VMware vSphere® Data Protection™ 5.8

TECHNICAL OVERVIEW
REVISED AUGUST 2014

vmware®

Table of Contents

Introduction

VMware vSphere® Data Protection™ is a backup and recovery solution from VMware. It is fully integrated with VMware vCenter Server™ and VMware vSphere Web Client, providing disk-based backup of virtual machines. vSphere Data Protection is based on the industry-leading EMC Avamar backup and recovery solution and is available in two editions:

- VMware vSphere Data Protection, included with VMware vSphere Essentials Plus Kit and higher
- VMware vSphere Data Protection Advanced, purchased separately

Features and Benefits of vSphere Data Protection

- Wizard-driven setup and management to quickly and easily implement a data protection solution for a vSphere virtual machine environment
- Significantly reduced backup data disk space requirements, with the patented, variable-length Avamar deduplication technology
- Use of VMware vSphere Storage APIs – Data Protection as well as Changed Block Tracking (CBT) to reduce load on the vSphere host infrastructure and minimize backup window requirements
- Agentless virtual machine backup and restore that reduces complexity and deployment time
- Secure, efficient backup data replication to Avamar for offsite data protection
- Direct-to-host emergency restore operation that enables virtual machine recovery even when vCenter Server and vSphere Web Client are offline
- File Level Restore (FLR), which enables granular file and folder restoration without the need for an agent in Microsoft Windows and Linux virtual machines
- Simple Web browser–based administration through vSphere Web Client
- Appliance and backup data protection via a checkpoint-and-rollback mechanism

Additional Features and Benefits of vSphere Data Protection Advanced

- Increased capacity of as much as 8TB of deduplicated backup data storage
- Deployment of external proxies enabling as many as 24 parallel backup operations
- Integration with EMC Data Domain Boost and EMC DD Boost for additional scale, efficiency, and reliability
- Microsoft Exchange Server agent for application-consistent backup and restore of databases and mailboxes, including those protected by a database availability group (DAG)
- Microsoft SQL Server agent that leverages the Virtual Backup Device Interface (VDI) feature for proper backup and restore of databases in standalone configurations and clustered environments
- Microsoft SharePoint agent that enables granular database backup and restore
- Reliable, efficient replication of backup data between vSphere Data Protection Advanced appliances for redundancy and offsite data protection
- Flexibility to restore replicated backup data at both the source and target locations
- Automated backup verification that provides the highest level of confidence in backup data integrity
- In-place upgrade of vSphere Data Protection to vSphere Data Protection Advanced

This paper presents an overview of the architecture, deployment, configuration, and management of vSphere Data Protection and vSphere Data Protection Advanced. Throughout the remainder of this document, vSphere Data Protection and vSphere Data Protection Advanced will be referred to collectively as vSphere Data Protection. Features exclusive to one version or the other will be called out explicitly.

Architectural Overview

vSphere Data Protection requires VMware vCenter Server 5.1 or higher, either the Windows implementation or the Linux-based VMware vCenter™ Server Appliance™. VMware vCenter Single Sign-On™ is also required.

vSphere Data Protection supports backing up virtual machines on vSphere versions 5.0 and later. Web browsers must be enabled with Adobe Flash Player to access vSphere Web Client and vSphere Data Protection functionality. See vSphere documentation for a list of Web browsers currently supported with vSphere Web Client.

vSphere Data Protection is deployed as a prebuilt, Linux-based virtual appliance. A maximum of 20 vSphere Data Protection appliances can be deployed per vCenter Server. Each appliance is deployed by default with four virtual CPUs and 4GB of memory. Storage capacity for deduplicated backup data is configured during deployment.

Optionally, as many as eight external proxies (virtual appliances) can be deployed per vSphere Data Protection Advanced virtual appliance. Proxies can be deployed to enable SCSI HotAdd transport backups of virtual machines running on datastores not directly accessible by the vSphere Data Protection Advanced virtual appliance. Examples include vSphere hosts utilizing local direct attached storage (DAS) and hosts deployed at remote locations. External proxies are required for the Linux logical volume manager (LVM) and EXT4 FLR. Deployment of external proxies is performed using the vSphere Data Protection configure user interface (UI).

