


# Microsoft® System Center

## Data Protection for the Hybrid Cloud

Shreesh Dubey, Vijay Tandra Sistla, Shivam Garg, Aashish Ramdas
Mitch Tulloch, Series Editor

www.dbooks.org

# Visit us today at



## microsoftpressstore.com

- **Hundreds of titles available** – Books, eBooks, and online resources from industry experts
- **Free U.S. shipping**
- **eBooks in multiple formats** – Read on your computer, tablet, mobile device, or e-reader
- **Print & eBook Best Value Packs**
- **eBook Deal of the Week** – Save up to 60% on featured titles
- **Newsletter and special offers** – Be the first to hear about new releases, specials, and more
- **Register your book** – Get additional benefits



Microsoft




Get the latest news from Microsoft Press sent to your inbox.

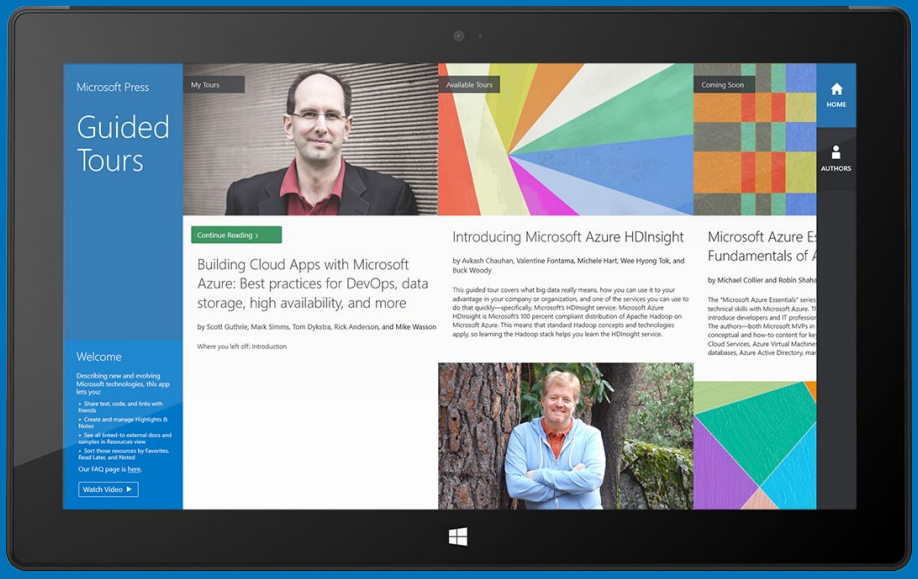
- New and upcoming books
- Special offers
- Free eBooks
- How-to articles

Sign up today at MicrosoftPressStore.com/Newsletters

Microsoft

www.dbooks.org

# Wait, there's more...




Find more great content and resources in the **Microsoft Press Guided Tours** app.

The Microsoft Press Guided Tours app provides insightful tours by Microsoft Press authors of new and evolving Microsoft technologies.

- Share text, code, illustrations, videos, and links with peers and friends
- Create and manage highlights and notes
- View resources and download code samples
- Tag resources as favorites or to read later
- Watch explanatory videos
- Copy complete code listings and scripts

Download from Windows Store

PUBLISHED BY
Microsoft Press
A division of Microsoft Corporation
One Microsoft Way
Redmond, Washington 98052-6399

Copyright © 2015 by Microsoft Corporation All rights reserved.

No part of the contents of this book may be reproduced or transmitted in any form or by any means without the written permission of the publisher.

Library of Congress Control Number
ISBN: 978-0-7356-9583-2

Printed and bound in the United States of America.

First Printing

Microsoft Press books are available through booksellers and distributors worldwide. If you need support related to this book, email Microsoft Press Support at mspinput@microsoft.com. Please tell us what you think of this book at http://aka.ms/tellpress.

This book is provided "as-is" and expresses the author's views and opinions. The views, opinions and information expressed in this book, including URL and other Internet website references, may change without notice.

Some examples depicted herein are provided for illustration only and are fictitious. No real association or connection is intended or should be inferred.

Microsoft and the trademarks listed at http://www.microsoft.com on the "Trademarks" webpage are trademarks of the Microsoft group of companies. All other marks are property of their respective owners.

**Acquisitions Editor:** Karen Szall
**Developmental Editor:** Karen Szall
**Editorial Production:** Megan Smith-Creed
**Copyeditor:** Megan Smith-Creed
**Cover:** Twist Creative • Seattle

www.dbooks.org

# Contents


Introduction vi

**Chapter 1 Data protection trends and challenges ..... 1**

Data growth trends ..... 1
Data protection scenarios and challenges ..... 2
Emergence of the public cloud ..... 2

**Chapter 2 Overview of hybrid cloud backup ..... 5**

Cloud design point for backup ..... 5
Azure Backup ..... 7
Microsoft workloads and enterprise client backup ..... 8
Hyper-V virtual machine backup at CPS scale ..... 8
De-duplication of backup storage ..... 9
System Center integration ..... 10

**Chapter 3 Protecting Microsoft workloads ..... 11**

Basic configuration ..... 11
- Adding disks to a DPM storage pool ..... 11
- Installing DPM agents ..... 12
- Configuring protection groups ..... 13

VSS framework ..... 14
Hyper-V protection ..... 15
- Hyper-V backup process ..... 16

Client protection ..... 17
- Client data recovery ..... 18

Exchange Server protection ..... 18
SQL Server protection ..... 18
- SQL Server backup process ..... 20
- Self-service recovery of SQL Server databases ..... 20

SharePoint protection ..... 21
- SharePoint backup process ..... 22
- SharePoint catalog ..... 23
- Scoped consistency check ..... 23
- SharePoint recovery ..... 23

**Chapter 4 Protecting Azure IaaS workloads ..... 27**

Setting up DPM in Azure ..... 27
- Creating a new VM for DPM ..... 28
- Joining the VM to a domain ..... 29


**What do you think of this book? We want to hear from you!**

**Microsoft is interested in hearing your feedback so we can continually improve our books and learning resources for you. To participate in a brief online survey, please visit:**

**http://aka.ms/tellpress**

Adding backup storage ... 29
Installing and configuring DPM and Azure Backup ... 31
Reviewing the post-deployment architecture ... 32
Protecting workloads ... 33
Discovering servers and installing the agent ... 33
Discovering workloads and creating a Protection Group ... 34
Workloads and configurations supported for backup ... 35
Considerations for performance and scale ... 35
Recommendations for better performance ... 36
Scaling up vs. scaling out ... 36
Tiering data to Azure Backup ... 38

**Chapter 5 Protecting Hyper-V virtual machines ... 39**

Customer scenarios and challenges ... 39
Planning for VM backup ... 40
What to back up (host level vs. guest level) ... 40
When to back up ... 40
Where to back up ... 41
How to back up ... 42
How to control costs ... 43
Protecting Hyper-V VMs ... 43
Protecting Hyper-V over SOFS ... 44
Protecting Hyper-V over CSV ... 46
Continued protection with VM migration ... 48
Protecting replica VMs ... 48
Protecting servers in workgroups and untrusted domains ... 49
Recovering Hyper-V data ... 50
How to restore a file from a VM ... 50
How to restore a VM ... 51
Recommendations ... 53
Case study: Real-world customer ... 54

**Chapter 6 VMware private cloud protection ... 55**

This information is not yet publicly available. It will be included when this ebook is re-issued in summer 2015.

**Chapter 7 Protecting the Microsoft Cloud Platform System ... 57**

Protecting the management cluster ... 57
Default protection policy ... 58
Recovering VMs and databases ... 60
Recovering from failures of management cluster features ... 63
Protecting tenant VMs ... 66
Using DPM servers for tenant backup ... 67
Adding tenant VMs to backup ... 67
Recovering tenant VMs ... 68
Monitoring backups ... 71
Case study: A real-world CPS customer ... 72



www.dbooks.org

**Chapter 8 Optimizing backup storage ..... 73**

Exponential growth in backup storage ..... 73
Containing the cost of backup storage ..... 74
Software approaches to reducing stored backup data ..... 74
Azure Backup ..... 76
Data Protection Manager ..... 76
Using deduplication with DPM ..... 76
High-level deployment architecture and constraints ..... 77
Understanding the backup-deduplication software stack ..... 77
Deduplication benefits: A real-world scenario ..... 78

**Chapter 9 Integration with System Center ..... 79**

Management and monitoring scenarios and challenges ..... 79
Enterprise reporting capabilities ..... 80
Management and monitoring solutions ..... 81
SLA-based alerts ..... 82
Client Auto Deployment ..... 84

**Chapter 10 Integration with Azure Backup ..... 87**

Advantages of Azure Backup ..... 87
Backup scenarios ..... 88
Tape replacement ..... 88
Branch office backup ..... 89
Windows client backup ..... 89
Protection of Microsoft Azure assets ..... 89
Getting started with Azure Backup ..... 90
Azure Backup capabilities ..... 91
Expanded workload support ..... 91
Long-term retention ..... 92
Offline seeding of initial replica ..... 94

**Chapter 11 Protecting Azure IaaS virtual machines ..... 95**

Why back up Azure VMs? ..... 95
Tradeoffs with VM backup ..... 96
Azure Backup vs. on-premises backup ..... 96
How VM backup in Azure works ..... 97
The backup extension ..... 97
Data transfer ..... 98
Learn more ..... 99

**What do you think of this book? We want to hear from you!**

**Microsoft is interested in hearing your feedback so we can continually improve our books and learning resources for you. To participate in a brief online survey, please visit:**

**http://aka.ms/tellpress**

# Introduction

If you are responsible for architecting and designing the backup strategy for your organization, especially if you're looking for ways to incorporate cloud backup into your business continuity scenarios, this book is for you. With the increasing trends in virtualization as well as the move to the pubic cloud, IT organizations are headed toward a world where data and applications run in on-premises private clouds as well as in the public cloud. This has key implications for data protection strategy, and it is important to choose the solution that provides the same level of data protection you have afforded so far while allowing you to harness the power of the public cloud.

We will cover how the Azure Backup service has evolved into a first-class platform-as-a-service (PaaS) service in Microsoft Azure that integrates with the on-premises enterprise class backup product, System Center Data Protection Manager (DPM), to provide a seamless hybrid cloud backup solution. Current backup products treat the cloud as a storage endpoint, which we see as a limited-use case for the public cloud. The approach we describe in this book allows you to exploit the full power of the public cloud and gives you the flexibility to manage your backups in a hybrid world.

We have made a steady set of investments in DPM over the last 18 months, and, as of this writing, we have released six update rollups, including customer hot fixes as well as new features in the areas of private cloud protection, storage optimization, and workload support. The last chapter focusses on the most recently released protection for infrastructure-as-a-service (IaaS) virtual machines, which went to preview release in March 2015 and is expected to be generally available by Q3 of calendar year 2015.

This book covers improvements added in DPM 2012 R2 as well as the integration with Microsoft Azure Backup service and assumes you have working knowledge of the DPM 2012 version. To get familiar with older versions of DPM, refer to the following:

- *http://social.technet.microsoft.com/wiki/contents/articles/7485.system-center-data-protection-manager-2012.aspx*
- *http://social.technet.microsoft.com/wiki/contents/articles/11867.system-center-2012-data-protection-manager-survival-guide.aspx*
- *http://blogs.technet.com/b/dpm/*



www.dbooks.org

# Acknowledgments

The authors would like to thank the following individuals for their help on this book project:

- The entire hybrid cloud backup engineering team, without whom the products this book talks about wouldn't be possible
- John Loveall and the Windows Server Deduplication team for excellent collaboration to make de-duplication work for DPM
- Vijay Tewari, Jim Pinkerton, and other folks on the CPS team for making DPM the certified backup product for CPS
- Corey Sanders and Guy Bowerman from the IaaS VM team for collaborating on the IaaS VM backup feature
- Mitch Tulloch for being a great coach, for being patient about the numerous delays and schedule resets, and above all, for helping us build a book that we are really proud of and one we think will really help the community on backup

# Free ebooks from Microsoft Press

From technical overviews to in-depth information on special topics, the free ebooks from Microsoft Press cover a wide range of topics. These ebooks are available in PDF, EPUB, and Mobi for Kindle formats, ready for you to download at:

*http://aka.ms/mspressfree*

Check back often to see what is new!

# Errata, updates, & book support

We've made every effort to ensure the accuracy of this book and its companion content. You can access updates to this book—in the form of a list of submitted errata and their related corrections—at:

*http://aka.ms/Scdatapro/errata*

If you discover an error that is not already listed, please submit it to us at the same page.

If you need additional support, email Microsoft Press Book Support at mspinput@microsoft.com.

Please note that product support for Microsoft software and hardware is not offered through the previous addresses. For help with Microsoft software or hardware, go to *http://support.microsoft.com*.

## We want to hear from you

At Microsoft Press, your satisfaction is our top priority, and your feedback our most valuable asset. Please tell us what you think of this book at:

*http://aka.ms/tellpress*

The survey is short, and we read every one of your comments and ideas. Thanks in advance for your input!

## Stay in touch

Let's keep the conversation going! We're on Twitter: *http://twitter.com/MicrosoftPress*.



www.dbooks.org

CHAPTER 1

# Data protection trends and challenges

Digital data is more important than ever before, and organizations depend on their IT departments to handle the deluge of data that gets generated in modern IT environments. While some of this data is ephemeral and does not need long-term protection, most of the data still needs traditional forms of protection.

This chapter starts by examining trends in data growth and key inflection points. It then explores these inflection points and how they affect scenarios related to data protection and management. Lastly, the chapter looks at the emergence of the public cloud and the unique opportunities and challenges that it poses for data protection.

## Data growth trends

Per estimates made by IDC and EMC (see *http://www.emc.com/leadership/digital-universe/2012iview/executive-summary-a-universe-of.htm*), digitally created data is expected to double roughly every two years and reach an astounding 40,000 exabytes by 2020. While the cost-per-gigabyte of storage is coming down, it is clearly not keeping up with the rate of data growth. In the data protection domain, the contribution to data growth comes from the need for traditional forms of protection. The 3-2-1 rule summarizes this perfectly—3 copies of data on at least 2 different forms of storage with at least 1 copy that is offsite.

**NOTE** **For more information on the 3-2-1 rule, see** *http://www.40tech.com/2012/06/12/follow-the-3-2-1-backup-rule-to-safeguard-your-files/*.

At the same time, other changes are sweeping through the datacenter. Per a study by EMC in 2013 (see *https://education.emc.com/content/_common/docs/articles/Managing_Storage_Trends_Challenges_and_Options_2013_2014.pdf*), server virtualization reached a critical inflection point in 2012 when the storage deployed for virtualized deployments surpassed the storage deployed for physical deployments. Virtualized deployments will accumulate storage steadily to reach about 50 percent of all storage deployed by 2015, as more and more organizations transition their storage from their physical

deployments. Like with any infrastructure inflection, it presents an opportunity for IT decision makers to re-examine their datacenter management tool-chain, and clearly backup technologies are one of the major considerations resulting from this inflection. The other important datacenter trend is that while private and public cloud deployments are on the rise, the majority of deployments are still on-premises. This means infrastructure management technologies such as backup need to support a hybrid deployment environment.

## Data protection scenarios and challenges

The growth in data and the change in infrastructure deployments in datacenters is a multi-dimensional problem that IT professionals need to deal with in their organizations. The 2013 study by EMC also surveyed IT professionals across 800 organizations about their pain points around storage. The top two challenges articulated were (in order):

1. Managing data storage (79 percent of respondents)
2. Designing, deploying, and managing backup, recovery, and archive solutions (43 percent of respondents)

These challenges span across deployments—physical, virtual, private cloud, or public cloud. Data protection products need to adapt to new rules:

- Deal exceptionally well with exponential growth in data and management of data at scale.
- Provide innovative new ways of bringing down the per-gigabyte cost of storage.
- Thrive in a hybrid world for the immediate foreseeable future in enterprise IT.

## Emergence of the public cloud

The public cloud has been transforming IT by a combination of exceptional technology improvements and attractive economics. Public cloud companies, notably Amazon and Microsoft, are accelerating their datacenter build-out and aggressively offering compute/storage/network at compelling prices. Every enterprise is looking for scenarios where they "offload" their infrastructure to the public cloud and reduce their CAPEX spend. According to a CIO of an IT service provider:

> *"If you don't adopt the public infrastructure and move your value up the chain to some other differentiator, your competitor will, and you will be soon filing for Chapter 11."*

This is changing the way companies think about their IT infrastructure, and public cloud is a central theme in strategic initiatives for most companies. A key driver of public cloud adoption is the backup scenario, and traditional products are evolving to embrace the public cloud. This is happening for two reasons:



www.dbooks.org

- Backup is the largest consumer of on-premises storage. In addition, backup also requires a tape-based infrastructure for long-term retention. The combined storage requirement can easily exceed several petabytes, and the total cost of ownership is very attractive with the public cloud when compared to an on-premises deployment.
- For infrequently accessed data, the risk associated with storing data on the public cloud is within acceptable limits.

