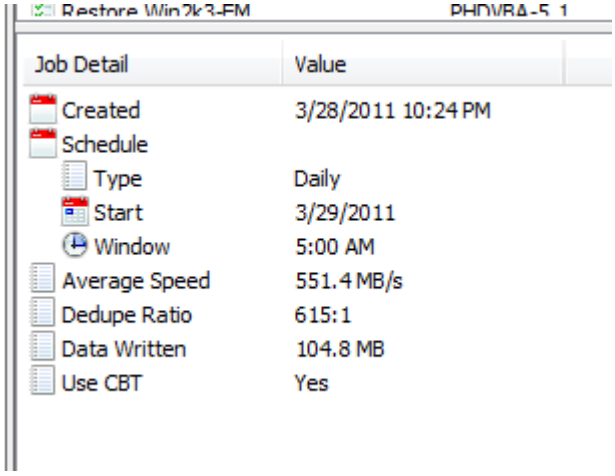
Amount of data to be backed up

Use the subsequent backup data from the calculations above.

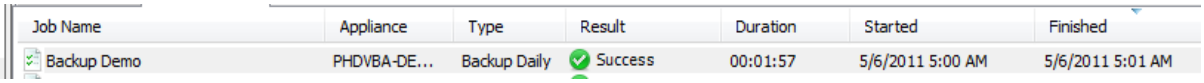
Calculated Throughput per VBA

Throughput per VBA can be determined be performing a PHD backup and calculating the throughput by using the "Data Written" and "Job Duration" data. This can be collected from the PHD Console, under the "Jobs" tab (see Below).

Data Written



Job duration



VBA Calculated Throughput = (Data Written) / (Job Duration)

- If the above Backup job cannot be run for whatever reason, then use 3MB/s as a default approximation.

Backup Window

The Backup Window is the total amount of time allocated for performing backups. Usually the backup jobs will be configured to use a subset of the total backup window, to accommodate for any issues.

Number of VBAs to be deployed

It is common practice to design the VBA count around the Subsequent Backups versus the first Full Backups. The reason for this is PHD virtual will only perform a Full Backup the very first time a VM is deployed. Every Subsequent Backup is considered a "Virtual Full" backup, which is similar in theory to an Incremental Backup.

What this means is that PHD will only write "Changed Data" to the backup data store. So after the first Full Backup is completed, only **unique, changed data** will be written to the backup data store, every subsequent backup.

Examples

Example #1:

- Total Provisioned Data = 3TB
- Data to be written (Change data) = 75GB
- Calculated Throughput per VBA = 3MB/s
- Backup Window = 10 Hours
- Number of VBAs = (Calculated – See Below)

Formula

VBA Count = (Data Written) / (Throughput) / (Backup Window)

VBA Count
Subsequent Backup

Provisioned Data	3	TB
Data Written	75,000	MB
Calculated Throughput per VBA	3.00	MB/s
Backup Window	10	Hours
❖ Number of VBAs	0.69	VBA

- ❖ In order to accommodate the backup window of 10 hours, one (1) PHD VBA will need to be deployed.
- ❖ The "Effective Throughput" for the above example equates to 120MB/s.

Example #2:

- Total Provisioned Data = 8TB
- Data to be written (Change Data) = 200GB
- Calculated Throughput per VBA = 3MB/s
- Backup Window = 10 Hours
- Number of VBAs = (Calculated – See Below)

Formula

VBA Count = (Data Written) / (Throughput) / (Backup Window).

VBA Count
Subsequent Backup

Provisioned Data	8	TB
Data Written	200,000	MB
Calculated Throughput per VBA	3.00	MB/s
Backup Window	10	Hours
❖ Number of VBAs	1.85	VBA

- ❖ In order to accommodate the backup window of 10 hours, two (2) PHD VBAs will need to be deployed.
- ❖ The "Effective Throughput" for the above example equates to 240MB/s.

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

PHD Virtual™