Implementing Backup Data Replication for Disaster Recovery

Using Rsync to replicate the PHD Virtual backup datastore for disaster recovery

Written by

Frank Musacchia
Director, Technical Services
PHD Virtual

March 2011

Table of Contents

Introduction 3

PHD Virtual Benefits 3

Overview 4

Rsync Overview 4

Grsync Overview 4

PHD Virtual Architecture 5

VBA™ Architecture 5

VBA™ Diagram 5

Backup Datastore 5

Datastore Architecture 5

Hardlinks 6

Rsync Installation/Configuration 6

Installation 6

Basic rsync Syntax 6

Using Cron for Automation 7

Replication Diagram 7

Recommendations 7

Restoring Virtual Machines 8

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 describe leveraging ***rsync*** for performing replication of the PHD Backup Storage Target, to a Secondary Backup Storage Target, whether located locally or at remote Disaster Recovery location. Having a Replica of the Backup Storage Target is helpful in situations where there is a failure of the Primary Storage Target. This will allow for recovery of Virtual Machines as well as Individual files from the Secondary Backup Storage Target, thus allowing organizations to meet their DR and/or Compliance requirements.

Rsync Overview

Rsync is a software application for Unix and Windows systems which synchronizes files and directories from one location to another while minimizing data transfer using delta encoding when appropriate. An important feature of rsync not found in most similar programs/protocols is that the mirroring takes place with only one transmission in each direction. Rsync can copy or display directory contents and copy files, optionally using compression and recursion.

While the rsync algorithm forms the heart of the rsync application that essentially optimizes transfers between two computers over TCP/IP, the rsync application supports other key features that aid significantly in data transfers or backup. They include compression and decompression of data, block by block, using zlib at the sending and receiving ends, respectively, and support for protocols such as ssh that enables encrypted transmission of compressed differential data. Instead of ssh, stunnel can also be used to create an encrypted tunnel to secure the data transmitted. Finally, rsync is capable of limiting the bandwidth consumed during a transfer, a useful feature that few other standard file transfer protocols offer.

In daemon mode rsync listens on the default TCP port of 873, serving files in the native rsync protocol or via a remote shell such as RSH or SSH. In the latter case, the rsync client executable must be installed on both the local and the remote host. Released under the GNU General Public License, rsync is open source software and is widely used. For further detailed information use the following link http://en.wikipedia.org/wiki/Rsync.

Use the following link to download rsync: http://rsync.samba.org/

Grsync Overview

Grsync is a Graphical User Interface (GUI) for the rsync synchronization tool under Linux / Unix System. There is also a port of Grsync on Windows platform. Grsync is released under the GPL license and is open source software, and makes use of the GTK+ UI toolkit. It doesn't support all of rsync features, but can be effectively used to synchronize local directories and supports remote targets in a limited way.

Use the following link to download Grsync: http://sourceforge.net/projects/grsync/

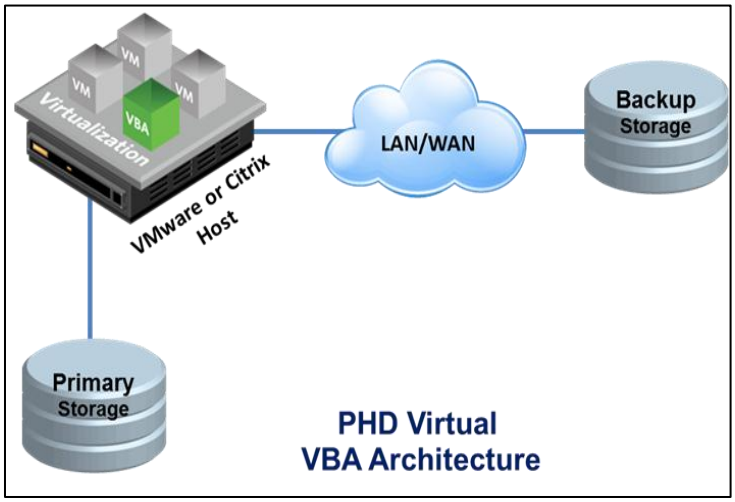
PHD Virtual Architecture

VBA™ Architecture

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

- Deploys as a small virtual appliance on either a VMware vSphere or Citrix XenServer host
- 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



Backup Datastore

PHD Virtual allows the flexible configuration to backup to different types of storage targets, for the backup storage data. The current backup storage targets supported are:

- Virtual Disk (VMDK or VHD file)
- NFS
- CIFS

Datastore Architecture

Within the backup storage target, the following folders comprise the PHD Backup Datastore structure:

- **Backups** - Metadata and hard links for backups
- **Blocks** - Deduplicated block store
- **CBT** – CBT block store
- **Exports** - Temporary files for iSCSI exports

Hardlinks

PHD Virtual utilizes Hardlinks for data mapping on the Backup Storage target and must be preserved when replicating the Backup Storage Target. Hardlinks are listings that contain information about the file. A file can have multiple hardlinks, appearing in multiple directories, but isn't deleted until there are no remaining hardlinks to it.