Looking back at the challenges faced by IT professionals around backup and backup-related storage, it is clear that the public cloud is uniquely poised to address the problems:

- It reduces the per-gigabyte cost of long-term storage by eliminating the CAPEX and OPEX associated with tape infrastructure. The cloud offers better economics and faster, more reliable restores. Companies can reduce their on-premises footprint by keeping minimal retention (for example, one week) on local disks and moving the rest of the data to the cloud for longer retention.
- The per-gigabyte cost of backup also decreases for backup of client operating systems if the public cloud is used instead of infrastructure that is deployed on-premises. The affordability of traditional backup solutions is one of the key reasons organizations don't deploy protection for their PCs (laptops and desktops), and the public cloud now makes this scenario affordable.
- Backup for branch offices becomes more cost effective by leveraging backup to the cloud. Typically, backup from branch offices is consolidated at the head office, and this involves additional network and storage infrastructure to handle the high-volume backup traffic. Backing up data directly to the cloud reduces the overhead of deploying costly infrastructure just for backup.
- Backup products that have a public cloud element can leverage the public cloud to deal with the problem of management at scale. The public cloud gives anywhere access for management and monitoring of backups, especially to end customers to perform self-service restores.

That being said, there are some concerns that need to be addressed for any scenario (like backup) that leverages the public cloud:

- **Security** The key question here is, "How secure is my data in the cloud?" Most cloud-based services offer some form of data encryption for the data to address this concern.
- **Network** The availability of bandwidth, the latency, and the cost of sending data over the wire are the factors that affect a customer's decision. Most of the countries/regions in the continents of North America and Europe have ubiquitous and affordable network connectivity. However, in many countries/regions, network bandwidth is not available or is quite expensive.

- **Availability** Reliance on a third-party cloud provider implies a set of risks around data availability, and even leading service providers like Amazon and Microsoft Azure have been hit with widespread service outages in the recent past. Cloud service providers offer service level agreements and compensation caused by down time which mitigates this to some extent.

When done right, backup to the cloud can be a game-changer for enterprises. The remaining chapters explore how Microsoft's hybrid cloud backup solution (Azure Backup and Microsoft System Center Data Protection Manager) addresses the challenges and mitigates the concerns.



www.dbooks.org

CHAPTER 2

# Overview of hybrid cloud backup

This chapter provides a high-level overview of Azure Backup and Microsoft System Center Data Protection Manager (DPM), which constitute the Microsoft hybrid cloud backup solution. Azure Backup is a hybrid platform-as-a-service (PaaS) service that leverages the Azure bottomless storage capacity to provide offsite backups and integrates with DPM, which is the enterprise-grade, on-premises backup solution included in the System Center Suite. While the emphasis is on the newer features, this chapter also covers other aspects of the hybrid cloud backup solution at a high level. Subsequent chapters drill down into some important areas in greater detail.

## Cloud design point for backup

Most backup products have evolved their cloud strategy by treating the cloud as just another storage medium, like tape or disk, for the backup data. Furthermore, for all experiences on the data protected in the cloud, either the data needs to be restored back to the on-premises datacenter or customers need to run the backup software in an infrastructure-as-a-service (IaaS) virtual machine (VM) in the cloud to deliver the experiences. While this has some advantages in terms of simplicity and ease of deployment, it has cost implications and is a limited-use case for the ultimate value that the public cloud offers.

Microsoft has taken a different approach by building a multi-tenanted PaaS backup service in Microsoft Azure called Azure Backup, which integrates seamlessly with the on-premises DPM backup product to provide an end-to-end solution optimized for the cloud. This approach leverages elastic storage in Azure as well as on-demand compute services to build rich backup experiences in a very economical, pay-as-you-go model. Advantages include:

- Anywhere, anytime access for managing and monitoring backups and self-service restores
- Efficient cloud storage architecture that provides low-cost, resilient data storage
- On-demand scaling up or down depending of ingestion of data
- Rich data services in Azure, like item-level restores and testing backups
- Consistent way to back up on-premises, hybrid, and IaaS/PaaS deployments

Table 2-1 shows an illustrative example of the advantages of a PaaS backup service over a backup application running in an IaaS VM.

**TABLE 2-1** Comparison of a PaaS backup service and a backup application running in an IaaS VM

| ITEM COMPARED | BACKUP APP IN IAAS A4 VM | PAAS BACKUP SERVICE |
|---|---|---|
| Maximum backup storage | 16 TB (max possible storage in IaaS A4 VM) | Unlimited |
| Compute costs borne by customer | $536/month | Included |
| Data transfer cost for restores | $0.087/GB to $.181/GB | Included |

***See also*** *For more information on data transfer pricing, see http://azure.microsoft.com /en-us/pricing/details/data-transfers/.*

DPM is recognized in the industry as a best-in-class enterprise backup, and together with Azure Backup, it provides a compelling hybrid cloud backup solution for three key classes of data:

- Enterprise client protection (PCs and desktops).
- Host-level VM backups for Microsoft Hyper-V, more popularly known in the backup domain as "agentless virtual machine backups."
- Workload-aware backup for Microsoft SQL Server, Microsoft SharePoint Server, and Microsoft Exchange Server. These workloads typically run on a physical server, but they could also be running on a guest operating system in a Hyper-V VM, VMware VM, or Azure IaaS VM.

The DPM roadmap also includes support for host-level VMware VM backup, to have parity with the backup that is supported for Hyper-V VMs. Future workloads that will be supported are Oracle and IBM DB2, both running on Linux. The goal is to diversify and become a truly heterogeneous enterprise backup solution.

Subsequent sections will describe features in Azure Backup and DPM, specifically:

- Overview of Azure Backup PaaS service
- Workload (SQL Server, Exchange Server, SharePoint Server) and Enterprise Client backup
- Enhancements to support Hyper-V VM backup at Cloud Platform System (CPS) scale
- De-duplication support using Windows De-Duplication capability
- Integration with System Center



www.dbooks.org

# Azure Backup

Azure Backup is a multi-tenanted hybrid PaaS service built in Azure to provide reliable, secure offsite backups in Azure. It can back up on-premises data in Azure coming either from Windows-based servers or from DPM servers as well as provide backups for IaaS or PaaS applications running in Azure. It offers the following key features:

- **Security** All Azure backups are encrypted at source, during transmission, and stored encrypted in Azure. The encryption key is stored at source and is the only way to restore any of the data stored in Azure, so the customer is in full control of access to the data in the service.
- **Reliability** Customers have the choice of using locally redundant storage (LRS), where three copies of the data are created in the Azure datacenter, or geo-replication storage (GRS), where an additional three copies are created at a geo-separated datacenter to ensure data is highly resilient and available even in the case of an Azure-level disaster.
- **Long-term retention** When customers consider Azure as a backup location, they typically want to use it as a long-term retention target driven by compliance and/or disaster recovery requirements. Traditionally, this is done using tape backups, but Azure provides a more compelling alternative, and there are various studies available that show total cost of ownership for a cloud backup is superior to tape backups in the long run.
- **Simplicity** It has a familiar interface that can scale to protect a few servers and to handle large-scale server deployments. Future enhancements include a central management that enables customers to centrally manage all the backup infrastructure.
- **Efficiency** After the initial copy is seeded in Azure, incremental backups are taken as per the backup policy, compressed, encrypted, and sent to Azure where they are stored as incrementals. This optimizes network utilization during transfer as well as storage consumption in Azure, both of which are key factors to keep in mind when sending data to Azure.
- **Cost effectiveness** Azure Backup pricing includes a per-instance backup management fee and the cost of storage consumed on Azure. A key value offered with Azure Backup is that restore operations do not incur egress charges. This information is updated frequently, so please visit the product page (*http://azure.microsoft.com/en-us/pricing/details/backup* ) for updated pricing information.

***See also*** *For more information on how to determine if cloud backup is right for your servers, see http://www.gartner.com/reprints/microsoft?id=1-2CV21H1&ct=150402&st=sb.*

## Microsoft workloads and enterprise client backup

DPM provides workload-aware backup for top Microsoft workloads, namely Exchange Server, SharePoint Server, and SQL Server. This includes working closely with the workload team to jointly validate backup methodology in all configurations and ensuring they are fully supported both from the backup as well as workload perspective. Following is a list of the key workload protection features, some of which are recent additions in DPM 2012 R2:

- Added support for workloads running on the VMware platform. This is the first step towards providing a heterogeneous backup solution.
- Automated discovery of backup artefacts that need to be protected. Exchange and SharePoint have very complex deployments with multiple servers and SQL instances to be protected, all of which is automated.
- Backup for highly available deployments for Exchange Database Availability Group and SQL AlwaysOn, which typically require higher levels of data protection.
- Integration with Volume Shadow Service (VSS) to ensure full-fidelity backups with recovery point objective (data loss tolerance) as low as 15 minutes.
- Granular restore capability, such as mailbox recovery for Exchange Server, database-level recovery for SQL Server, and item-level recovery for SharePoint Server.

On the Client Backup front, support was added in Azure Backup to directly protect Windows Clients (desktops and clients) in Azure. There has been wide acceptance of this feature since customers like the fact that they can reduce their on-premises storage infrastructure and leverage Azure for it, but still take advantage of the enterprise scale management for client machines, such as central management, enforcement of compliance policies, auto-provisioning for new machines, and so on.

## Hyper-V virtual machine backup at CPS scale

As the adoption of Hyper-V picked up in the enterprise segment, there has been a focused effort to provide best-in-class support for protecting Hyper-V VMs. These features started with the System Center 2012 R2 release and subsequent update rollups aligned with the Windows Server 2012 R2 releases. With these features, DPM is the most performant, robust, and scalable Hyper-V VM backup solution in the market currently. These are some of the key features:

- Leverages Hyper-V native VHD snapshots in concert with the host-level VSS infrastructure to reduce the overall impact on the system and provide efficient backups for VMs.
- Certified to protect a fully provisioned CPS "stamp" consisting of four racks, 8,000 VMs, and 32 DPM servers with a daily VM backup at a backup SLA goal of 99 percent. (Backup SLA implies one successful daily recovery point for every VM in CPS.)



www.dbooks.org

- Support for running DPM in a virtualized environment. DPM supports a scale-out architecture, which requires deployment of multiple DPM servers to handle large-scale DPM deployments. Having the ability to run them virtualized was a key enabler for this. Some of the large DPM deployments have over 100 DPM servers in a single datacenter deployment.
- Automated provisioning, management, and monitoring for managing the VM backup process in a hosted environment for CPS. These were released as Windows PowerShell-based runbooks customized for the CPS deployment and currently only available as part of CPS. It is on the roadmap to release them for a non-CPS Windows Server 2012 R2 Hyper-V environment.
- Support for scale-out file server (SOFS) that uses storage spaces and just a bunch of disks (JBOD) to create low-cost storage alternatives, ideal for deploying as backup storage.
- Support for file-consistent backup of Linux guests running on Hyper-V without needing to take them offline. Service providers typically run a combination of Windows and Linux guest operating systems, so providing a consistent and uniform backup policy is essential.
- Ability to specify backup and consistency-checking time windows to provide backup administrators more control of the backup process.

## De-duplication of backup storage

Backup storage is one of the top consumers of storage infrastructure, so storage optimization techniques such as compression and de-duplication have always been priorities for backup IT administrators. De-duplication involves locating duplicate blocks of storage and replacing them with a reference and a single instance of the duplicate block. Depending on the workload that is writing to the storage and the block sizes used to perform the de-duplication, storage savings can range anywhere from 50 to 90 percent

In typical enterprise storage deployments, backup storage is provisioned on SAN devices, which have built-in block-level de-duplication capabilities, and DPM works seamlessly in this deployment. With Storage Spaces and SOFS in Windows Server 2012 R2, customers can create commodity storage built natively on a Windows-based server and JBODs, which can be a viable alternative to traditional SANs. In this deployment, it is important for DPM to inter-operate with the native Windows de-duplication in Windows Server 2012 R2. There are some deployment best practices that need to be followed to ensure maximum storage savings, and these are covered in more detail in a subsequent chapter.

One more note about different kinds of de-duplication when it comes to backup storage. There are generally two ways to do this—inline and offline. Inline de-duplication is done as part of the backup process where every backup set is de-duplicated as it is stored in the

backup storage. Offline de-duplication is done as a post-processing step after the backup has completed. De-duplication is an I/O- and performance-intensive operation, so the number of backups running and the size of the backup storage pool determines which approach works better. For large scale deployments with a large number of backups, offline de-duplication gives better backup throughput (number of backup jobs/hour). It also affords better storage savings since you can aggregate the de-duplication across a backup data sources set as opposed to individual data sources in inline de-duplication. The approach used in DPM is offline de-duplication that provides excellent throughput and storage savings up to 70 percent in large-scale private cloud deployments.

## System Center integration

System Center is the Microsoft enterprise management suite, and the backup process is one of the key fabric management activities provided by DPM, which is included in this suite. As part of this integration, there is support for some scenarios that drive high customer value propositions:

- VM mobility, a key benefit of virtualization, creates management overhead to ensure all housekeeping activities, such as backup, continue to run seamlessly. DPM integrates with the Virtual Machine Manager to ensure it can perform initial discovery as well as ongoing tracking to ensure all scheduled backups continue to work seamlessly during and after migration scenarios.
- A central console that leverages System Center Operations Manager to manage and monitor multiple DPM servers. This is the recommended best practice to monitor backups in an enterprise environment, which typically has multiple DPM servers.
- A new reporting framework based on the Operations Manager data warehouse that collates alerting and monitoring data across multiple DPM servers and allows customers to create custom reports using simple SQL queries. A reporting monitoring management pack is available separately to set up the process of updating the data warehouse and some sample queries to generate a sample dashboard. Since de-duplication is done at the Windows File Server level, a separate file server management pack is available that integrates to provide de-duplication savings on the backup storage. This is a fundamentally different approach, where all monitoring data is documented and updated in the Operations Manager data warehouse on a regular basis so IT administrators can generate custom reports very easily.



www.dbooks.org

CHAPTER 3

# Protecting Microsoft workloads

Most organizations rely on Microsoft server workloads to run their businesses. Ensuring protection of Microsoft workloads is a critical part of the business continuity strategy, and organizations need a backup tool that ensures their workload data is protected from various kinds of data-loss scenarios. With Microsoft System Center Data Protection Manager (DPM), organizations can protect workloads such as Microsoft SQL Server, Microsoft SharePoint Server, Microsoft Exchange Server, and Hyper-V virtual machines (VMs) not only to disk or tape media, but also to Microsoft Azure, Microsoft's cloud platform. This chapter describes how DPM orchestrates backup and recovery of Microsoft workloads.

## Basic configuration

To start protecting Microsoft workloads with DPM, it is essential to get a high-level overview of how protection is configured. DPM discovers the set of data sources to protect, enables the selection of the data sources, and protects the data to disk, tape, or Azure. DPM also orchestrates the recovery of Microsoft workloads from the backup media to the production server or an alternate location as specified.

After DPM is installed, there are essentially three steps for configuring protection:

- Configure the replica storage to store all the backup data.
- Install DPM agents on each of the servers that need to be protected.
- Create a protection group to pick the data sources and configure the backup schedule and backup storage target for the protection group.

### Adding disks to a DPM storage pool

To take advantage of built-in Windows Server de-duplication capability, it is recommended that you deploy the DPM server in a virtualized environment. In a virtualized environment, virtual disks can be added to the VM from the Hyper-V Manager. Add the disk to a storage pool using the instructions available at *https://technet.microsoft.com/en-in/library/hh758075.aspx*.

## Installing DPM agents

If the servers to be protected are in the same domain as DPM, the agent install can be pushed from the DPM administrator console. If the servers to be protected are in a different domain or behind a firewall, the agent must be installed separately on each server and attached to the DPM server. Complete the following steps to install the agent from the DPM administrator console:

1. Open DPM.
2. Click the Management tab located at the bottom-left corner of the console, and then click the Agents hyperlink.
3. Click Install to launch the Protection Agent Installation Wizard. There are two options: Install Agents or Attach Agents (for servers that are behind the firewall or for computers in a workgroup or untrusted domain). Click Install Agents and select the set of target servers on which to install the agents. If the workload to be protected spans multiple servers or a cluster, the agent must be installed on each server or node of the cluster.