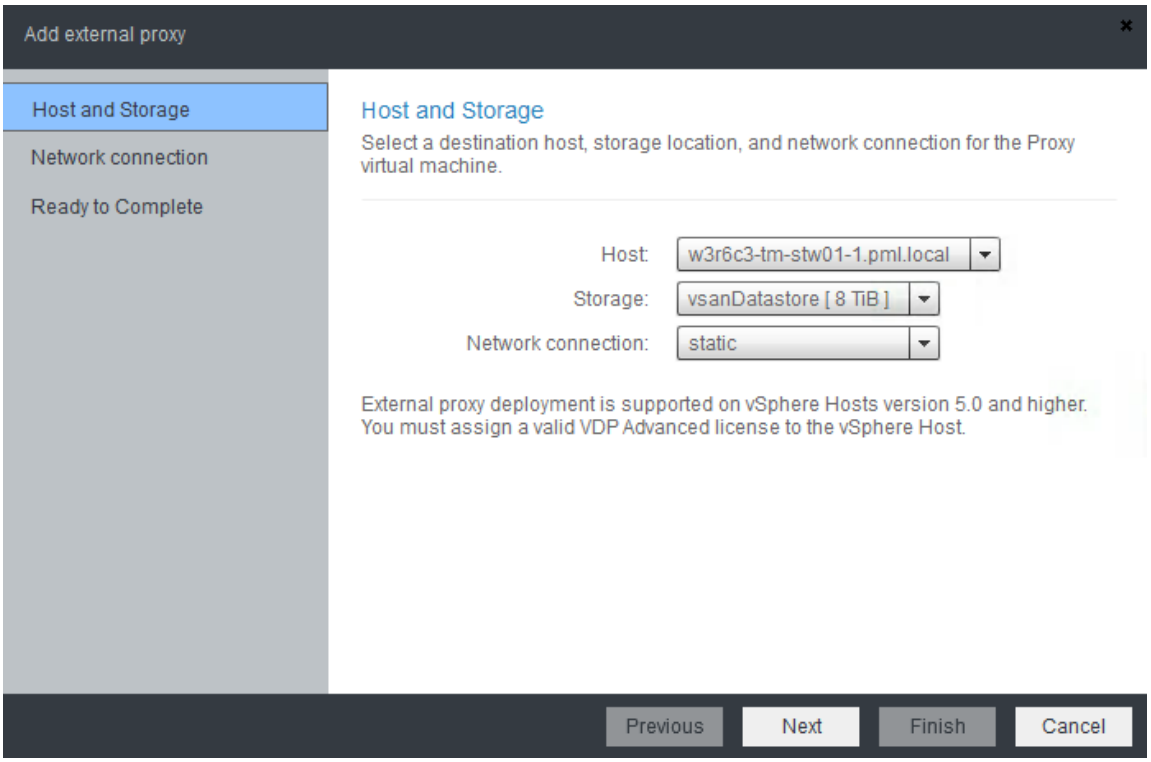

Figure 1. Adding an External Proxy

vSphere Data Protection Advanced application agents are downloaded using vSphere Web Client and are installed in the guest operating system (OS) of the virtual machines running Exchange Server, SQL Server, and SharePoint.

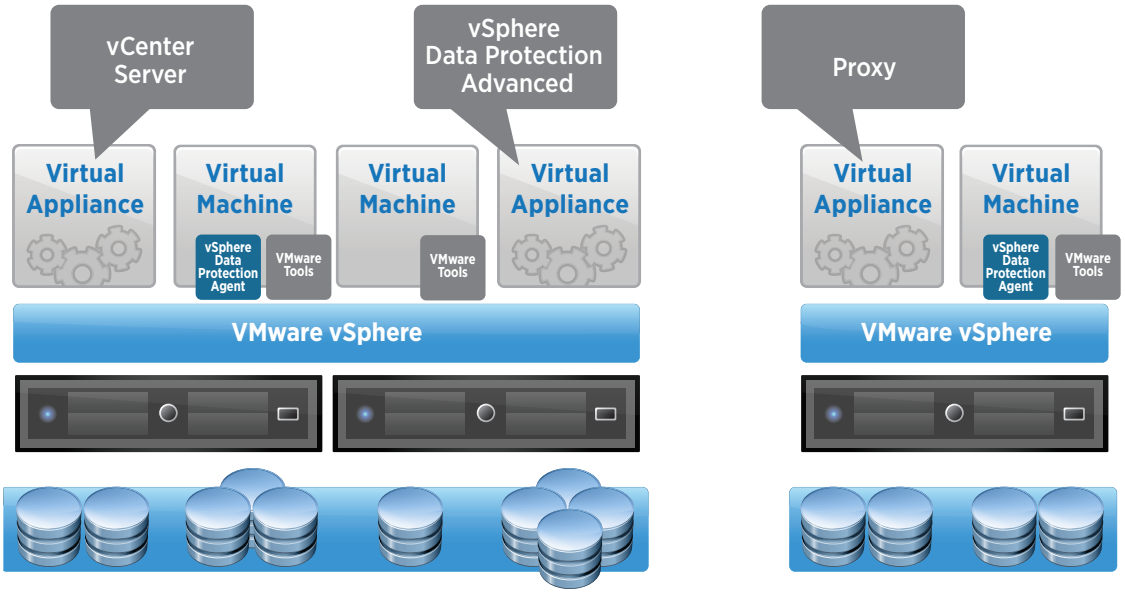

Figure 2. vSphere Data Protection Components

vSphere Data Protection supports as much as 2TB of deduplicated backup data capacity. Assuming average virtual machine sizes, average data change rates, and a 30-day retention policy, approximately 40 to 50 virtual machines can be protected with a 2TB appliance. vSphere Data Protection Advanced scales to 8TB of deduplicated backup data capacity for protection of 150 to 200 virtual machines on average, given the previously mentioned assumptions. Every environment is different, so actual results will vary.

The virtual machine disk (VMDK) files for a vSphere Data Protection virtual appliance can be stored together on the same vSphere datastore or distributed across multiple vSphere datastores. It is also possible to detach VMDK files that make up an existing vSphere Data Protection virtual appliance backup data partition and attach them to a newly deployed appliance.

Extra storage capacity cannot be added after the appliance has been deployed, so users should plan adequately prior to deploying vSphere Data Protection to help ensure proper sizing. In contrast, vSphere Data Protection Advanced enables dynamic provisioning—as much as 8TB of total capacity.

When determining storage capacity requirements, several factors—number of protected virtual machines, amount and formats of data being backed up, retention periods, data change rates, and others—should be considered.

Deployment and Configuration

vSphere Data Protection is deployed using vSphere Web Client from a prepackaged Open Virtualization Archive (OVA) file. The same OVA file is used for both vSphere Data Protection and vSphere Data Protection Advanced deployments. vSphere Data Protection Advanced functionality is enabled by entering a license key either during or after deployment.

After the appliance has been deployed and powered on, a Web browser is used to access the vSphere Data Protection configure utility to perform the initial configuration. The first time a user connects to the vSphere Data Protection configure UI, it runs in "install mode." With the "install mode" wizard, items such as IP address, host name, DNS, time zone, vCenter Server connection information, and storage are configured. A performance storage test can also be run at this time, which is highly recommended to validate that the storage on which vSphere Data Protection is running meets or exceeds recommended performance levels. Upon successful completion of these tasks, the appliance must be rebooted, which will take several minutes as the appliance automatically finalizes its initial configuration.

After initial configuration, the vSphere Data Protection configure utility runs in "maintenance mode." In this mode, it is utilized to perform functions such as starting and stopping services in the appliance, deploying proxies (vSphere Data Protection Advanced only), collecting logs, performing emergency restores, upgrading the vSphere Data Protection appliance, and rolling back the appliance to a previous valid configuration state, which will be discussed later in this document.