Rsync supports the copying of hardlinks and is it critical to have this enabled when replicating the PHD Virtual backup datastore for proper data recovery.

To preserve Hardlinks in rsync use the following command:

- --hard links (Preserve hard links)

Rsync Installation/Configuration

Installation

You can install and run Rsync on any one of the following:

1. On a Physical Server
2. On a Virtual Machine
3. On a Storage Array

Basic rsync Syntax

The following is the basic command structure with a handful of basic options attached:

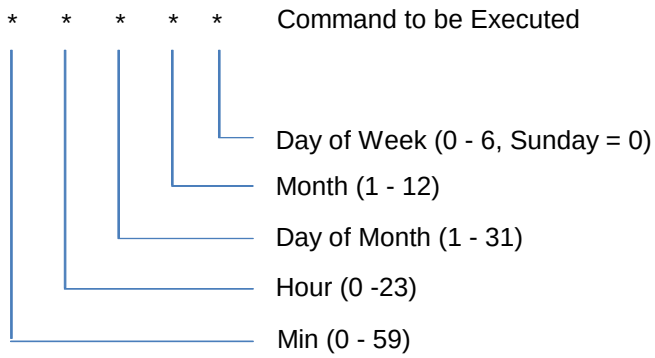
rsync --recursive --links --hard-links --perms --owner --group --devices --times --sparse --delete /mnt/source/ /mnt/dest/

--recursive	(Recursive mode)
--links	(Copy symlinks)
--hard-links	(Preserve hard links)
--perms	(Preserve permissions)
--owner	(Preserve Owner)
--group	(Preserve Group)
--devices	(Preserve device files)
--times	(Preserve modification times)
--sparse	(Handle sparse files efficiently)
--delete	(Delete extraneous files from destination directories)
/mnt/source/	(Source folder, PHD backup directory)
/mnt/dest/	(Destination folder)

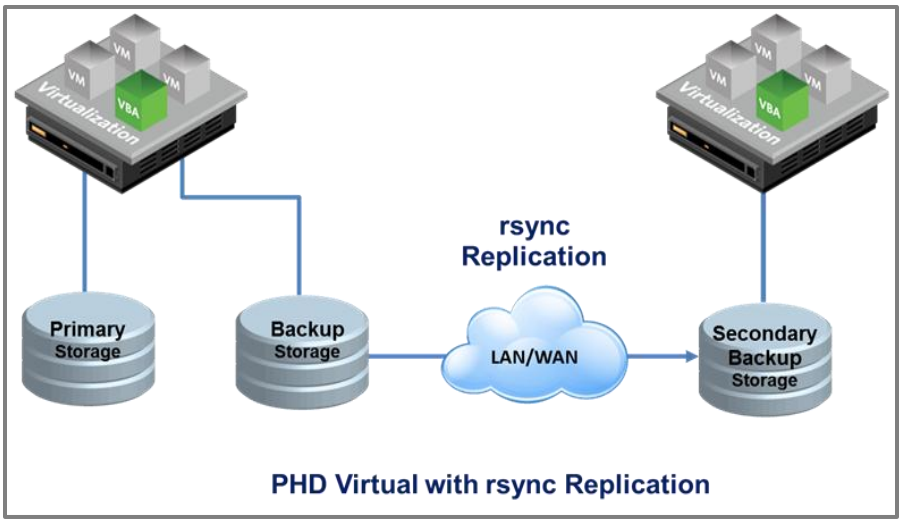
Using Cron for Automation

Cron is a linux utility which can be used for scheduling rsync. An example of a rsync cron job which would allow rsync to start every night at 12:30am would be as follows:

30 0 * * * /usr/bin/rsync --recursive --links --hard-links --perms --owner --group --devices --times --sparse --delete /mnt/source/ /mnt/dest/



Replication Diagram



Recommendations

1. It is highly recommended that the rsync process is started ***"After"*** all PHD backup jobs have completed.
2. The VBA at the Secondary site should stay in a ***"Powered Off"*** state, until it is needed to perform a restore.

Restoring Virtual Machines

In order to restore Virtual Machines and/or Individual Files (FLR) from the Secondary Storage Target, the following step must be taken:

1. The Secondary Backup Storage Target, which contains the PHD replicated datastore, needs to be connected to a Hyperisor Host (VMware or XenServer).
2. A PHD Virtual VBA™ needs to be installed on the Hypervisor Host.
3. The VBA™ needs to have the Secondary Backup Storage Target mounted. This is done by configuring the storage target in the Storage tab of the Configuration settings of the VBA™.
4. The VBA™ needs to be rebooted in order to recognize the Secondary Backup Storage Target.
5. Utilize the Restore Wizard to restore a Virtual Machine or and/or an Individual file.