4. Provide inputs such as user name, password, and domain name. The user must have administrator privileges on the target computer.
5. Click Restart Method and complete the agent installation.

***See also*** *For agent installation in different domains, follow the steps outlined in the TechNet documentation at https://technet.microsoft.com/en-us/library/bb870934.aspx.*

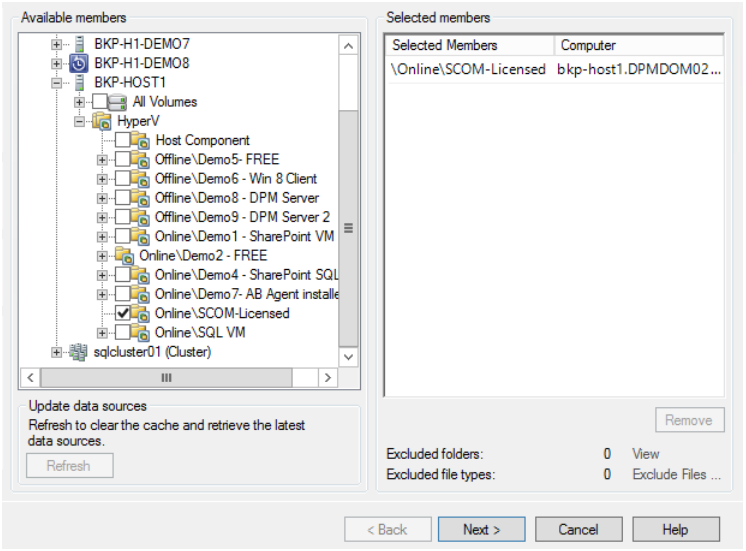


www.dbooks.org

## Configuring protection groups

With the agents installed on the servers that need to be protected, you can configure protection for the workloads. Complete the following steps to configure protection groups.

1. Open DPM.
2. Click the Protection tab, and then click New at the top-left corner.
3. Select the server to protect. DPM uses Active Directory to identify the server and the clusters in the domain that it is part of.
4. When the server is identified, select the specific set of data sources, referred to as members, to back up. You can select multiple types of data sources in a single protection group, but it is recommended that you segregate data sources based on their type and protection goals since a protection group is a means to logically group data sources that have the same protection intent.




5. Select the data protection method. DPM supports short-term retention to disk as well as to tape. For short-term retention, specify the retention range as well as synchronization frequency. For long-term retention goals, select tape or Azure Backup. If a tape library is configured, the tape is shown as a valid target for long-term backup. Similarly, if Azure Backup is configured, you can select Azure as an online protection target. Subsequent chapters in this book cover how Azure Backup can be configured and used as a long-term retention target.
6. DPM carves out replica storage for each protection group. Select Co-Locate Data Source On The Same Disk if the type of data is a Hyper-V VM or client computer or SQL Server databases to gain storage efficiencies.

7. After the storage pool is selected, select the mechanism to transfer the initial copy of the data to the DPM server.
8. Select the consistency check frequency. A consistency check can run only when the replicas become inconsistent or on a daily schedule. With the latest update rollup improvements in DPM 2012 R2, the maximum duration for a consistency check job can be specified. Because the consistency check job consumes additional IOPS in the production machine, this capability is critical for administrators who want to limit the spill-over of the consistency check job past the backup window.
9. After all inputs are provided for short-term protection, a new protection group is configured and can be monitored from the Monitoring tab.
10. The protection group can be modified later to add new data sources, to modify the backup schedule or retention policy, or to add or modify long-term protection goals.

**NOTE** **When you select data sources, it is critical that you select the right set of members for enabling protection. For instance, when selecting clustered resources, do not pick data sources from individual nodes; instead, point to the cluster to select data sources. Similarly, when you select a SharePoint data source, it is critical to point the selection to the web front end server as opposed to the back end SQL Server instance machine.**

## VSS framework

DPM leverages Volume Shadow Copy Services (VSS) and filter bitmaps to make the backup process efficient. DPM leverages file filter technology to maintain a bitmap of changes between two synchronization events. With VSS volume snapshots, the set of changed blocks as tracked by the bitmap are read and transferred from the production server to DPM replica storage. All major Microsoft workloads, such as Hyper-V, SQL Server, Exchange, and SharePoint, support VSS writers, and the Generic VSS writer enables third parties to participate in the VSS protocol. Figure 3-1 illustrates the VSS architecture.



www.dbooks.org

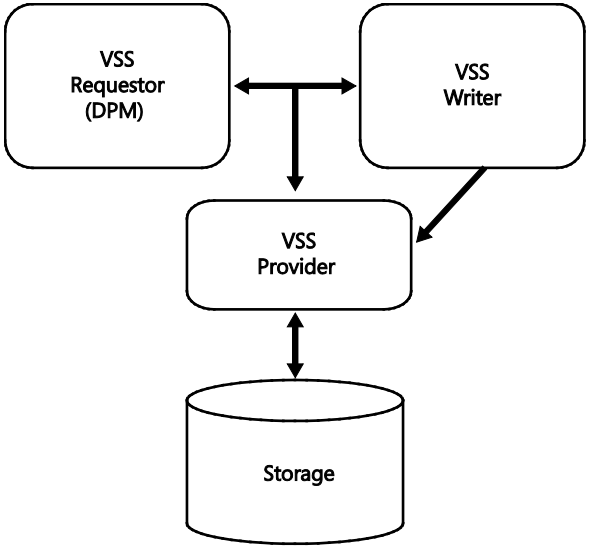



**FIGURE 3-1** VSS architecture diagram

# Hyper-V protection

DPM 2012 R2 UR3 Hyper-V protection features improved scalability and reliability to meet customer backup SLA. VMs can be protected to disk, to tape, or to Azure Backup for long-term retention. Hyper-V protection includes protection of Windows as well as Linux VMs in both standalone and clustered environments. Details on how Hyper-V protection is configured are described in subsequent chapters.

In a clustered environment, a protected VM can move to another Hyper-V host within the same cluster without its storage, or the VM's storage can migrate without the VM compute node, or both the compute and the storage can migrate.

With System Center Virtual Machine Manager (VMM) integration, DPM is able to continue protection of a VM during live migration without requiring user intervention as long as the DPM agent is installed on all the target Hyper-V hosts.

1. Install the VMM console on the DPM server and associate the VMM server with the DPM server from DPM PowerShell running in Administrator mode:

   ```
   Set-DPMGlobalProperty -DPMServerName <> -KnownVMMServers <>
   ```

2. Start the DPM-VMM Helper Service from the control panel.

When the protected VM's storage is migrated, a consistency check is required to make the data on the production server consistent with the data on the replica storage. This is because the file filter tracking information on the production server is not migrated along with the protected VM's VHD files, thereby causing a break in the tracking logic.

## Hyper-V backup process

The Hyper-V backup process includes host-side features as well as guest-side features (see Figure 3-2).

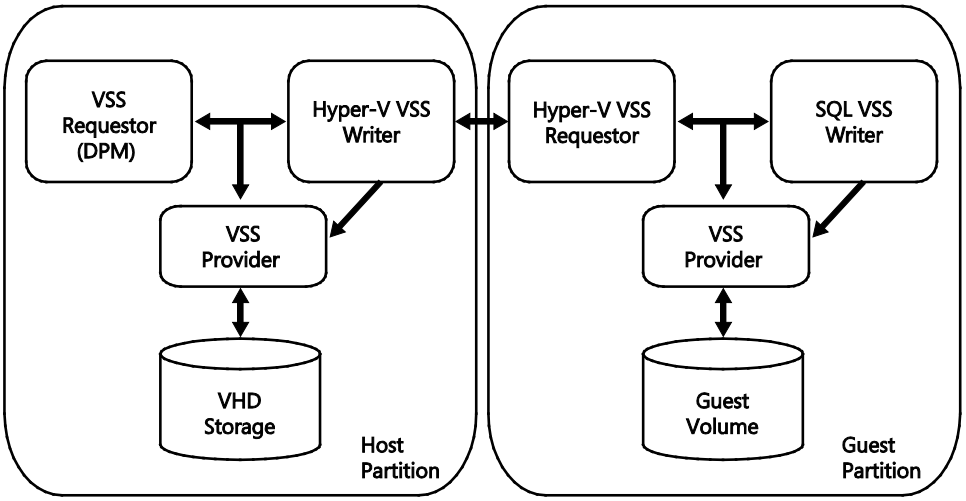



**FIGURE 3-2** Hyper-V backup process

The host side includes Hyper-V VSS writer, the VSS provider, and the backup application, which acts as the VSS requestor. DPM acts as the VSS requestor, which initiates the backup operation periodically and triggers the Hyper-V VSS writer to quiesce the VM.

To enable an application-consistent snapshot of the application running within the VM, Hyper-V communicates to the guest operating system through the Hyper-V integration service. Hyper-V inside the guest acts as the VSS requestor and requests the workloads to quiesce. For instance, if SQL Server is running inside the VM, the SQL Server writer participates in the VSS protocol and flushes in-flight data buffers to the disk, and when it is done, the VSS provider takes a volume snapshot.

After the volume snapshots have been created inside the guest, the VSS provider on the host creates a shadow copy of the volume that contains the VHDs. The volume snapshot enables the application, the VM in this case, to continue to make changes to the VHDs that are attached to the VM while the backup operation takes place.

DPM maintains a bitmap file per VHD that is being tracked, and it can easily identify the blocks that have been modified since the last synchronization event. The block size is maintained as 16 KB to optimize the amount of data that is transferred and stored on the replica. Since the bitmap only indicates which 16-KB blocks are modified, the DPM backup agent reads the modified data from the storage snapshot. These changes are then transported to the DPM replica server and stored on the replica volume. To maintain versions of backup data, DPM maintains recovery points, which are essentially snapshots of the data on the replica storage pool. When a new recovery point is created on the DPM replica storage, a volume



www.dbooks.org

snapshot is taken, and only new changes are written to the replica volume, while shadow copies are maintained for older recovery points. This mechanism of synchronization is called an express full backup.

Because DPM maintains a bitmap for each file that is being backed up, it requires a list of files to back up. The Hyper-V VSS writer not only participates in the backup operation, but it also enables the VSS requestor to track the list of files that need to be backed up. For instance, if a VM has two VHDs, the file specification with VSS will maintain a list of VHDs and their locations. This not only enables DPM to maintain the right tracking mechanism, but also enables DPM to capture the corresponding volume on the host machine. This enables the backup operation to detect any changes in the file specification. For instance, when a new VHD file is attached to the VM, DPM automatically makes an initial copy of the VHD as part of the next synchronization event.

During the restore process, an administrator can recover the entire VM or restore files and folders that are within the VM. For a VM, the restore can be to the same host, to the same cluster, or to a different cluster.

To ensure consistent backups are captured for Linux VMs, it is critical to install Linux Integration Services. The mechanism to quiesce the workloads within the guest operating system exists only in a Windows operating system. Hence, for Linux VMs, the Hyper-V VSS writer uses a different approach to quiesce the workloads. The Hyper-V writer leverages the file system-level operations, such as freeze and flush, to ensure that the data is file-system consistent. Therefore, from a user's perspective, the file-level consistency is always maintained by the backup process.

## Client protection

Since DPM 2010, DPM has enabled protection of client computers, and with DPM 2012 R2, the latest versions of Windows 8.0 and Windows 8.1 can be protected too. With the DPM 2012 R2 latest update rollup, a backup administrator can configure protection groups to back up client data to Azure Backup. DPM enables on-the-go customers to back up their data using DPM servers when they are connected to the corporate network through wired or wireless LAN.

While the laptop or the desktop computer is disconnected from the network, the data changes to the files and folders are tracked and stored on the local hard drive of the client computer. When the client machine is connected to the corporate network, only delta changes since the most recent backup are synchronized to DPM, thereby ensuring efficient storage of incremental data changes on the DPM replica storage.

If the client computer fails to synchronize to the DPM server within the policy definition, the local DPM client user interface notifies the user to connect to the corporate network. When configuring backup of client data, an administrator can define standard folders to back up, such as My Documents, and also enable end users to add additional folders for protection. The data is automatically backed up to the DPM server at the set backup frequency.

## Client data recovery

As with all other Microsoft workloads, data can be restored using the DPM console. However, the most common mechanism for restoring data is self-service recovery. After the data is backed up, users can browse and restore the recovery points using the Windows client DPM applet on the client computer. Users can search previous versions of files or browse the recovery points on a particular DPM server for all computers that they have access to.

To configure client end-user recovery, the computer on which end user recovery is desired, as well as the DPM server, must be registered to the Active Directory using DPMADSchemaExtension.exe. The following TechNet article describes the specific steps for configuring Active Directory for end-user recovery and for recovering file data: *https://technet.microsoft.com/en-us/library/jj627988.aspx*.

# Exchange Server protection

With Exchange Server, customers enjoy 14 days of backup with database availability groups. However, DPM provides an excellent way to retain data for an even longer period of time. DPM enables protection to tape or to Azure for long-term backup archiving.

DPM 2012 R2 and DPM 2012 SP1 support backup of Exchange Server 2013, Exchange Server 2010, and Exchange Server 2007. With Exchange Server 2013, DPM supports backup of not only servers running Exchange but also databases configured in a database availability group (DAG). Each node of a DAG can be backed up individually with the same DPM server or with a different DPM server.

With Exchange recovery, DPM can recover a single mailbox. The mailbox database is recovered to a recovery database, and then the individual mailbox is recovered. An entire Exchange database or the entire Exchange server can be recovered if it is protected with bare metal backup.

> ***See also*** *For detailed steps to configuring an Exchange database from DPM backup, see the TechNet article at https://technet.microsoft.com/en-us/library/jj628013.aspx.*

# SQL Server protection

DPM enables protection of SQL Server in various configurations. Typical SQL server configurations include standalone SQL Server, SQL Server deployed in a clustered environment with Windows Server Failover Clustering as well as SQL Server deployed as an AlwaysOn availability group. With the latest update rollup, DPM 2012 R2 also supports SQL Server 2014 along with all other major versions of SQL Server, such as SQL Server 2012 and SQL Server 2008, on all major Windows Server versions. In earlier versions of SQL Server, such as SQL Server 2008 where SQL AlwaysOn technology is not available, DPM enables protection of SQL Server in a database mirroring configuration. In a mirroring configuration, protection of the



www.dbooks.org

principal database is available, although not on the mirrored database. DPM doesn't support SQL Server backup when it has database files stored on a remote SMB file share or Windows scale-out file server. DPM also doesn't support databases whose data is stored on Windows Azure blob storage.

SQL Server can be deployed in a physical machine or inside a VM. The DPM agent must be installed on the SQL Server machine. If SQL Server is configured in a clustered mode or as a SQL AlwaysOn cluster, the DPM agent needs to be installed on all the nodes that are failover targets for the SQL Server instance. If cluster members are added, DPM needs to be installed on the newly added cluster nodes as well.

In SQL Server AlwaysOn deployments, DPM honors Preferred Replica, Replica Only, Any Replica, and Primary preferences set by the SQL Server administrator (see Figure 3-3). However, for Preferred Replica, DPM always backs up only from the replica node. When Availability Group is selected for protection, all databases that are added to the availability group are automatically backed up.

> ***See also*** *For a description of how to configure protection of SQL Server AlwaysOn configuration, see https://technet.microsoft.com/en-us/library/hh780998.aspx.*

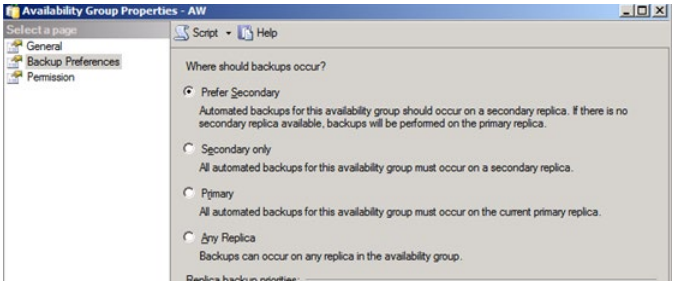



**FIGURE 3-3** SQL Server AlwaysOn configuration showing backup preferences

In a SQL Server deployment, a user can enable auto protection of all databases within the SQL Server instance. This enables managing backup in a dynamic environment where databases are added or deleted from a SQL Server instance without requiring backup administrator intervention. In an auto protection mode, there is no mechanism to turn off backup of subset of databases, for example, a master or model database. Also, with auto protection, it is important to manage the size of the replica storage pool so that it doesn't run out of space.