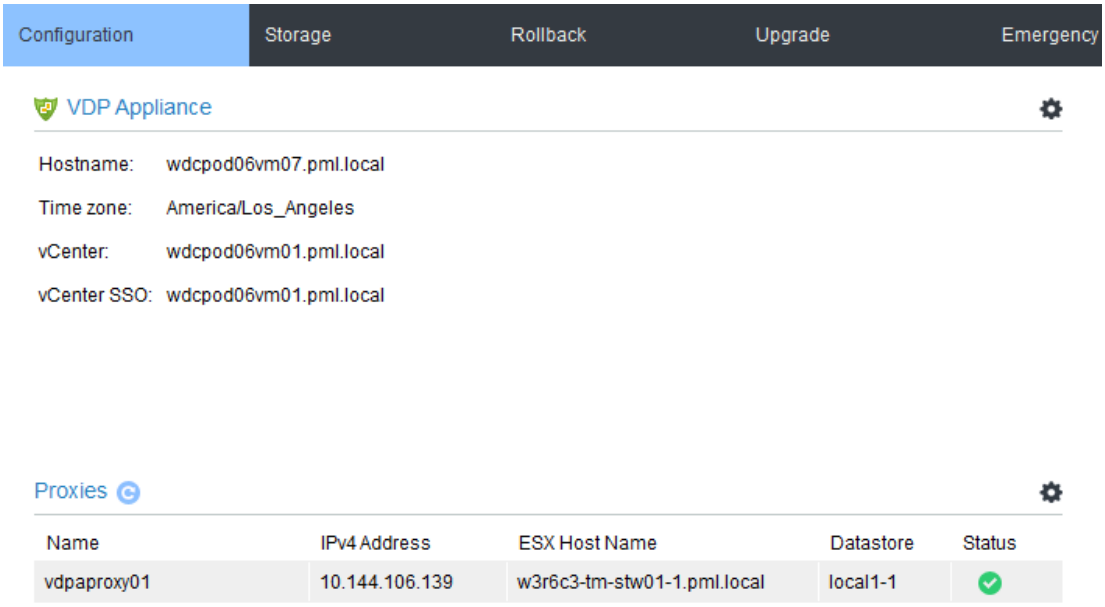

Figure 3. vSphere Data Protection Advanced Configure User Interface Running in Maintenance Mode

Backup

Creating and editing a backup job is accomplished using the **Backup** tab of the vSphere Data Protection UI in vSphere Web Client. Individual virtual machines can be selected for backup. Users also can select specific VMDK files for backup. For example, an administrator might want to put the OS, applications, and data on separate VMDK files. If the requirement is only to protect the data, the administrator can select only the virtual machine disk containing the data, which helps minimize backup capacity consumption. Figure 4 shows the selection of **Hard disk 2** for the virtual machine named **dbserver02**.

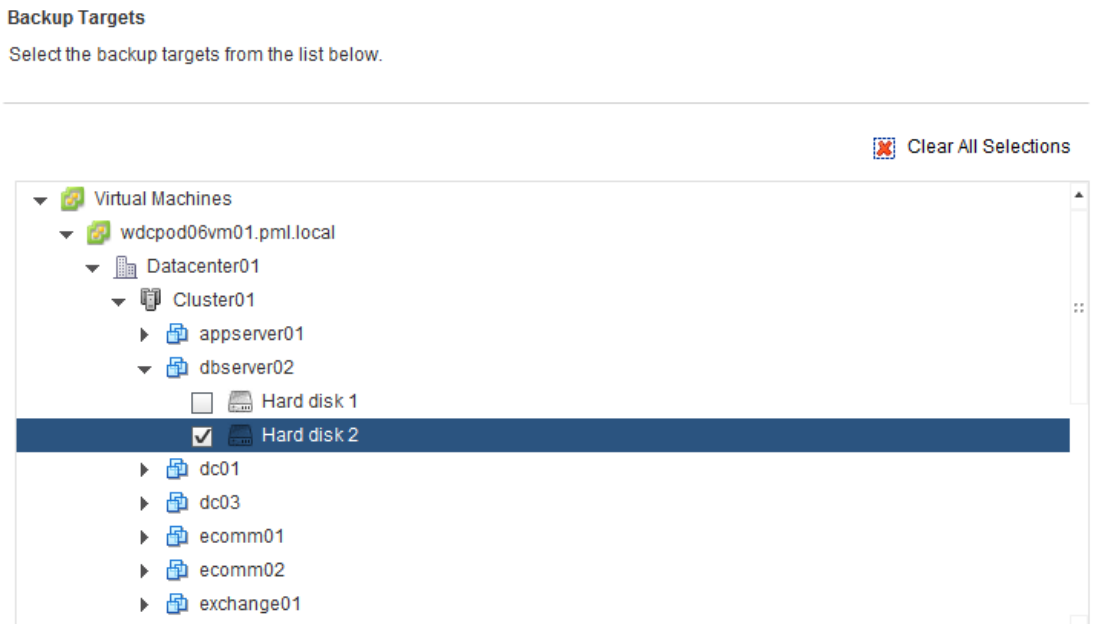

Figure 4. Selecting an Individual Virtual Machine Disk for Backup

Containers of virtual machines such as data centers, clusters, and resource pools can also be selected for backup. When a virtual machine is added to the protected container, it automatically is backed up. Likewise, when a virtual machine is removed from the container, it no longer is included in the backup job. Previous restore points are preserved until expired by the retention policy.

Backup jobs can be scheduled daily, weekly, or monthly. Each job starts at its scheduled time and runs once on the day it is scheduled.

The retention policy can be defined in a few ways; for example, retention for 60 days or until a specific date. A custom retention policy also can be defined.