During the restore process, an administrator can take one of the following actions:

- Recover the entire database to the original SQL Server instance
- Recover the database and rename it
- Recover to an alternate database instance
- Copy the database backup files to a restore folder in a network share

## SQL Server backup process

Similar to Hyper-V VM backup, DPM leverages VSS technology to take application-consistent snapshots of SQL Server. The VSS file specification provides a list of all the .mdf, .ndf, and .ldf files that are associated for a given database that is enabled for backup. DPM maintains a bitmap file filter that tracks the blocks of SQL Server database files that are changed. For each synchronization event, SQL Server is quiesced, a volume snapshot is taken, and a stable point in time of .mdf, .ndf, and .ldf files are copied to the replica server. With DPM file filters and the change bitmap, only delta changes between the current synchronization and previous synchronization are copied to the DPM replica. With express full technology, customers can essentially do a full back up every day efficiently, both in terms of data transfer as well as storage on the replica server.

In addition, to express full backup, DPM ships a transaction log to the DPM replica storage, thereby minimizing data loss up to 15 minutes. While express full backup is efficient in terms of data transfer and storage on the replica, it is expensive on the disk IOPS since storage snapshots are maintained on the production server while the backup data is being copied. Transaction log backup, alternatively, is lightweight and enables up to a 15-minute recovery point objective (RPO).

**NOTE** **When you use DPM, it is critical that no other process backing or truncating transaction logs is enabled because it would interfere with DPM. After the transaction logs are backed up, they are truncated, and this could lead to a break in the transaction log chain.**

## Self-service recovery of SQL Server databases

DPM can be configured to enable self-service recovery for a group of users. The first step is to configure a DPM role using the Configure Self Service Recovery option in the Protection view in the DPM console. The list of users that are allowed to perform self-service recovery are added to the role from Active Directory. The specific set of databases that are allowed for self-service recovery are added to the DPM role. The target recovery SQL Server instances are added to the DPM role.

When this option is configured, the end users can install the Self-Service Recovery Tool (SSRT) from System Center 2012 on their client computers and perform SQL Server database restore without the intervention of the backup administrator.

***See also*** *For more information on the specific steps to configure self-service recovery and how to use the tool, see https://technet.microsoft.com/en-us/library/jj674322.aspx.*



www.dbooks.org

# SharePoint protection

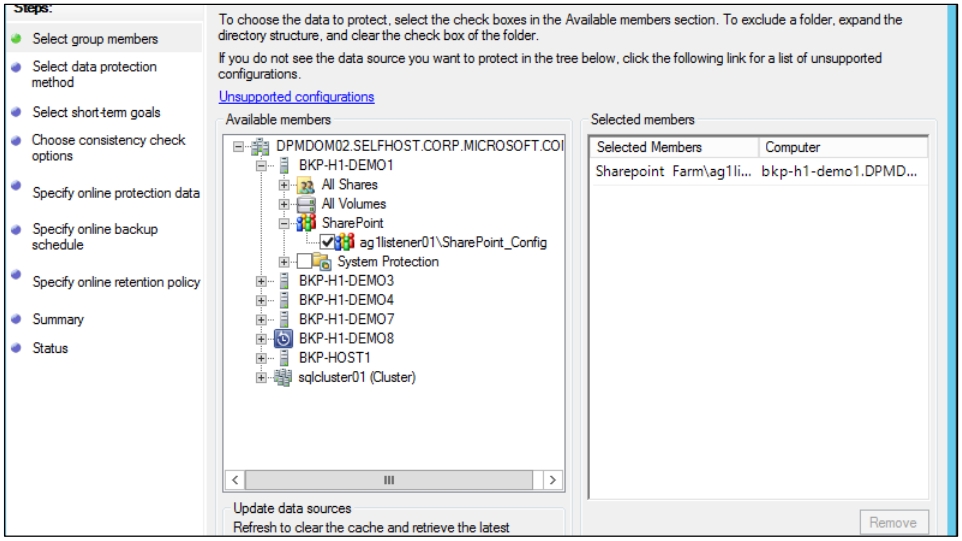
At a high level, SharePoint consists of front-end web servers, a SharePoint configuration database, and SharePoint content databases. The main goal of protecting a SharePoint farm is to protect the content that is stored in the SQL Server content database, as well as the configuration of the SharePoint farm so that the SharePoint farm can be recovered in the event of a disaster, data loss, or corruption.

To protect a SharePoint farm, complete the following steps:

1. Install the DPM agent on the front-end servers and on each of the SQL Server instances that back the SharePoint farm. If the SQL Server instance is configured in AlwaysOn configuration, install the DPM agents on each of the servers that span the availability group.
2. Configure the front-end server for SharePoint protection using the ConfigureSharePoint cmdlet:

   ```
   ConfigureSharePoint [-EnableSharePointProtection] [-EnableSPSearchProtection] [-ResolveAllSQLAliases] [-SetTempPath <path>]
   ```

3. Use the DPM console to create a new protection group and select the member server as the front-end web server to be configured for protection.



## SharePoint backup process

DPM coordinates the backup across multiple servers in the farm to back up data. After backup is done, DPM queries the SharePoint object model and gets all the information about the site and items from the SharePoint server. SharePoint protection uses the SharePoint VSS writer to protect the entire farm. SharePoint VSS writer is a referential writer, and DPM uses the SharePoint VSS writer to obtain the SharePoint topology, such as the SharePoint content databases and the configuration database that are part of the SharePoint farm. With this information, the SQL Server configuration and content databases are backed up. Figure 3-4 illustrates a SharePoint 2013 configuration with backup agents.

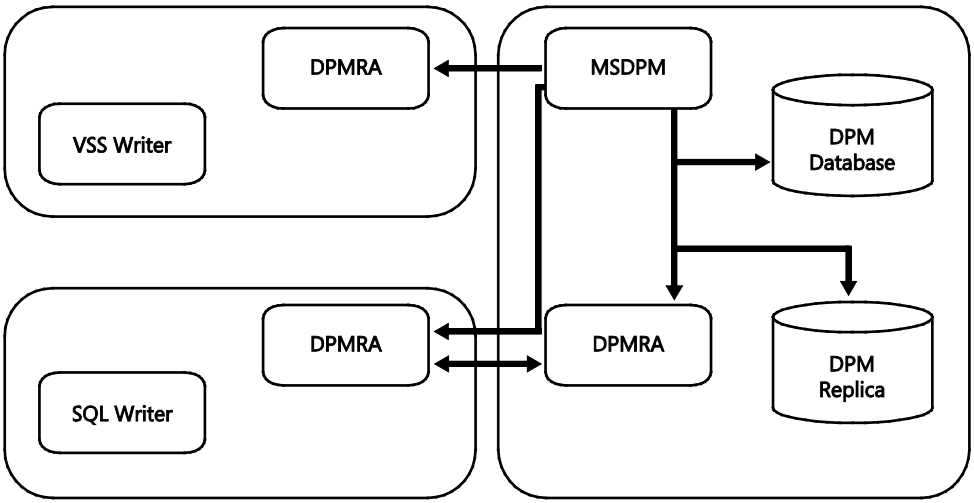



**FIGURE 3-4** SharePoint 2013 configuration with backup agents

As far as SQL Server backup is concerned, it uses exactly the same workflow as backing up SQL Server databases as discussed in the previous section. The data movement for the SQL Server databases happen from the SQL Server machine to the DPM replica server directly, and the SharePoint front-end server and the VSS writer are not involved. Each content database is backed up independently since there is no referential integrity or consistency requirements across the database in a SharePoint farm.

For example, assume the SharePoint farm contains contentDB1, contentDB2, and configDB on SQL1 instance and contentDB3 and contentDB4 on SQL2 instance. SQL1 instance databases contentDB1.mdf, contentDB1.ldf, contentDB2.mdf, contentDB2.ldf, and configDB.mdf are backed up directly from SQL Server SQL1 machine, whereas contentDB3.mdf, contentDB3.ldf, contentDB4.mdf, and contentDB4.ldf are backed up directly from the SQL Server SQL2 machine. It is critical to note that although SQL backup is taken by the SQL VSS writer, transaction log backup is done as part of SharePoint SQL server backup.



www.dbooks.org

SharePoint supports partial backup. Since SharePoint farms can be large, failure to back up a single database in the farm doesn't result in complete farm backup failure. Backups for one or more databases can fail, but a recovery point is created with all other databases. Similarly, when a new content database is added, its initial replica is seeded to the DPM replica server through a nightly job. To immediately add a content database to the protection group, you can add it manually.

**NOTE** **Filestream data that is local to SQL Server are backed up, but filestream data in a remote share are not backed up.**

**TIP** **If SharePoint configuration files need to be backed up, it is advised to back up the SharePoint server with System State backup.**

## SharePoint catalog

In addition to the content database backup, DPM also maintains a catalog of all the items that are backed up as part of a recovery point. This information is used to identify what items are available for item-level recovery from a SharePoint recovery point. When the database backup is complete, the cataloging process starts, although it can be scheduled to run separately. The cataloging process obtains the list of all the SharePoint items that are part of the last recovery point. There is one catalog across all the recovery points in the DPM database. Failure to query the object model for the catalog doesn't result in failure of the backup operation.

## Scoped consistency check

Since SharePoint farms can be large, SharePoint backup supports scoped consistency check. This enables consistency checks only on databases that are inconsistent and other databases go through normal synchronization. The recovery point is created on the DPM replica when all databases are synchronized regardless of whether a consistency check ran on some databases while normal synchronization took place on others.

## SharePoint recovery

DPM enables various levels of SharePoint recovery. Since the entire SharePoint farm, including the configuration and content databases, is backed up, different parts of the farm can be restored as follows (see Figure 3-5):

- Restore entire farm
- Restore specific content databases
- Restore site collection
- Restore site
- Restore individual documents

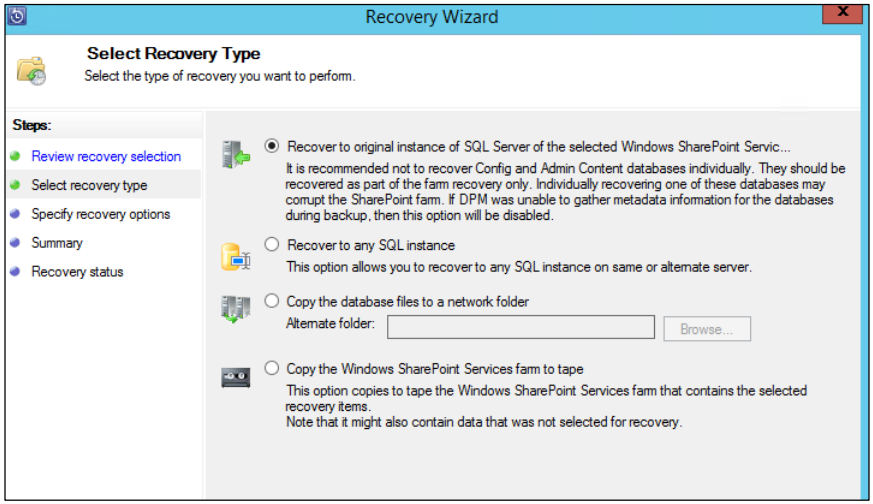



**FIGURE 3-5** SharePoint 2013 Recovery Wizard

The SharePoint data can be recovered to the original site or an alternate site, or it can be copied to a network folder or to a tape. Restoring an entire site to the original site is not recommended without restoring the entire farm since it could lead to inconsistencies. The recommended practice is to recover to an alternate site or an alternate server for configuration or content database recovery. However, to restore a particular item only, it is safe to restore to the original site since the entire content is not impacted by the restore operation. If the entire site needs to be recovered from scratch, it is recommended that you recover the SharePoint server from bare metal and then recover the configuration and content databases using DPM.

To recover the farm, SharePoint recovery provides two methods. One method is to recover using a recovery farm. Use this option when the SharePoint farm has changed since the particular recovery point was created. This involves creation of an alternate farm similar to the production farm called the Recovery Farm Server. To use this option, you must specify the front-end web server name of the recovery farm as well as the SQL Server instance name and a temporary location to copy the content database to (see Figure 3-6).



www.dbooks.org

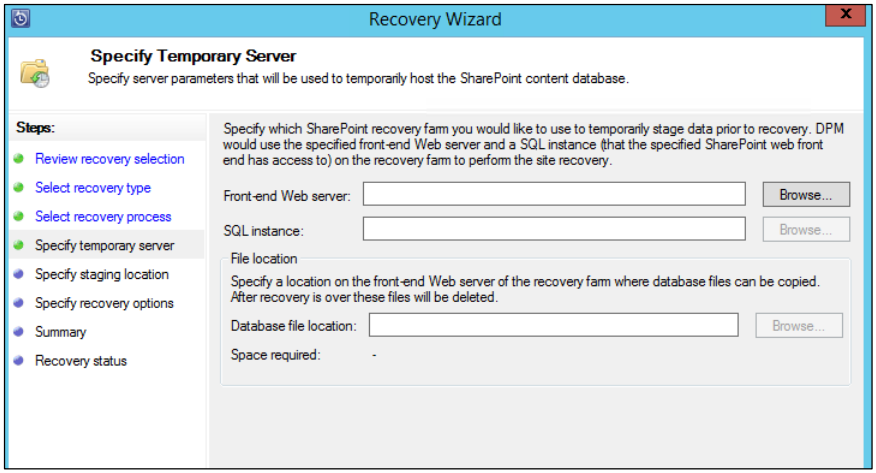



**FIGURE 3-6** SharePoint 2013 recovery using a recovery farm

For item level recovery, you browse each recovery point to find the item. Information about which items are present in each recovery point and the corresponding content database mapping is included in the DPM catalog. Using the DPM catalog, the particular content database is restored first, attached to the recovery farm. Then the particular item from the content database is transferred to the original farm. Along with the contents, the URL and its security attributes are also recovered.

The other recovery method does not involve a recovery farm. This option avoids copying data to the recovery farm and then to the original farm. Instead, the SQL server contents are copied to a share and made visible to a SQL Server instance. The content database is temporarily attached to the SQL instance and then the required item is restored to the target farm (see Figure 3-7).

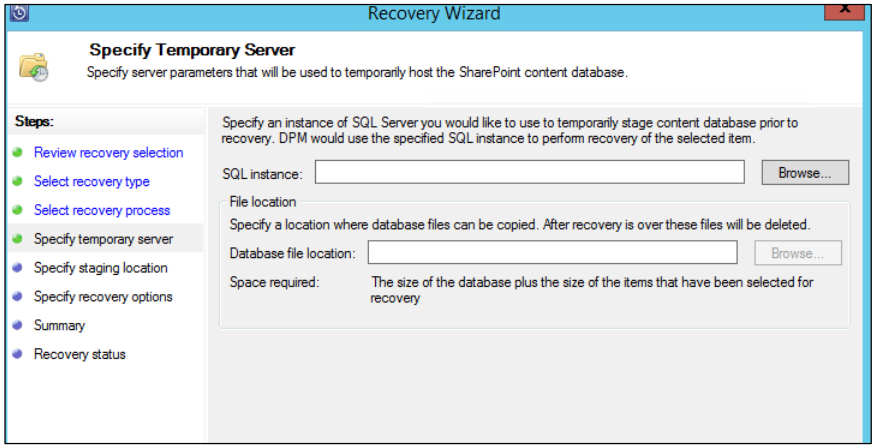



**FIGURE 3-7** SharePoint 2013 recovery without using a recovery farm



www.dbooks.org

CHAPTER 4

# Protecting Azure IaaS workloads

Organizations build business continuity procedures to encompass all infrastructure that is critical to their business. Backup is one such business continuity procedure, and organizations can choose from an extensive set of vendors who can cater to their on-premises backup needs.

However, as organizations move more of their infrastructure pieces into the public cloud and treat the public cloud as an extension of their on-premises infrastructure, backup procedures are extended to include public cloud elements as well. Microsoft Azure is no different. Deployment of Tier I workloads like Microsoft SQL Server and Microsoft SharePoint on the public cloud depends on having a good backup and recovery story. Microsoft System Center Data Protection Manager (DPM) is Microsoft's answer to this problem. By running DPM in an Azure infrastructure-as-a-service (IaaS) virtual machine (VM), you get the ability to back up workloads running in other Azure IaaS VMs.

This chapter starts by covering the nuances of setting up DPM in Azure. It then describes the protection of workloads using DPM and Azure Backup. Lastly, the chapter delves into the best practices for deploying DPM at scale in Azure and the constraints that need to be kept in mind.

## Setting up DPM in Azure

There are three ways to set up and use DPM in Azure:

- Download the System Center binaries to a new Azure IaaS VM and install DPM.
- Use the DPM evaluation virtual hard disk (VHD) as the image for a new Azure IaaS VM.
- Upload the VHD to Azure from an existing DPM VM on-premises, and use the uploaded VHD as the image for a new Azure IaaS VM.

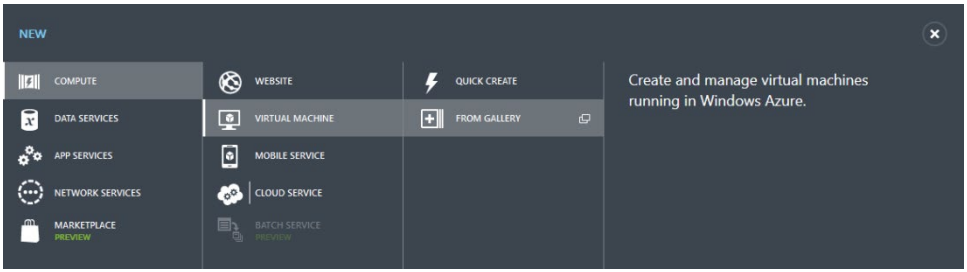
These three options have varying degrees of simplicity and utility, depending on how comfortable you are with using DPM. For example, existing users of DPM might find it easier to configure everything on-premises and just upload the preconfigured VHD to Azure for use. The evaluation VHD is ideal for users trying out DPM who want to get started quickly. This chapter covers the first option in greater detail, that is, a full installation of DPM after downloading the System Center binaries.

For any of the three methods listed, you first need to create a VM and configure it correctly. The following section covers the creation and configuration of the VM.

## Creating a new VM for DPM

To begin the process of setting up DPM in Azure, you need to first create an Azure IaaS VM. The following steps walk you through the process of creating a VM using the Azure management portal.

1. Go to *http://manage.windowsazure.com* and log in to the Azure management portal.
2. Click New at the bottom-right corner of the management portal, click Compute, and then click Virtual Machine. You will have two options for creating a VM: Quick Create and From Gallery. Click the From Gallery option to open the Virtual Machine Creation Wizard.




3. Search for the image named Windows Server 2012 R2 Datacenter, and click the arrow to go to the next screen.
4. The second page of the Virtual Machine Creation Wizard collects the VM name, size, and login information. Ensure that you pick the Standard tier. The size of the VM should be at least A2 (2 cores, 3.5 GB of memory). The size of the VM can be changed later; choosing the right size is explained later in this chapter.
5. The third page of the Virtual Machine Creation Wizard has the most important configuration settings. You can retain the default values for most of the inputs. However, pay close attention to the Subscription, the Virtual Network, and the Storage Account options.
6. If you have multiple subscriptions, pick the subscription that also contains the VMs and workloads that need to be protected. The DPM VM should be placed in this subscription.
7. The DPM VM should have an independent storage account. Select the Use An Automatically Generated Storage Account option. To find the generated storage account name later, you click the VM to show the details and then navigate to the Disks section on the Dashboard tab.

   The default replication setting for the automatically-created storage account is Geo-redundant. This setting is also recommended for your backup data because it acts as insurance against Azure disasters. Even if your production data is stored on locally



www.dbooks.org

redundant storage and is lost when a disaster strikes the Azure datacenter, the backup data is replicated to another region and is available to you.

8. The DPM VM should have proximity to the workloads that need to be protected. At the very least, the DPM VM should be placed in the same region as the VMs whose workloads need to be protected.

   If you have created a virtual network for your VMs and workloads, the DPM VM must be placed in the same virtual network. Between selecting a region and a virtual network, a virtual network gets preference.

9. In the Region/Affinity Group/Virtual Network combo box, enter the option that best fits your deployment.
10. Complete the remaining fields and finish the Virtual Machine Creation Wizard.

**NOTE** **DPM uses the Active Directory domain controller to identify servers that are joined to the same domain. You can then select a subset of these servers to protect by installing the backup agent. When adding the DPM VM to a virtual network, you must ensure that the domain controller is accessible.**

## Joining the VM to a domain

After the VM is created, wait for it to boot up and reach the Running state, and then connect to the VM over RDP with the username and password that you provided at the time of VM creation.

When you are logged in, join the machine to the same domain that the yet-to-be-protected workloads are joined to. This trust relationship allows DPM to discover the servers easily and makes the backup agent installation process simple. Alternatively, you can use certificate-based authentication and manual installation of the agent.

## Adding backup storage

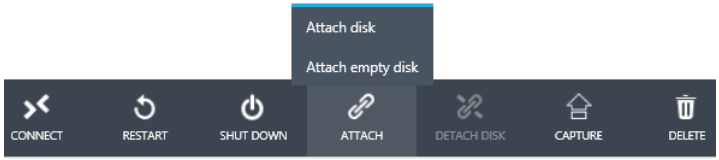
All backup products need storage media to keep the backup data and the recovery points, and DPM is no exception. There are two broad types of storage media supported by DPM: disk and tape. When it is deployed on-premises or in your datacenter, there are multiple ways to make disk storage available to DPM:

- Direct attached storage (DAS)
- iSCSI disks
- VHDs on an SMB Share
- Fiber channel attached SAN storage

However, when DPM is running as a VM in Azure, the only type of storage that can be used is a VHD. The VHDs that are created and attached to the VM together form the backup storage pool for DPM consumption.

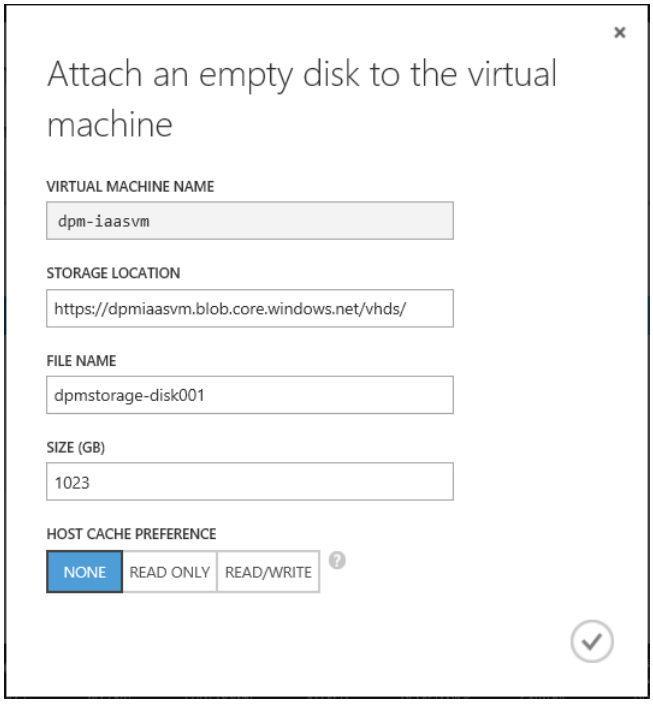
The following steps walk you through the process of creating a VHD using the Azure management portal and attaching it to the DPM VM.

1. Go to *http://manage.windowsazure.com* and log in to the Azure management portal.
2. In the left pane, click the Virtual Machines tab.
3. From the list of VMs displayed, select the VM that will host DPM.
4. From the menu bar at the bottom, click Attach and select Attach Empty Disk.




5. A dialog box with details about the disk that needs to be attached appears. Enter a file name and the size of the disk. (How to choose the correct amount of storage for the backup storage is covered later in this chapter.)

   For the remaining options, use the default values. Click the check mark on the bottom-right corner of the dialog box to complete the action.

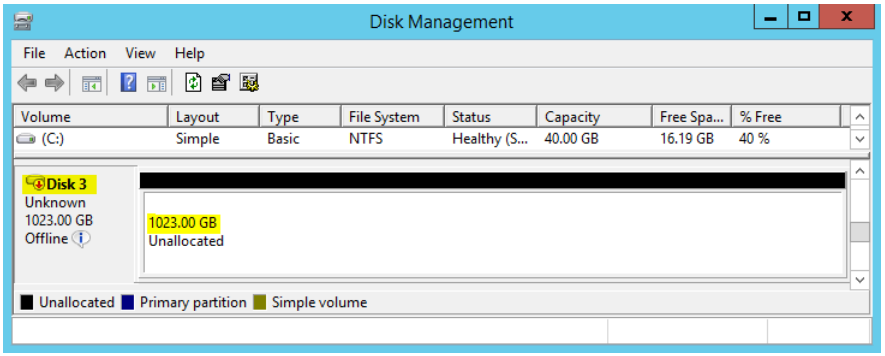





www.dbooks.org

**NOTE** The storage location that is pre-populated is the automatically generated storage account from the VM creation procedure.

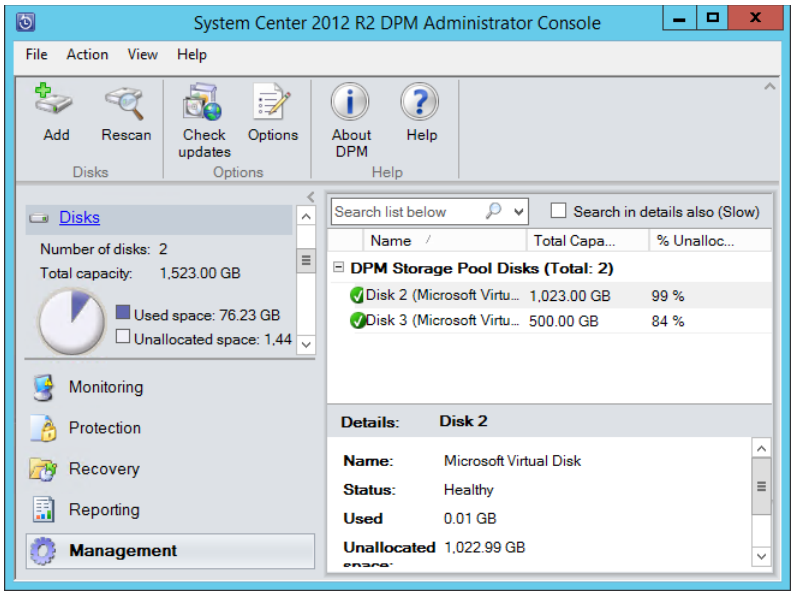
6. Azure creates the empty disk and attaches it to the VM. When the operation is complete, use the Disk Management tool in the guest to see the disk. Right-click the disk and select Online. Later, you will configure this disk as backup storage for DPM.




## Installing and configuring DPM and Azure Backup

When the VM is ready, download and install DPM. Complete the following steps to get a fully equipped version of the software up and running.

1. An evaluation version of System Center can be downloaded from the TechNet Evaluation Center at *http://www.microsoft.com/en-us/evalcenter/evaluate-system-center-2012-r2*. Run the Upgrade Wizard along with the product key to move to a full version.
2. Follow the steps outlined in the TechNet documentation for installing DPM, found at *http://technet.microsoft.com/en-us/library/hh758153.aspx*.
3. Ensure that the backup storage is correctly recognized and used by DPM. Click the Management tab at the bottom-left corner. Click the Disks link in the left pane. This opens the list of disks that form the backup storage pool. On the menu bar at the top, click Add to start the workflow necessary to use the attached Azure disks as the backup storage.




4. Ensure that the Azure Backup Agent is downloaded and installed. Ensure that the agent is configured to send data to the Azure Backup vault and that the appropriate scratch space is also provisioned in DPM. Chapter 10, "Integration with Azure Backup," covers the benefits of using Azure Backup with DPM in great detail and offers guidance on how to set it up correctly.

## Reviewing the post-deployment architecture

After the VM has been deployed and configured and the backup storage for the VM has been attached, the overall architecture should be similar to the diagram shown in Figure 4-1.

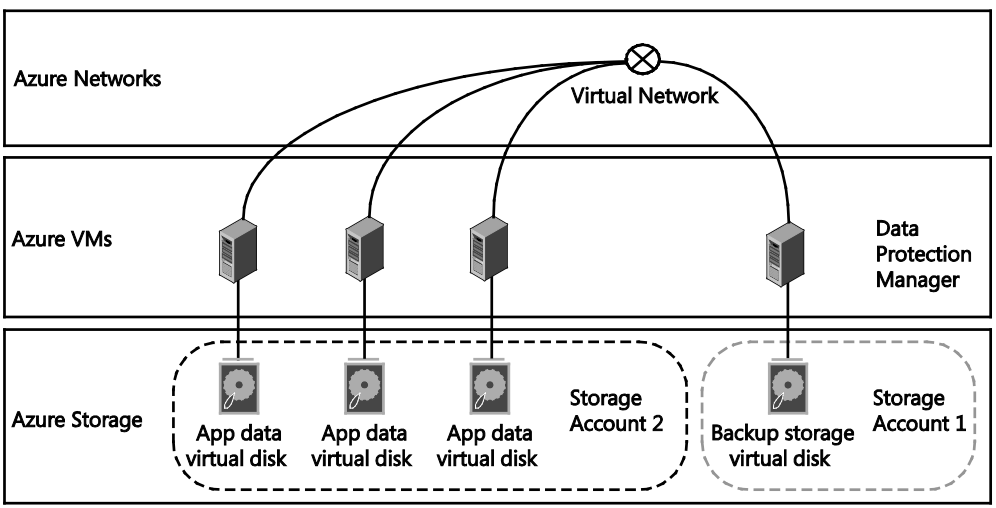



**FIGURE 4-1** Relationship between DPM and the protected workloads in the Azure infrastructure



www.dbooks.org

# Protecting workloads

Protecting workloads in Azure using DPM is very similar to protecting the same workloads on-premises. This section covers the key aspects of protecting workloads in Azure and highlights differences from your on-premises experience that you might encounter.

Whether in Azure or on-premises, protecting workloads requires two key steps:

1. Discovering servers and installing the agent
2. Discovering workloads and creating a protection group

## Discovering servers and installing the agent

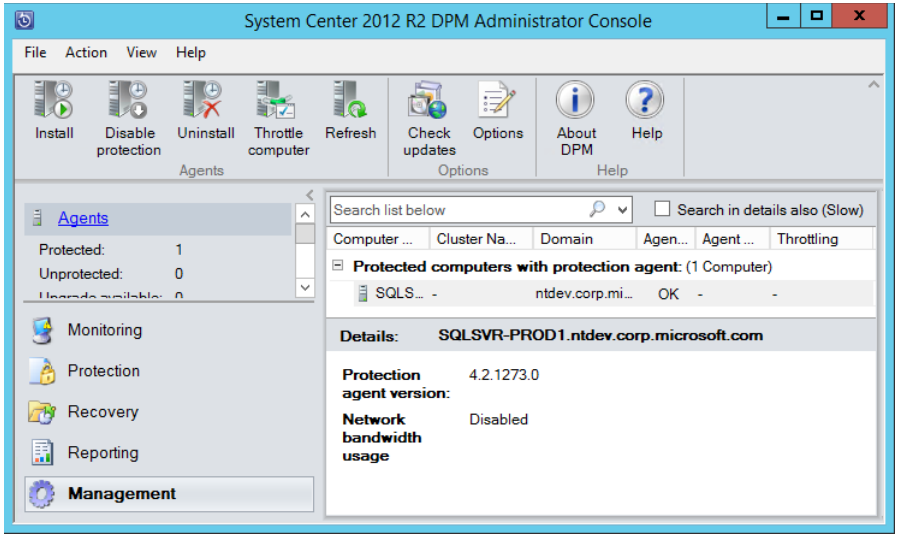
DPM depends on the Active Directory domain controller for server discovery. It queries the domain controller and gets a list of servers that are joined to the same domain as the DPM server. From this list, you can select one or more servers that have workloads needing protection and install the backup agent on those servers. This process is simple and is preferred by the majority of customers using DPM. This process is applicable to the workloads in Azure too; the workloads in Azure and the DPM instance in Azure need to share the same domain.

Typical customer deployments also employ Azure Virtual Network to partition and classify the VMs into access groups and control the external access. With Azure Virtual Network, you can provision private networks in Azure and optionally connect these to your on-premises datacenters to form a hybrid deployment. With an Azure virtual network in place, DPM should be a part of the virtual network to get access to the workloads it needs to protect.