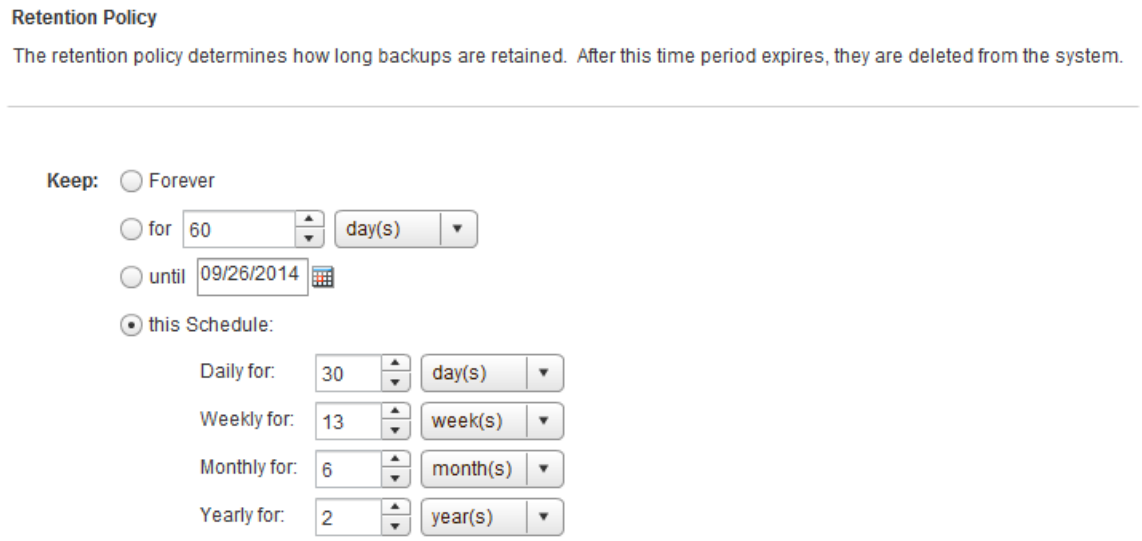

Figure 5. Custom Retention Policy

After a backup job has been created, it can be edited or deleted. It is also possible to clone a backup job. Cloning can be useful if, for example, the backup administrator wants to easily duplicate an existing custom retention policy for a new set of virtual machines.

The initial backup of a virtual machine can take some time because all data blocks that make up that virtual machine must be backed up. Subsequent backups typically take much less time because vSphere Data Protection utilizes CBT in vSphere.

Application Backups

vSphere Data Protection Advanced adds the ability to properly back up Exchange Server, SQL Server, and SharePoint application databases. SQL Server clusters and Exchange Server database availability groups are also supported. A vSphere Data Protection Advanced application agent is installed in the guest OS of each virtual machine running these applications. These agents enable application-consistent backups and provide support for other options such as full, differential, or incremental backups; multistream backups; and database log management.

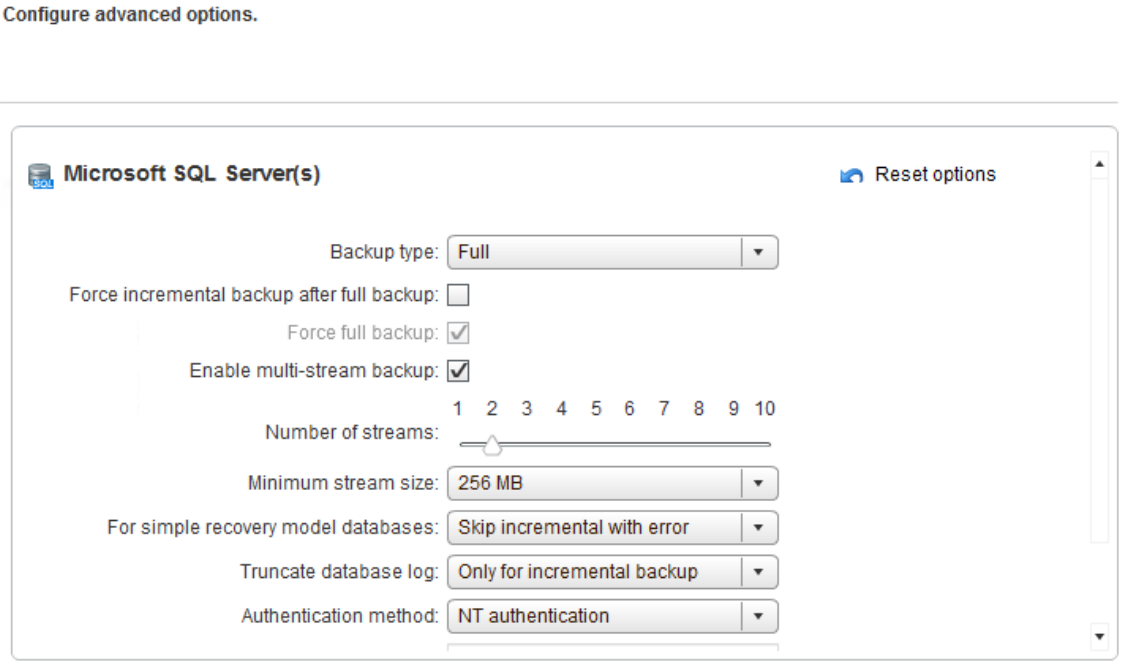

Figure 6. SQL Server Backup Job Options in vSphere Data Protection Advanced

Backup Data Replication

vSphere Data Protection Advanced can replicate backup data between vSphere Data Protection Advanced appliances and to Avamar. This capability is especially useful to move backup data offsite in a secure and reliable manner. Because the backup data is deduplicated at both the source and target, only unique backup data segments are replicated. The replicated data is encrypted and compressed to secure it and to further minimize network bandwidth consumption.

When creating a replication job, it is possible to define specific criteria for which backup data is replicated. Individual clients—virtual machines and applications—can be selected for replication; specific backup types—weekly backups, for example—can be chosen and date restrictions can be defined.

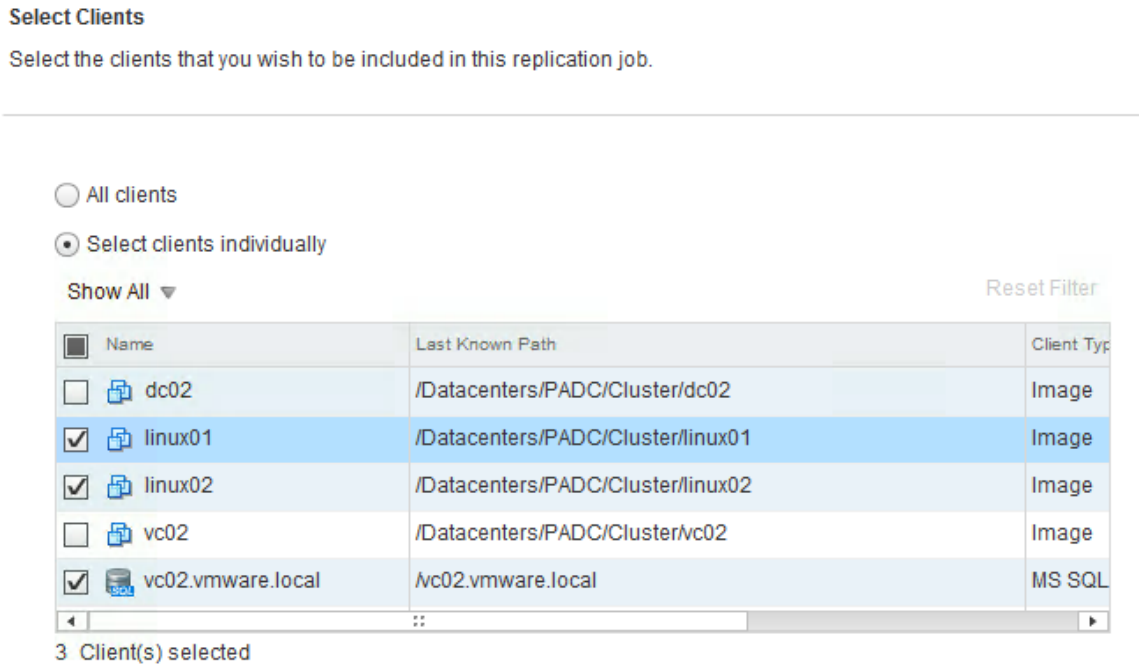

Figure 7. Backup Selection in a Replication Job

The replication job can be scheduled to run daily, weekly, or monthly. By default, the retention policy of the replicated backup data is the same as what was defined in the backup job(s) for that backup data. A different retention policy for replicated backup data can be defined. For example, an administrator might want to retain backup data locally for 30 days and retain the same replicated backup data offsite for 180 days.

There are numerous replication topology options with vSphere Data Protection Advanced. Replication can be one to one or, as shown in Figure 8, a more robust backup and replication strategy can be implemented. Replicated backup data can be "rereplicated"—for example, backing up virtual machines at site A, replicating this backup data to site B, and then replicating the replicated backup data from site B to site C. This approach results in backup data availability at all three sites.

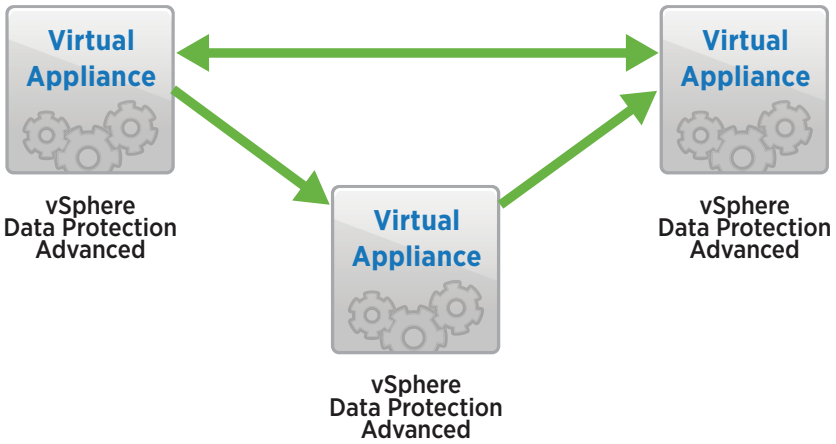

Figure 8. Replication Topology Example

NOTE: vSphere Data Protection—the edition included with vSphere—can replicate backup data only to Avamar.

Restore

Restore an entire virtual machine by using the **Restore** tab in the vSphere Data Protection UI. The administrator can browse the list of protected virtual machines and select one or more restore points. Individual virtual machine disks can also be selected for restore.

vSphere Data Protection offers fast and efficient recovery by leveraging CBT. When restoring an entire virtual machine to its original location, the workloads of both a full image restore and a restore leveraging CBT are evaluated. vSphere Data Protection intelligently determines which method will result in the faster virtual machine recovery time.

It is also possible to restore virtual machines from replicated backup data at the target location and locally. Example scenario: A vSphere Data Protection Advanced virtual appliance protects virtual machines in a primary data center. Backup data is replicated by vSphere Data Protection Advanced from the primary data center to a vSphere Data Protection Advanced virtual appliance at a disaster recovery data center. Disaster strikes the primary data center; virtual machines, including the vSphere Data Protection Advanced virtual appliance, are lost. When the primary data center is back online, a new vSphere Data Protection virtual appliance is deployed and connected to vSphere Data Protection at the disaster recovery site. The new vSphere Data Protection virtual appliance can perform restores at the primary data center by using backup data in vSphere Data Protection at the disaster recovery site.