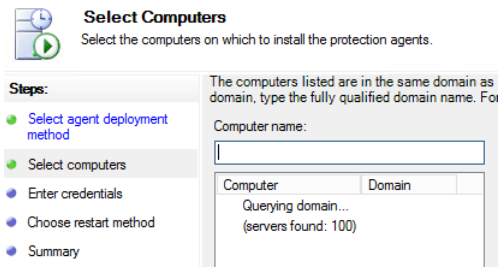
**NOTE** **The virtual network is set when the VM is created. Ensure that you place the DPM VM in the correct network.**

DPM should be able to discover servers without problems if the VM is connected to the correct Azure virtual network and is joined to the proper domain. Start the agent installation workflow in the DPM console by completing the following steps:

1. Click the Management tab at the bottom-left corner.
2. Click Agents in the left pane. This opens the list of servers with the agent installed and the menu bar options to modify the list of installed agents.
3. Click Install in the top-left corner of the DPM console.




4. As a part of the agent installation workflow, DPM discovers the other servers in the same domain. Select the server you need to install the agent on from the list of servers discovered.




## Discovering workloads and creating a Protection Group

When the agent is installed on the selected servers, you can protect the workloads running on those servers by creating a Protection Group. The Protection Group is a set of data sources that are configured to have the same backup configuration parameters, such as time of backup, synchronization frequency, retention period, and whether the backup data should be written to other media, like cloud and tape. The data sources can be simple files and folders or workload-specific entities, such as Microsoft SQL Server databases and Microsoft Exchange mailboxes. Any given data source can be a part of only one Protection Group, but a Protection Group can manage multiple data sources.

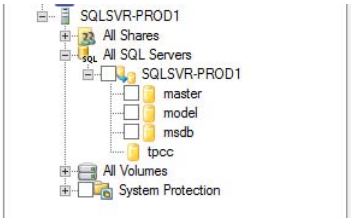
Complete the following steps to select and protect data sources:

1. Click the Protection tab at the bottom-left corner.
2. On the menu bar, click New to trigger the creation of a new Protection Group. A wizard appears to guide you through the process of protecting data sources.



www.dbooks.org

3. On the Select Group Members page, select the data that you want to protect with this Protection Group. Browse through the servers on which you have installed the backup agent and select the objects that you want to include for protection.




4. Ensure that you also select the I Want Online Protection option on the Select Data Protection Method page. This ensures that you can continue tiering data to Azure Backup and keep the storage footprint of the DPM VM to a minimum.
5. Continue to the end of the wizard to finish creating the Protection Group.

With this procedure complete, you have successfully selected the Azure workloads to be protected and have created a Protection Group to encapsulate them.

## Workloads and configurations supported for backup

While the experience of using DPM in Azure is very similar to the on-premises experience, the workloads that are supported for protection by DPM in Azure are a little different. This is fundamentally because the workloads and configurations that Azure supports are different from the workloads and configurations that are supported on-premises on Windows Server or Hyper-V. For example, at the time of writing this chapter, Microsoft Exchange is not a supported workload in Azure. By extension, backup of Microsoft Exchange mailboxes is not a supported scenario for DPM running in Azure. While there is no hard block in the software, planning the backup strategy for your Azure infrastructure needs a clear understanding of what support you will get from Microsoft.

The list of workloads supported by Azure is a continuously evolving one. You can find the latest and up-to-date information at *http://support.microsoft.com/kb/2721672*. Additionally, the detailed list of workloads, versions, and capabilities supported by DPM can be found in the TechNet documentation at *http://technet.microsoft.com/en-us/library/jj860400.aspx*.

# Considerations for performance and scale

Using DPM in Azure to protect your Azure workloads also means that you need to pay for the additional resources consumed by DPM. This includes the VMs, the Azure disks used as backup storage, and Azure Backup for longer retention. IT administrators have the dual responsibility of providing a performant backup environment and of controlling the costs. With the pay-as-you-use model in Azure, it is possible to "right size" the DPM setup and

grow your resources as you need them, without having the reserve capacity up front. This section delves deeper into the parameters that tune the performance and scale DPM in Azure.

## Recommendations for better performance

DPM is an I/O-intensive application, and though it is not sensitive to latency, it consumes significant network bandwidth. It is therefore important that the right resources are made available to meet your backup needs. Here is a list of recommendations based on performance tests run with DPM in various configurations:

1. Use the Standard tier when creating a VM in which DPM will be installed. The IOPS per attached disk is higher for the Standard tier (500 IOPS) than for the Basic tier (300 IOPS).
2. Do not use the same storage account for the DPM disks and the disks attached to the production VMs. Azure places a limit of 20,000 IOPS per storage account. If the data to be backed up is read and written to the same storage account, you have effectively halved the IOPS available from the storage account to the workloads or DPM.
3. The minimum VM size should be A2 with 3.5 GB of RAM. DPM needs at least 2 GB of RAM to work correctly, and A2 is the smallest VM size where the RAM is greater than 2 GB.
4. Since DPM and the workloads are in the same region, there is no network egress cost incurred during backup or restore.

## Scaling up vs. scaling out

DPM in Azure supports both scale out and scale up. This section provides guidance on when to scale up and when to scale out.

Scaling up your DPM setup deals primarily with the need for more backup storage. Different VM sizes in Azure support a different number of attached virtual disks. Given that the maximum possible size of an Azure disk is just shy of 1 terabyte (TB), each VM size has a limited amount of storage that can be used as directly attached virtual disks.

The other factor that influences the scaling decision is the number of servers that are being protected. A single VM cannot scale up infinitely to provide resources for backing up workloads, and when you reach the limit for a specified size, you need to choose whether to scale up or scale out.

Table 4-1 lists the operating limits that are suggested for DPM VMs of different sizes. The maximum number of protected servers that a given size can support is derived from scale tests that have been run, while the maximum raw backup storage is a limit imposed by Azure and can be explored in greater detail at *http://msdn.microsoft.com/en-us/library/azure/dn197896.aspx*.



www.dbooks.org

**TABLE 4-1** Limits on raw backup storage and number of protected servers for each VM size

| DPM VM SIZE | MAXIMUM RAW BACKUP STORAGE | MAXIMUM NUMBER OF PROTECTED SERVERS |
|---|---|---|
| A2 | 4,092 GB (4 disks of 1,023 GB each) | 20 |
| A3 | 8,184 GB (8 disks of 1,023 GB each) | 40 |
| A4 | 16,368 GB (16 disks of 1,023 GB each) | 60 |

For example, assume you start with an A2 VM having a single 500-GB virtual disk attached, protecting 10 servers. If you need more storage, the first step is to exhaust the storage capability of the A2 VM. The easiest way to do this is to create a new virtual disk and attach it to the running VM. If you need to protect more servers, then you can install up to 20 agents on various Azure VMs to protect the workloads on those servers. When you reach the maximum storage capacity of the VM or exhaust the number of servers that the DPM instance should be handling, you can either scale up or scale out:

- Change the VM size. Click the VM to open the detailed view. Click the Configure tab, and scroll down to the Virtual Machine Size drop-down list. Change the VM size to the next higher size, and click Save.
- Create another VM, and configure DPM with the same steps that have been highlighted in this chapter.

**NOTE** **There is no easy way to increase the size of an existing virtual disk in Azure. In order to utilize the full storage capacity of a VM, ensure that each disk is created with size of 1,023 GB—the maximum possible. Azure charges you for the data consumed on the virtual disk rather than on the created size of the virtual disk.**

The advantage of changing the VM size is that the software is already set up and you need to manage one less backup server. However, when you reach an A4 size and exhaust either the maximum raw backup storage or the number of protected servers, then you have to scale out and create a new VM to accommodate your backup growth.

## Tiering data to Azure Backup

Protecting large amounts of data can quickly degenerate into a situation where a large number of DPM VMs are needed for longer retention. For example, protecting 100 GB of data for 30 days and having 5-GB daily incremental backup data means that DPM ends up storing 250 GB of data. (This is based on an initial size of 100 GB plus 30 days of incremental backups at 5 GB per backup.) Working backwards with the same assumptions, it means that you cannot protect more than 1.6 TB of data using an A2 VM. This data growth can be severely limiting as you incur additional costs of spinning up and running new VMs.

The recommendation is to use Azure Backup as the bottomless storage pit for anything more than a few days. Thus, the data retained with DPM is always the freshest and can be used quickly for any immediate recovery scenarios. This also postpones the decision to scale up or scale out with respect to hitting the maximum raw backup storage limit.



www.dbooks.org

CHAPTER 5

# Protecting Hyper-V virtual machines

Virtualized servers have already surpassed physical servers in terms of the number of deployed instances, and this is primarily due to the cost savings achieved with server virtualization. With virtualized servers becoming first-class citizens of the datacenter, building a virtualized private cloud is the next step in the cloud computing journey to get all the benefits of the cloud while keeping control within an IT department. Whether the private cloud is built on Hyper-V or VMware, Microsoft System Center 2012 R2 Data Protection Manager (DPM) protects both deployments. This chapter covers the protection and recovery of Hyper-V VMs at private cloud scale using DPM 2012 R2. Different protection configurations that apply to Hyper-V VMs and the different restore options that you can exercise are covered in detail.

## Customer scenarios and challenges

Organizations build the virtualized private cloud primarily to offer anything-as-a-service to their end customers (also called *tenants*) in a managed or self-serve way. This includes renting the infrastructure and providing value-added services (for example, backup and disaster recovery as a service). When customers offer backup-as-a-service (BaaS), they promise certain service level agreements (SLAs), which have financial implications if not met. The ability to back up thousands of VMs daily at scale and recover a particular tenant VM or a file is the primary requirement. Backup is all about data insurance, and the biggest challenge is to get reliable backups at scale while keeping the costs in control.

With the release of DPM 2012 R2, Hyper-V VMs could be protected, but not efficiently at scale (for example, 1,000 VMs in a 24-node cluster). Customers—especially hosting service providers—reported issues with Hyper-V protection at scale. The biggest issue was the inordinate amount of time taken to finish all the backup jobs, and overshooting the backup window resulted in SLAs being missed. An investigation of the issue revealed that the culprit was the host-level volume snapshot mechanism.

With Update Rollup 3 (UR3), significant enhancements were made for the backup of Hyper-V VMs to ensure that the backup was guaranteed within SLAs. The enhancements also made backups much more resource efficient and allowed a group of DPM servers to back up thousands of VMs within a virtualized deployment. The core change that enabled this performance and scale boost was the elimination of host-level volume snapshots during backups. Going forward in this chapter, you will find details and best practices for setup, protection, and recovery of Hyper-V data.

# Planning for VM backup

Before installing DPM 2012 R2 and performing a series of required configuration tasks, you should devise a strategy for what to back up, when to back up, and where to back up. This section helps answer some of the questions you might have before starting the data protection.

## What to back up (host level vs. guest level)

It is recommended that you protect your Hyper-V environment by combining host-level backup of Hyper-V VMs with the existing backup strategy for your in-guest applications, like Microsoft SQL Server, Microsoft Exchange, and Microsoft SharePoint.

Host-level backup of VMs is equivalent to protecting a physical server using bare-metal recovery. It is recommended that you protect your application data more frequently than your VMs. For example, VMs can have a schedule that backs up data once per day or once per week, while Microsoft SQL Server databases could be backed up as frequently as every 15 minutes.

## When to back up

Typically, production workloads experience peak load during specific time windows during the work day and are less loaded during off-peak hours (maintenance window). Customers (usually the hosting service providers) need the flexibility to run backups during off-peak hours, especially for backup at scale. They need the ability to contain overhead of backup as these affect the performance characteristics of the production workloads. IOPS consumption, network bandwidth, and CPU utilization during backup are some of the key performance parameters that affect the production workload performance. So the concept of a backup window for VM data sources was introduced in DPM 2012 R2 UR3.



www.dbooks.org

With UR3, you can now create specific time windows within which to run the scheduled backup jobs and consistency check (CC) jobs to achieve the SLAs. These time windows are configured through Windows PowerShell at the level of a protection group to strictly ensure that all backup and CC jobs run only during the set window. Jobs that are actively transferring backup data after the window has ended are allowed to continue, but all other jobs are automatically cancelled. The backup/CC windows do not affect ad hoc jobs triggered by the user.

The following Windows PowerShell commands demonstrate how you can configure the backup window. Set the values for $pgName, $startTime, and $duration. The backup schedule should align with the StartTime parameter used in the Set-DPMBackupWindow command.

```
$pg = Get-ProtectionGroup -DPMServerName $env:computername | ?{$_.FriendlyName -like
"*$pgName*"}
$mpg = Get-ModifiableProtectionGroup $pg
$sched = Get-DPMPolicySchedule -ProtectionGroup $mpg -ShortTerm | ? { $_.JobType -eq
"FullReplicationForApplication" }
Set-DPMBackupWindow -ProtectionGroup $mpg -StartTime $startTime -DurationInHours
$duration
Set-DPMPolicySchedule -ProtectionGroup $mpg -DaysOfWeek $sched.WeekDays -TimesOfDay
$sched.TimesOfDay -Schedule $sched
set-dpmprotectiongroup $mpg
```

Similarly, to set the consistency check window, the following Windows PowerShell commands can be used.

```
Set-DPMConsistencyCheckWindow -ProtectionGroup $mpg -StartTime $startTime -
DurationInHours $duration
Set-DPMProtectionJobStartTime -ProtectionGroup $mpg -JobType ConsistencyCheck -StartTime
20:00
-MaximumDurationInHours 3
set-dpmprotectiongroup $mpg
```

## Where to back up

DPM supports longer retention (multiple years) since the release of DPM 2012 R2 UR5. It is recommended that you use Azure Backup as the bottomless storage for host-level VM backups for longer retention. You can still have a few days' worth of backup on-premises to ascertain faster operational recovery jobs. Offloading backup data to Azure gives you the illusion of infinite storage capacity; you no longer need to worry about managing the tape infrastructure.

## How to back up

With the enhancements delivered in UR3, a single DPM server can protect up to 800 VMs. With customers running thousands of Hyper-V VMs in a cluster or rack, a single DPM server does not suffice. Multiple DPM servers are needed to protect these VMs, and DPM scale-out architecture allows this without any issues. When DPM servers are deployed, these can be managed and monitored through the System Center Operations Manager console.

A DPM server can protect up to 800 VMs with co-location turned on and up to 300 VMs with co-location turned off. In general, you have one replica volume and one recovery point volume per protected data source. Co-location enables you to have multiple data sources mapping on a single replica and recovery point volume. It allows you to locate data from different protection groups on the same disk or tape storage.

Here are some high-level issues with the co-location feature and the workarounds:

- Keeping the replica volume too small defeats the purpose of co-location. Default replica volume size is 250 GB (which is highly debatable since one size doesn't fit all customers/data sources). This number is configurable through a registry setting, HKEY_LOCAL_MACHINE\SOFTWARE\Microsoft\Microsoft Data Protection Manager\Collocation\HyperV\CollocatedReplicaSize. The only condition is that it must be a multiple of 10 MB. It is very difficult for a customer to know the optimal number.
- Keeping the replica volume too big means that a lot of data sources will be co-located, and this will mean that the customer will have less flexibility due to the nature of disk colocation, which includes:
  - If one data source is not able to create a recovery point for a number of days equal to the retention range, then it loses all recovery points due to garbage collection of shadow copies. In non-colocation cases, this does not happen since at least one recovery point is preserved.
  - A user can't stop protection with retain data and re-protect in a different protection group more than twice for a co-located data source. The third time produces an error and the user must wait for days to perform the operation a third time.
- When protecting very few data sources, large replica volumes lead to wasted storage space. You can tweak the number of data sources on a volume by changing the registry value HKEY_LOCAL_MACHINE\SOFTWARE\Microsoft\Microsoft Data Protection Manager\Collocation\HyperV\DSCollocationFactor to fit more VMs on the replica. It is safe to increase it to a large number (for example, 50).

If you understand the implications and find the right settings, the recommendation is to use the colocation feature. Why not leverage it and protect more data sources per DPM server?



www.dbooks.org

***See also*** *For an explanation of how to leverage the DPM scale-out feature to protect VMs deployed on a big cluster, see the TechNet blog entry at http://blogs.technet.com/b/dpm/archive/2013/05/01/sc-2012-sp1-dpm-leveraging-dpm-scaleout-feature-to-protect-vms-deployed-on-a-big-cluster.aspx.*

**TIP** **Use the default provider shipped with DPM 2012 R2 UR3 or later.**

## How to control costs

Since backup is all about data, storage needs to be optimized for backup purposes. Double-parity (and not three-way mirrored) is recommended for backup storage, leveraging Windows Server de-duplication to reduce the amount of backup storage consumed. De-duplication provides a great opportunity to realize storage savings. These savings will vary depending on the workloads running within the VM and the amount of churn created. The high-level steps to realize these savings are as follows:

1. Run DPM in a virtualized deployment.
2. Provision backup storage through VHDs residing on scale-out file server (SOFS) shares.
3. Enable the De-duplication role on the SOFS volumes hosting the backup storage VHDs.
4. Use DPM VHD/VHDX files of 1 TB.