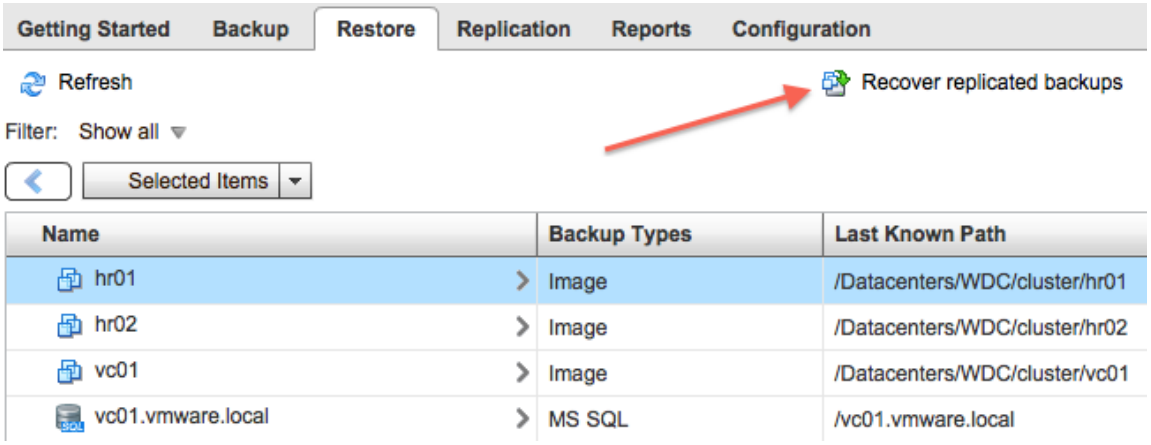

Figure 9. Recover Replicated Backups

File Level Restore

With vSphere Data Protection, it is also possible to restore individual files, folders, and directories within a virtual machine. An FLR operation is performed using a Web-based tool called vSphere Data Protection Restore Client. The process enables end users to conduct restores on their own, without the assistance of an administrator, by selecting a restore point and browsing the file system as it looked at the time that backup was done. They locate the item(s) to be recovered, select a destination for the restored items, and start the recovery. The progress of the restore job can be monitored in vSphere Data Protection Restore Client.

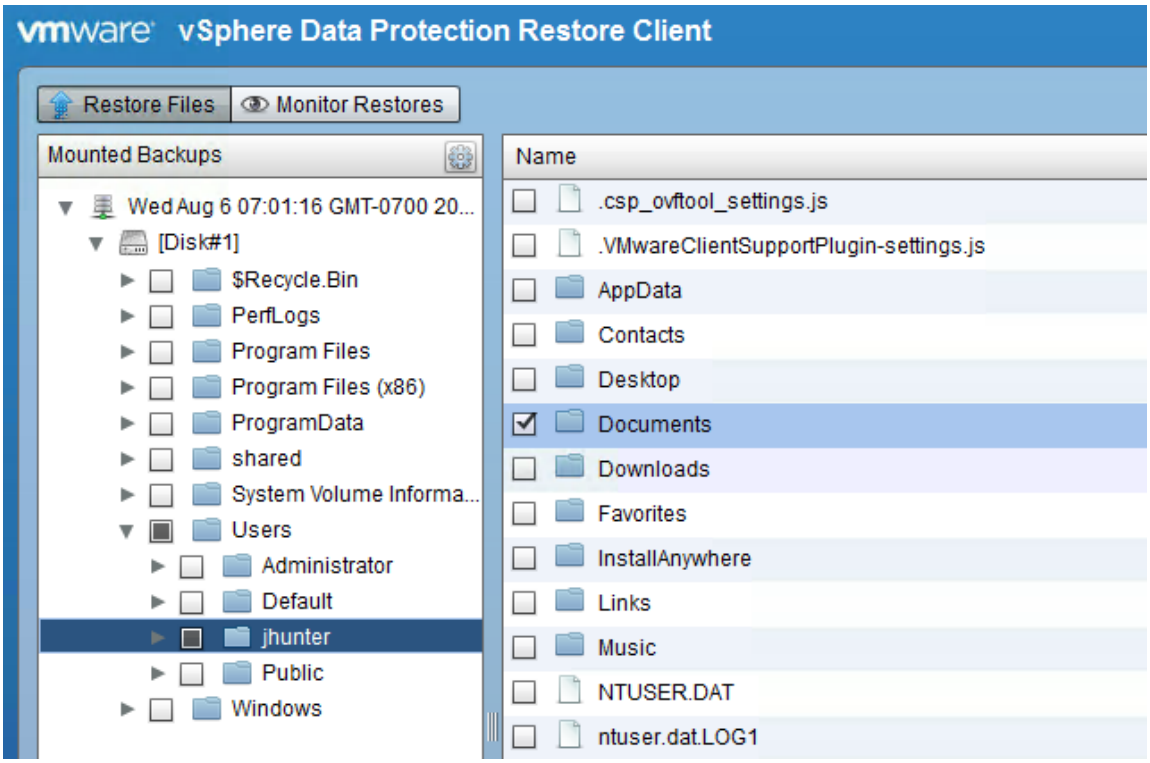

Figure 10. vSphere Data Protection Restore Client

Application Restore

vSphere Data Protection Advanced provides the ability to restore individual SQL Server, Exchange Server, and SharePoint application databases. SQL Server clusters and Exchange database availability groups are supported.

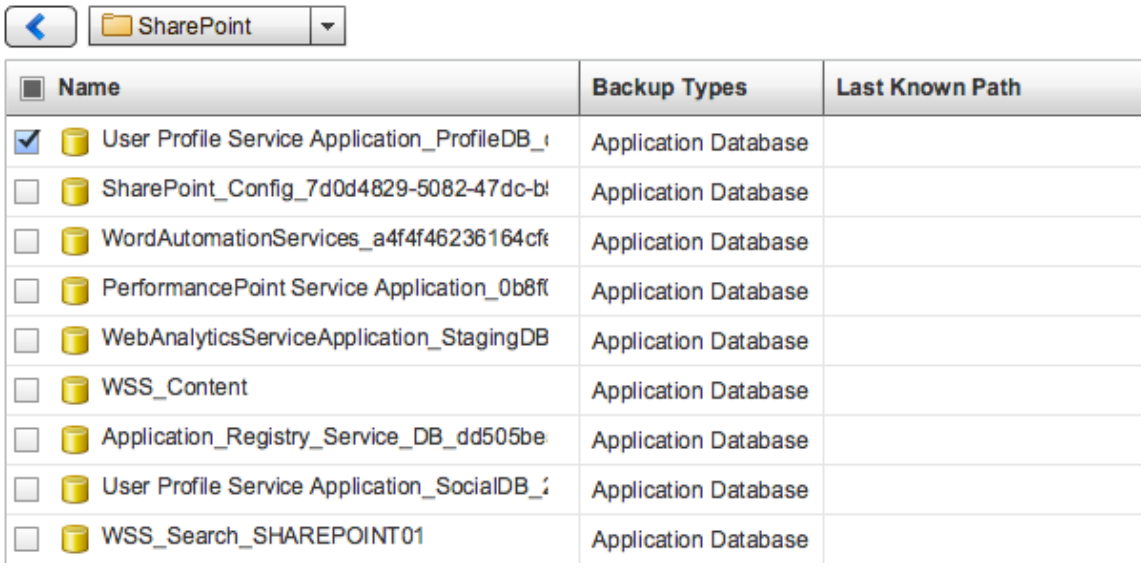

Figure 11. SharePoint Database Selected for Recovery in vSphere Data Protection Advanced

Because vSphere Data Protection Advanced leverages specific application agents, various restore options can be defined. Figure 12 shows options available with the SQL Server agent.

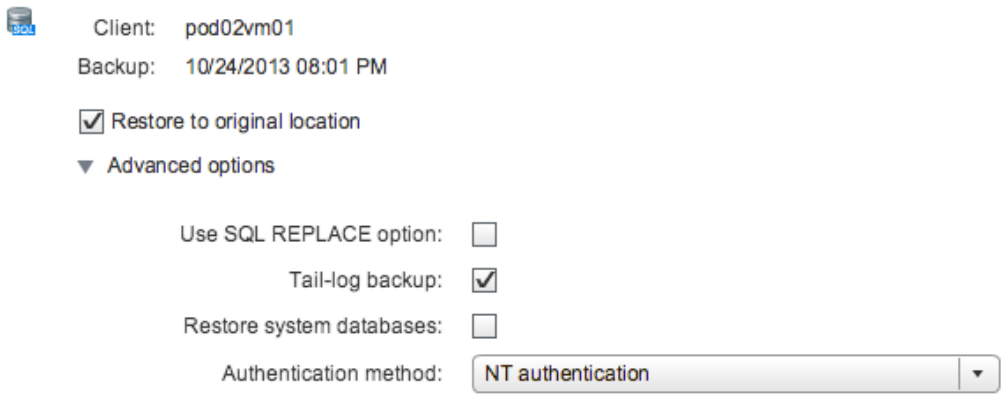

Figure 12. SQL Server Restore Options in vSphere Data Protection Advanced

The vSphere Data Protection Advanced agent for Exchange Server enables individual database backup and restore. It is also possible to select individual mailboxes for restore. A user's mailbox is recovered as a folder named "Recovered Items." The user can then browse the folder's contents by using an Exchange Server client such as Microsoft Outlook to retrieve any needed items.

Automated Backup Verification

Backup verification jobs can be created in vSphere Data Protection Advanced. These jobs automate the process of restoring a virtual machine: powering it on; verifying the guest OS booted, by detection of VMware Tools™ "heartbeats"; and, optionally, confirming an application started successfully by means of a custom script. The restored virtual machine is disconnected from the network to prevent interference with production systems. After the restore and verification have been completed, the restored virtual machine is deleted to free up capacity. Backup verification jobs can be scheduled at specific times daily, weekly, or monthly.

Direct-to-Host Emergency Restore

vCenter Server and the vSphere Web Client server must be online to perform restores using vSphere Web Client. When these components are offline, an emergency restore can be utilized to restore a virtual machine directly to the host on which vSphere Data Protection is running. **Emergency Restore** is one of the tabs in the vSphere Data Protection configure UI.

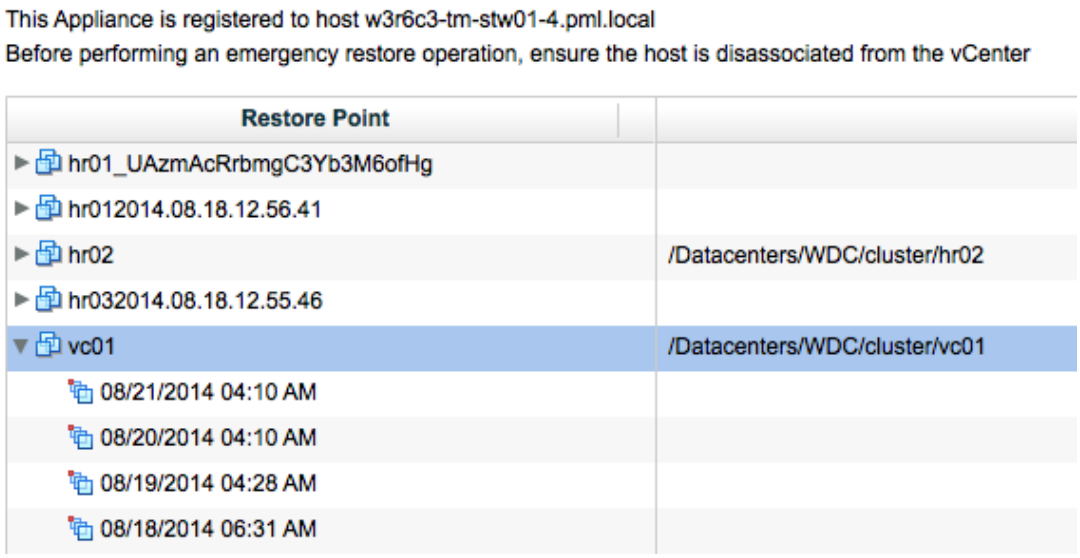

Figure 13. Direct-to-Host Emergency Restore

Reporting

The **Reports** tab in vSphere Data Protection and vSphere Data Protection Advanced displays a variety of information: appliance status, used capacity, backup and replication job details, and so on. If there were errors during a job, clicking **Task Failures** enables an administrator to view specific information about the failures, including client logs. The **Job Details** section provides information about backup, replication, and backup verification jobs. The list of clients—protected virtual machines and applications—can be filtered to quickly locate a specific client. A list of clients that have not been backed up can also be viewed by clicking **Unprotected Clients**. All three views can be exported to comma-separated values (CSV) files.