De-duplication for Data Protection Manager is described in greater detail in Chapter 8, "Optimizing backup storage."

***See also*** *A white paper describing all the pre-requisites as well as guidance on how to ensure you get the most de-duplication savings from your setup is available at http://technet.microsoft.com/en-us/library/dn891438.aspx.*

# Protecting Hyper-V VMs

This section covers what can be protected, different deployment topologies, and performance and scale numbers. Hot backups of Microsoft workloads (for example, SQL Server, Exchange Server, SharePoint Server, and file servers) and Linux workloads is supported. Note that you get app-consistent backups for the VMs running Microsoft workloads while you get file-consistent snapshots for the VMs running Linux workloads due to no Volume Shadow Copy Service (VSS) support in the guest.

DPM can be used to protect VMs hosted on the following:

- Standalone Hyper-V hosts that use local or direct-attached storage. Note that this option does not provide continuous availability and is not recommended for production deployments.
- Hyper-V cluster with storage on Server Message Block (SMB) shares backed by an SOFS cluster. This deployment type is referred to as "Hyper-V over SOFS."
- Hyper-V cluster with the virtual hard disks stored on Clustered Shared Volumes (CSV). This deployment type is referred to as "Hyper-V over CSV."

## Protecting Hyper-V over SOFS

Hyper-V over SOFS configuration (with Storage Spaces) enables cost-effective, highly available, scalable, and flexible storage solutions for business-critical virtual deployments. It leverages industry-standard storage and allows customers to use Windows Server for highly available storage that can cost-effectively grow with demand. In this configuration, compute and storage are decoupled, and you can independently scale one without the need to scale the other. This configuration provides the lowest acquisition and operations cost. It allows highly available VMs, continuously available file servers, and fault-tolerant storage. The only challenge with this configuration is hardware setup and software installation unless you decide to go with the Cloud Platform System option.

Here are some considerations when using Hyper-V over SOFS and what they mean from a backup perspective:

- The file server must have Windows Server 2012 R2 with the new SMB 3.0 protocol. In order to get scalable backups with DPM, you cannot use non-Microsoft file servers that implement the SMB 3.0 protocol (because the DPM agent doesn't work with these file servers).
- You must have separate failover clusters for Hyper-V and for the file server. DPM agents should be installed on the Hyper-V cluster nodes and on all the storage nodes (since the storage server is also clustered). You'll need full-share and folder-level permissions for the local $ account of the file server on the SMB share.

***See also*** *For a description of a few different Hyper-V over SMB configurations with increasing levels of availability, see the blog post at http://blogs.technet.com/b/josebda/archive/2013/01/26/hyper-v-over-smb-sample-configurations.aspx.*

Now consider the failure scenarios. What if the compute node goes down? There is no impact to the protected VMs on that node since the tracking is on the storage node. If there is a DPM VM on that node, all running jobs fail and are re-tried automatically.

What if the storage node goes down? All VMs that are touched by this storage node (since the last backup) go into CC mode since the tracking is on the storage node. In Cloud Platform System, where one rack can host up to 2,000 VMs on a four-node SOFS cluster, roughly 500 VMs go into CC mode when one storage node goes down.



www.dbooks.org

How do the scale numbers look? In a scale test conducted by the DPM Product Group, continuous daily backups for three weeks (using virtualized DPM servers) were taken where the workload running inside each of the VMs was spread across multiple I/O profiles (SQL OLTP, Exchange, File Server, Video Streaming, and SQL Decision Support System). The guest operating system used for the protected VMs was Windows Server 2012 R2. Figure 5-1 shows the typical deployment used for protecting Hyper-V over SOFS, and Table 5-1 shows the details of the scale test conducted in-house for the Hyper-V over SOFS deployment.

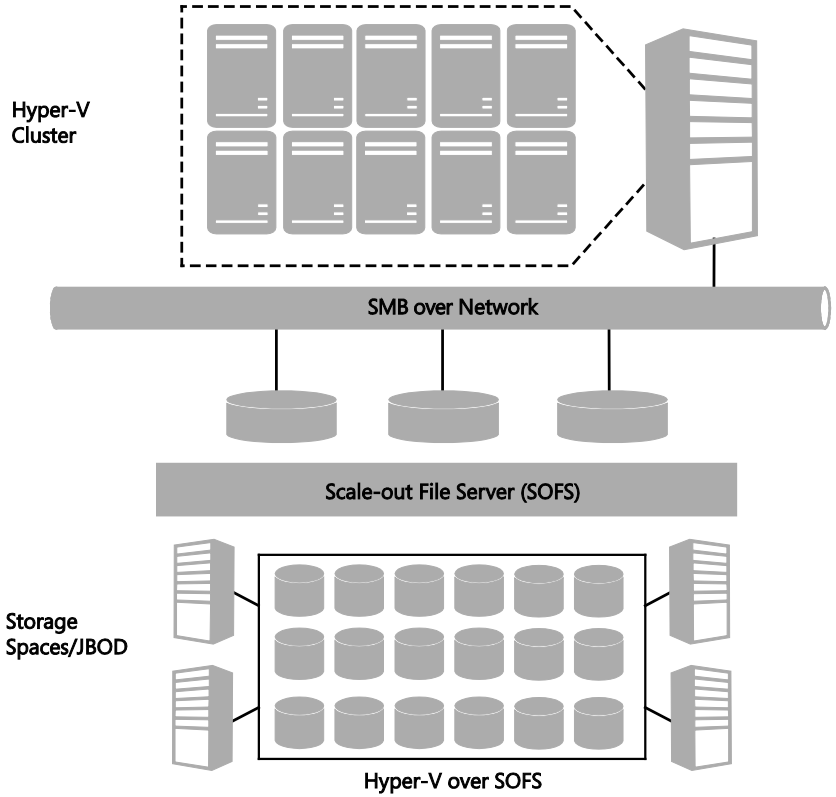



**FIGURE 5-1** Hyper-V over SOFS deployment

**TABLE 5-1** Internal scale test results for the Hyper-V over SOFS configuration

| CONFIGURATION | HYPER-V OVER SOFS |
|---|---|
| Number of Hyper-V hosts | 24 |
| VM config (RAM) | 2 to 8 GB |
| VM disk size | 120 GB (20 GB for operating system + 100 GB for data) |
| Total number of VMs | 1,000 |
| VM churn per day | 5% |
| SOFS cluster nodes | 4 |
| Number of virtual DPM servers | 8 |

Scale testing with each DPM server protecting between 50 to 250 VMs was performed. DPM VMs were deployed in scale-out configuration to protect VMs from the same Hyper-V cluster nodes. Results were pivoted around the following criteria:

- **Backup success rate per day** This signifies the percentage of VMs having successful backups in a single day.
- **Overall backup success rate** This signifies overall percentage of successful backups across all VMs for a three-week duration.

More than 98 percent success was achieved for both the metrics. It also implies that there were more than 20,000 jobs that ran successfully during this three-week duration. The few errors encountered were due to known auto-recoverable failures, such as "Out of storage space" and "Retry-able VSS errors."

## Protecting Hyper-V over CSV

Hyper-V over CSV (backed by SAN) is the most predominant deployment where VMs are hosted on a Hyper-V cluster with CSV storage. There is no limit to the number of disks a Hyper-V cluster can be configured to use, which allows much flexibility in designing the storage architecture of Hyper-V host clusters. For backing up the VMs, the DPM agent is installed on each cluster node, and you get reliable backups at scale with the latest version of DPM 2012 R2. There are no more host-level volume snapshots; it calls guest-level VSS to get application-consistent backups. DPM supports express full backups and parallel backups.

Now, consider the failure scenarios. What if the compute node goes down? All protected VMs on that host go into CC mode since the filter tracking is on that node. If there is a DPM VM on that host, all running jobs fail and are re-tried automatically. There is no impact on the jobs that have succeeded, and the jobs in queue continue to be in the queue even after the DPM VM moves to the new node in the cluster.



www.dbooks.org

Figure 5-2 shows the typical deployment used for protecting Hyper-V over CSV, and Table 5-2 shows the details of the scale test conducted in-house for the Hyper-V over CSV deployment. Scale testing for Hyper-V over CSV was performed, and the results were comparable to using Hyper-V over SOFS.

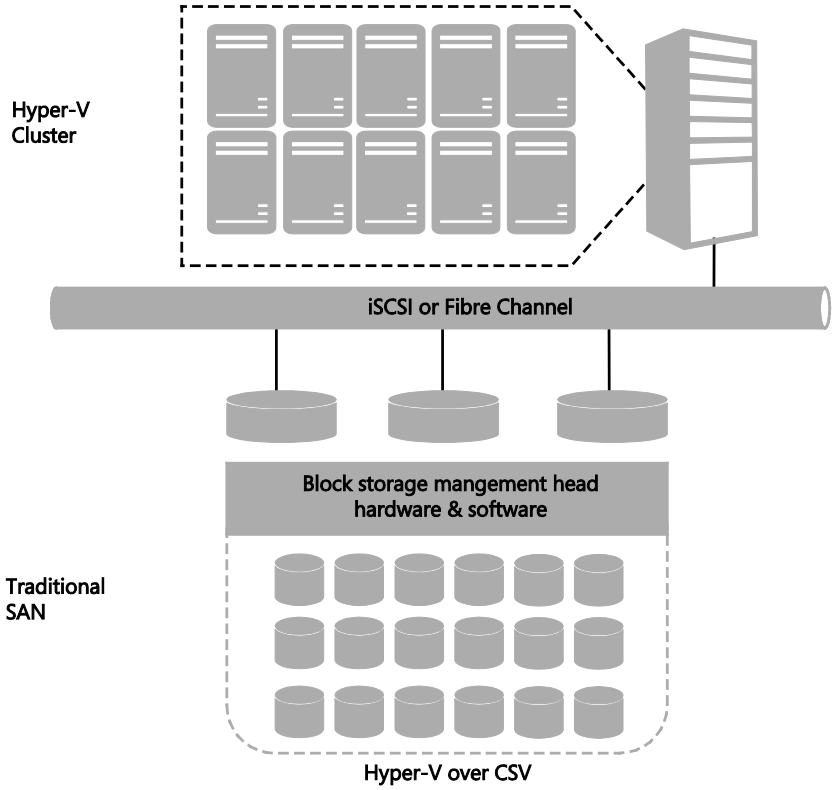



**FIGURE 5-2** Hyper-V over CSV deployment

**TABLE 5-2** Hyper-V over CSV configuration on which the scale tests were run by the product group

| CONFIGURATION | HYPER-V OVER CSV |
|---|---|
| Number of Hyper-V hosts | 12 |
| VM config (RAM) | 1 to 8 GB |
| VM disk size | 50 GB (20 GB for operating system + 30 GB for data) |
| Total number of VMs | 600 |
| SAN make/model | Dell Compellent SC8000 |
| Number of CSV cluster nodes | 12 |
| Number of virtual DPM servers | 2 |

***See also*** *Learn more about Hyper-V backup at private cloud scale at http://blogs.technet.com/b/dpm/archive/2014/08/12/hyper-v-backup-at-private-cloud-scale.aspx.*

## Continued protection with VM migration

VM migration refers to the process in which a running VM moves to another physical machine, keeping memory, network connectivity, and storage intact, which implies the application continues to run as-is without any disruption. This can happen due to a host server crash or due to re-balancing the resources in a cluster. So the expectation is that backups should run uninterrupted even if the VM moves to another node in the same cluster or even a different cluster. The same DPM server continues to protect the VM even after live migration. If the VM moves to a different cluster, DPM integration with Virtual Machine Manager helps discover the VM on a node in the new cluster.

***See also*** *Learn more about how to get uninterrupted data protection during live migration of VMs at http://blogs.technet.com/b/dpm/archive/2013/04/24/sc-2012-sp1-dpm-windows-2012-vm-mobility-uninterrupted-data-protection.aspx.*

## Protecting replica VMs

The first question that comes to customers' minds is whether they should protect the replica VMs. In customer conversations, three scenarios are found to be useful when protecting replica VMs:

- **Reduce the impact of backup on the production workloads** In today's world, workloads need to run 24 hours a day, 7 days a week with high performance. Workloads run on differencing disks during the backup operation and do impact the system.



www.dbooks.org

- **Limited network bandwidth between the branch office and head office** Network bandwidth is expensive, and it is redundant to send data to the head office twice, once for disaster recovery and then for backup. A more efficient way is to support backup from a replica site so that customers can manage the backup infrastructure from a single site.
- **Enterprise to hoster scenario** Most customers don't want to build another datacenter for disaster recovery and prefer to leverage a hosting service provider or public cloud. Hosting service providers offer SLAs around backing up customers' VMs on a regular frequency (for example, daily), and this can be easily achieved using replica VM backup.

There is some limitation with these scenarios: you can only get crash-consistent backups. However, most customers are comfortable with this. Remember, crash-consistency doesn't mean inconsistency. It's equivalent to the state when the power plug is pulled. Applications know how to recover from this state.

***See also*** *Learn more about the motivations, scenarios, and the product guidance at http://blogs.technet.com/b/virtualization/archive/2014/04/24/backup-of-a-replica-vm.aspx.*

## Protecting servers in workgroups and untrusted domains

DPM supports protecting Hyper-V VMs on the servers that use local user account (NTLM authentication) or that use certificates. For Hyper-V clusters, only certificate authentication is supported, not the NTLM certification. Even protection of a primary DPM server to a secondary DPM (in an untrusted domain) is supported through certificate authentication.

**NOTE** **Only standalone VMs are supported with certificate-based authentication, not the clustered VMs.**

***See also*** *Learn how to protect computers in untrusted domains and how to set up protection with certificate authentication at http://technet.microsoft.com/en-in/library/hh757801.aspx.*

# Recovering Hyper-V data

Protection is useless if you cannot recover in the case of disaster! While recovering a VM in case of a corruption seems to be the important restore use case, the most important scenario from a customer perspective is often that of recovering a file from a VM. That's because when a file is inadvertently deleted, the user tries to go back to an older point-in-time copy to get the file back. DPM server itself provides the additional capability of performing item-level recovery (ILR), which allows you to recover individual files, folders, volumes, and VHDs from a host-level backup. An advantage of using ILR is that the protection agent does not need to be installed on the guest operating system of the VMs on the host. Files recovered using ILR can be restored either to a network share or to a volume on a protected server.

> ***See also*** *Learn more about how to setup protection for VMs with SMB storage at http://technet.microsoft.com/en-us/library/hh757866.aspx.*

## How to restore a file from a VM

Restoring a file from a VM is the marquee feature of DPM. Based on customer conversations, it was found that most recoveries happen for files that were backed up within the last 7 to 14 days and most tenant requests are for files that were deleted inadvertently. The advantage of DPM is that the entire VM doesn't need to be recovered to get a single file. Instead, the VHD (which has the file) is mounted on the DPM server and the file is recovered to an alternate location.

> **NOTE** **Item-level recovery does not support recovery of an item to its original location.**

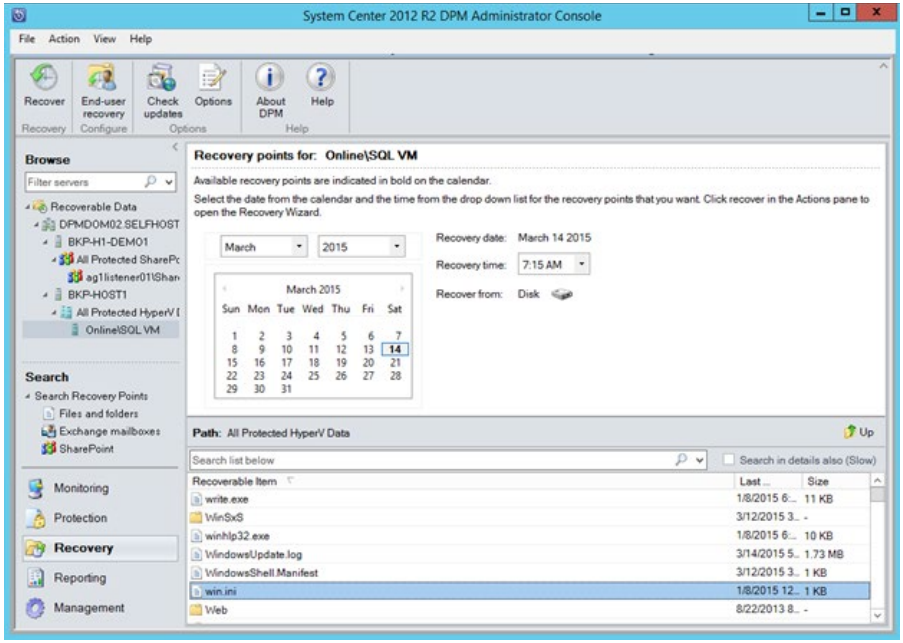
Complete the following steps to recover a file. These steps provide the recovered file when the VMs don't use the single parent disk.