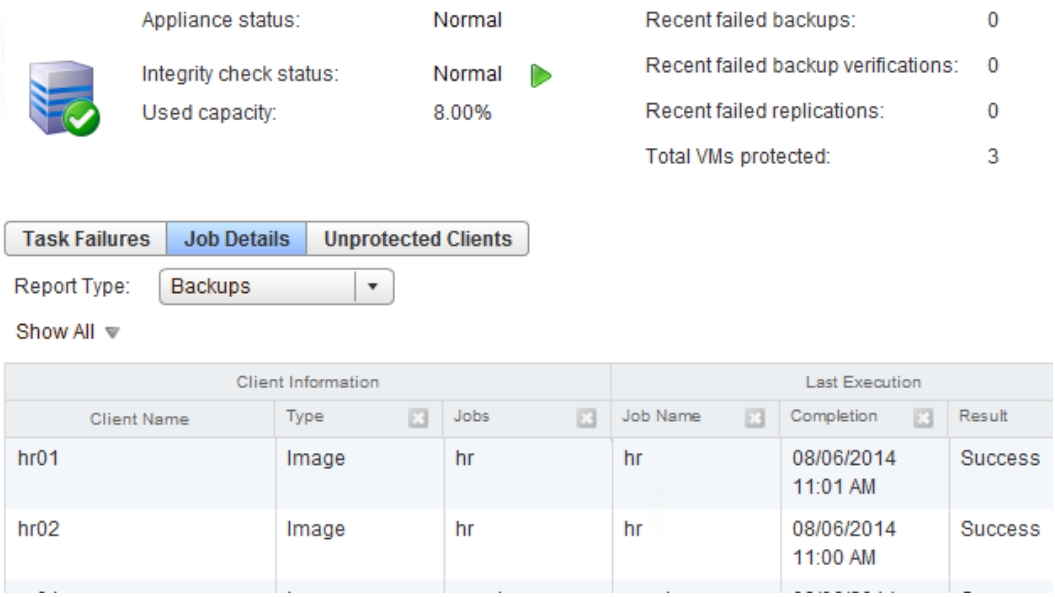

Figure 14. Reports Tab in the vSphere Data Protection UI

In addition to having UI reporting capabilities, vSphere Data Protection can be configured to send email reports scheduled at a specific time, once per day on any or all days of the week. Similar to the UI, these email messages contain details on the vSphere Data Protection appliance, backup jobs, and protected virtual machines.

Integration with EMC Data Domain

vSphere Data Protection Advanced can be configured to use Data Domain as a backup data target. The advantages of doing this include additional scale, global deduplication, and improved backup efficiency due to the DD Boost libraries built into vSphere Data Protection Advanced.

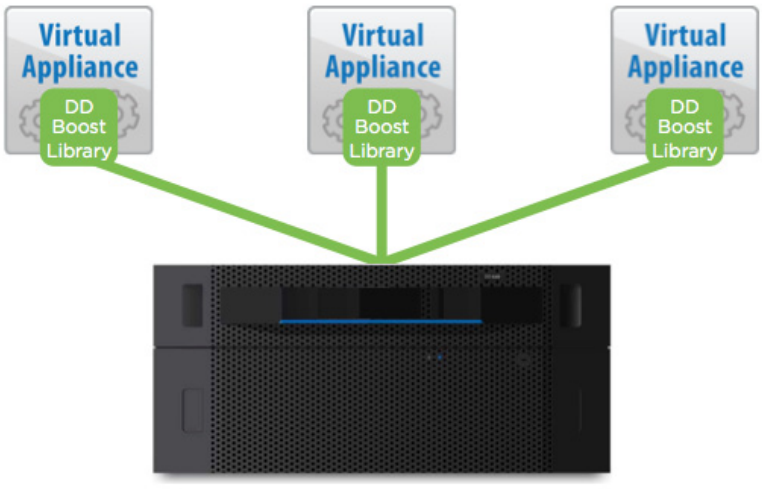

Figure 15. Multiple vSphere Data Protection Advanced Appliances with Data Domain

Backup data replication can be configured between two or more vSphere Data Protection Advanced appliances with separate Data Domain targets. Replication is configured and backup metadata is maintained in vSphere Data Protection Advanced, but replication of the backup data occurs directly between the Data Domain appliances. Data at the source and target is already deduplicated. Only unique data segments must be replicated, minimizing the amount of network utilization. Replicated data is encrypted and compressed for additional security and efficiency.

Avoiding Backup Data Corruption

vSphere Data Protection and vSphere Data Protection Advanced contain a checkpoint-and-rollback mechanism. A checkpoint is a system-wide backup of the vSphere Data Protection appliance that is performed to help protect the appliance from risks that might cause data corruption, such as an unexpected appliance power-off. In this case, the appliance would roll back to the last validated checkpoint. Any backup jobs performed after that checkpoint would be lost, but data corruption—that is, loss of all backup information—likely would be avoided.

Summary

Data protection is a key component of any business continuity plan. VMware vSphere Data Protection provides an efficient solution for protecting a VMware virtual machine infrastructure. VMware vSphere Data Protection Advanced adds more deduplicated backup data capacity as well as the capability to properly protect mission-critical applications using agents. Backup data can be securely and efficiently replicated to an offsite location for disaster recovery purposes. vSphere Data Protection Advanced can be integrated with EMC Data Domain. Deployment is quick and simple. Administration is easily performed using VMware vSphere Web Client.

About the Author

Jeff Hunter is a senior technical marketing architect at VMware with a focus on business continuity and disaster recovery solutions. He has been with VMware for more than six years, prior to which he spent several years implementing and administering VMware virtual infrastructures at two Fortune 500 companies.

Follow Jeff on Twitter: @jhuntervmware

VMware, Inc. 3401 Hillview Avenue Palo Alto CA 94304 USA Tel 877-486-9273 Fax 650-427-5001 www.vmware.com

Copyright © 2014 VMware, Inc. All rights reserved. This product is protected by U.S. and international copyright and intellectual property laws. VMware products are covered by one or more patents listed at http://www.vmware.com/go/patents. VMware is a registered trademark or trademark of VMware, Inc. in the United States and/or other jurisdictions. All other marks and names mentioned herein may be trademarks of their respective companies. Item No: VMW-WP-vSPHR-Dta-Prot-5.8-USLET-102 Docsource: OIC-FP-1189