1. Go to the Recovery workspace in the DPM console.
2. Select the VM from which you would like to recover the file.
3. Pick the date and time for the recovery.
4. Select the VHD and browse to the file you want to recover (e.g., win.ini).

> **NOTE** **You can recover multiple files from a single VHD but not from across multiple VHDs at the same time.**



www.dbooks.org




5. Right-click the selected file and select Recover. The Recovery Wizard opens.
6. On the Review Recovery Selection page, review the recovery item, date, time, media type, and then click Next.
7. On the Select Recovery Type page, specify the type as Copy To A Network Folder.
8. On the Specify Destination page, select the destination (volume or share) and it automatically populates the available space on the destination.
9. On the Specify Recovery Options page, configure specific options (e.g., security settings, network bandwidth usage throttling) for the recovery.
10. On the Summary page, review all the information, and then click Recover. If recovery succeeds, the recovery status shows as Successful on the Recovery Status page.

***See also*** *A description of various options for item recovery is available at https://technet.microsoft.com/en-us/library/hh757981.aspx.*

## How to restore a VM

In a private cloud deployment, it is recommended that you deploy all VMs with a golden image (a single parent VHD). Leveraging the same parent VHD and creating differencing disks for the VMs saves disk space. It is recommended that you keep the parent VHD on a high performant storage (for example, SSD) because most read operations will come from this VHD. DPM supports restoring the VM to the original location with or without a parent disk, but there are more steps involved with the latter option. The following steps show an example Alternate Location Recovery where you use DPM to restore a VM that has only a single parent disk and recover it to an alternate location. Start by getting the information from Virtual

Machine Manager on which DPM server protects the VM that needs to be recovered to the original location. (This is DPMServer01 in this example.)

1. Delete the corrupt VM:

```
Stop-VM -ComputerName HyperVHostName -Name VMName
Remove-VM -ComputerName HyperVHostName -Name VMName
```

   In the preceding command, *HyperVHostName* is the host on which the corrupt VM is located.

2. Create a symbolic link <SymbolicLinkOnHyperVHost> on the target server:

```
Enter-PSSession -ComputerName <RecoveryHost>
cd c:\
cmd /c "mklink /d link <SharedLocation>"
exit
```

   In the preceding command, *RecoveryHost* is the host on which the VM needs to be recovered and *SharedLocation* is the storage location for the recovered VM.

3. Connect to the DPM server, get the protection group on the DPM server named DPMServer01, and store the results in the $Pg variable:

```
$Pg = Get-DPMProtectionGroup -DPMServerName "DPMServer01" | where {$_.Name -eq
"PGName"}
```

4. Get the list of protected and unprotected data in the protection group and store the data source object:

```
$Ds = Get-DPMDatasource -ProtectionGroup $Pg | where {$_.Computer -eq
<ClusteredVMRoleName>}
```

5. Specify the recovery points for the given data source:

```
$Rps = Get-DPMRecoveryPoint -Datasource $Ds
```

6. Specify a particular recovery point:

```
$Rp = $Rps[$Rps.Length - 1]
```

   In the preceding command, *- 1* indicates the latest recover point; *- 2* would be the recovery point before that, and so on.

7. Recover the item to the alternate location (Hyper-V host):

```
$Rpo = New-DPMRecoveryOption -HyperVDatasource -TargetServer HyperVHostName -
RecoveryLocation AlternateHyperVServer -RecoveryType Recover -TargetLocation
<SymbolicLinkOnHyperVHost>
Recover-RecoverableItem -RecoverableItem $Rp -RecoveryOption $Rpo
```



www.dbooks.org

8. Connect to the host where the VM has been recovered and perform the following steps:

   a. Perform storage migration

   ```
   Move-VMStorage -ComputerName HyperVHostName -VMName VMName -
   DestinationStoragePath <SharedLocation>
   ```

   b. Re-parent the recovered VM to its original parent:

   ```
   Get-VMHardDiskDrive VMName | Get-VHD | where {$_.parentPath -ne $null} | Set-
   VHD -ParentPath <FullyQualifiedParentVHDLocation>
   ```

   c. Delete the local parent VHD that was just recovered.

   d. Delete the symbolic link ("c:\link") by opening Windows PowerShell session as an elevated user.

   e. Configure the recovered VM as a highly available source:

   ```
   Get-VM –Name VMName | Select VMId, ConfigurationLocation
   $res = Get-ClusterResource -Name "VMClusterResourceName" -Cluster
   ComputeClusterName
   Set-ClusterParameter -InputObject $res -Name VMId -Value <VMId> -Cluster
   ComputeClusterName
   ```

# Recommendations

The following list of recommendations is based on the changes that were introduced in DPM 2012 R2 UR3 to support Hyper-V backup at scale and on the results of various tests that were specifically conducted to develop this guidance:

- Use the default provider shipped with DPM 2012 R2 UR3 and later since the host-level volume snapshot on the production servers (Windows Server 2012 R2) was eliminated. There is no additional benefit if hardware providers are used.
- Deploy DPM on VMs running on Windows Server 2012 R2 Hyper-V servers.
- Provision backup storage through VHDs residing on SOFS shares and run DPM virtual to get the storage savings. DPM VHD/VHDX files should be 1 TB.
- Use multiple DPM servers, each protecting 250 to 300 VMs and leverage System Center Operations Manager for alerting, monitoring, and reporting. If you are especially concerned about reducing the DPM footprint, leverage the co-location feature and protect around 800 VMs, but note that there is a trade-off in terms of space savings.
- Use the Backup And Consistency Check window to control the backup.
- Do not overlap the backup/CC and de-duplication window since both are resource-intensive operations.

## Case study: Real-world customer

SaaSplaza, a hosting service provider of cloud systems for Microsoft Dynamics solutions, decided to switch to DPM. SaaSplaza used the Hyper-V technology in the Windows Server 2012 R2 operating system to create 5,000 VMs that run as a multisite private cloud between its Amsterdam; San Diego, California; Ashburn, Virginia; and Singapore datacenters. The biggest challenge was to meet the SLA of daily backups to SaaSplaza customers while controlling costs. The CEO said, "If you're a service provider, you have to have complete backups. You can't tell your customer, 'Sorry, we missed that day.'" The SaaSplaza staff spent tense hours going through backup logs trying to figure out which data was backed up and which wasn't.

The company decided to do a proof-of-concept with the latest version of DPM, which had three major improvements in areas that were important to them: speed, reliability, and de-duplication. They found the latest version of DPM to be much faster than previous versions and the backup product they were using currently. The CEO claimed, "Backups that took 10 hours before now take 3 hours. That saves us 7 hours in a 24-hour period. Because Data Protection Manager is faster, backups finish well within the allotted timeframe, which eliminates the problems that occur when backups carry over into business hours." DPM helped reduce backup times by 70 percent and costs by $500,000 a year.

***See also*** *Access the published case study at https://customers.microsoft.com/Pages/CustomerStory.aspx?recid=15016.*



www.dbooks.org

CHAPTER 6

# VMware private cloud protection

This information is not yet publicly available. It will be included when this ebook is re-issued in summer 2015.

www.dbooks.org

CHAPTER 7

# Protecting the Microsoft Cloud Platform System

Microsoft has extensive experience running some of the largest datacenters and cloud services across the globe. Microsoft Cloud Platform System (CPS) is an effort to bring Microsoft's cloud expertise to the customer's datacenter in an easy-to-deploy, fully validated converged system. CPS combines Microsoft's proven software stack of Windows Server 2012 R2, System Center 2012 R2, and Windows Azure Pack, with Dell's cloud server, storage, and networking hardware. As a scalable building block for your organization's cloud solution, CPS shortens the time to value and enables a consistent cloud experience. This chapter focuses on protecting the management cluster and tenant virtual machines (VMs).

> ***See also*** *You can learn more about CPS at http://www.microsoft.com/en-us/server-cloud/products/cloud-platform-system/.*

## Protecting the management cluster

The management cluster is the heart of CPS. It is a six-node physical Hyper-V cluster that hosts the resources for fabric and service management for a CPS stamp. A CPS stamp is a complete management and hosting domain that can range from a minimum of one rack to a maximum of four racks. All of the physical nodes must run Windows Server 2012 R2 Datacenter edition, typically in the Server Core configuration.

> ***See also*** *For more information on the management cluster and its performance, see http://blogs.technet.com/b/privatecloud/archive/2015/02/09/cps-management-cluster-and-its-performance.aspx.*

It is important to protect all of the features of the management cluster to ensure the fabric is up and running in a consistent manner in case of disaster (user errors, data corruption, and so on). CPS includes deployment of Microsoft System Center Data Protection Manager (DPM) pre-configured to back up the databases, VMs, and other critical data in all features of the management cluster. The DPM server on the management cluster uses a SQL Server database that is local to the DPM server VM. The only features that are not protected by DPM are the infrastructure VMs (AD/DNS/DHCP) since DPM has Active Directory dependency.

After CPS installation, regular backups automatically start and continue according to predefined settings. You can use these backups to restore data and functionality if there is a failure in the CPS environment.

Figure 7-1 shows the various features in the management cluster and how they are protected. The management cluster features include Active Directory, Domain Name Service (DNS), Dynamic Host Control Protocol (DHCP), Active Directory Federation Services (ADFS), Windows Deployment Services (WDS), Windows Server Update Services (WSUS), Windows Azure Pack (WAP), Infrastructure-as-a-Service (IaaS) Resource Provider (RP), Service Management Automation (SMA), Virtual Machine Manager (VMM), Operations Manager (OM), and other features. A SQL Server cluster consisting of four VMs is used for storing data relating to these features, and DPM is used for backing up the databases. Numbers in parenthesis indicate the number of VMs for that feature.

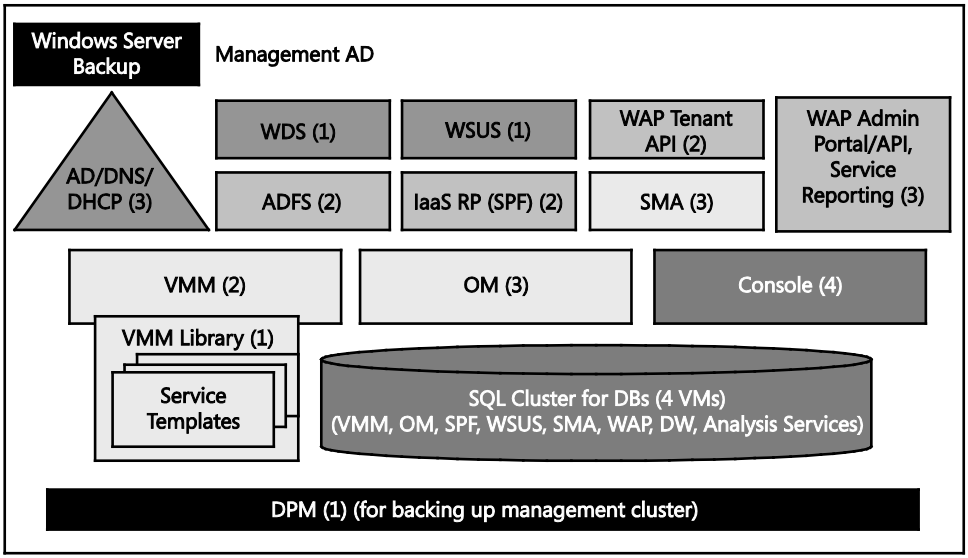



**FIGURE 7-1** Management cluster features in CPS

## Default protection policy

Table 7-1 summarizes the default schedule for backups and the retention periods for the various features of the CPS management cluster.



www.dbooks.org

**TABLE 7-1** Backup schedule and retention period for management cluster features

| FEATURE | SCHEDULE | RETENTION PERIOD |
|---|---|---|
| All VMs<br>WSUS<br>WDS<br>All System Center features<br>Console VMs<br>Windows Azure Pack<br>Features running in management cluster<br>Features running in compute cluster | Once per week (8:00 PM every Saturday) | Two weeks |
| All databases:<br>WSUS<br>All System Center features<br>Windows Azure Pack | Express full backup every four hours (starting at 12:00 AM) | Five days |
| Active Directory/DNS/DHCP using Windows Server Backup | Once per day | Five days |

Typically, you should not have to change any of these settings. However, you can change the day of backup or time of day if you need to. If you do change the schedule, make sure that you maintain the following backup frequency:

- **VMs** Once per week backup with a two-week retention period.
- **Databases** Once every four hours backup with a five-day retention period.

**TIP Don't change the number of recovery points because this will impact storage calculations in CPS. However, do use the DPM Create Recovery Point Now option to back up VM databases before and after you make any configuration change.**

DPM is used to back up the system databases in the SQL Server instance, but it does not back up its own database. Instead, the DPM database is backed up using Microsoft SQL Server Backup. The backup frequency is every four hours, and the retention period is one day. Details concerning the DPM server database for the DPM server that backs up the management cluster are as follows:

- **DPM server name** <Prefix>-DPM-01
- **SQL server instance name** SCDPM
- **Database name** DPMDB_<Prefix>_DPM_01

## Recovering VMs and databases

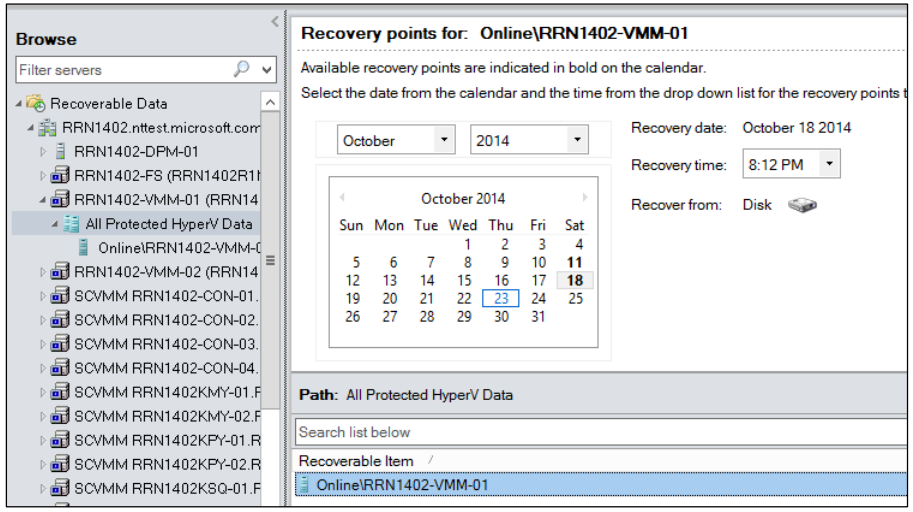
This section presents the general steps for VM and database recovery for management cluster features and for the Windows Azure Pack public features that are on the compute cluster.

> **TIP** When you recover a CPS feature, you must follow the feature's specific recovery steps as described in the "Recovering from management cluster component failures" section in the CPS Admin Guide.

### Recovering VMs to their original location

The following steps use the DPM administrator console to recover VMs to their original location:

1. Open the DPM administrator console, click Recovery, and then, in the Recoverable Data pane, browse to the VM instance that you need to recover.
2. Select the respective Hyper-V host and expand that node.
3. Click All Protected Hyper-V Data, and then, in the Recovery Points pane, select the VM that you want to recover, as shown in the following image.




4. Click any date and time in the calendar to see available recovery points. Dates that show as bold have active recovery points. To minimize data loss, it is important to choose to recover from the latest possible recovery point.
5. Click Recover in the Actions pane to launch the Recovery Wizard.
6. In the Recovery Wizard, on the Select Recovery Type page, select Recover To Original Instance.
7. On the Specify Recovery Options page, leave the Network Bandwidth Usage Throttling and SAN Recovery selections without any modifications.



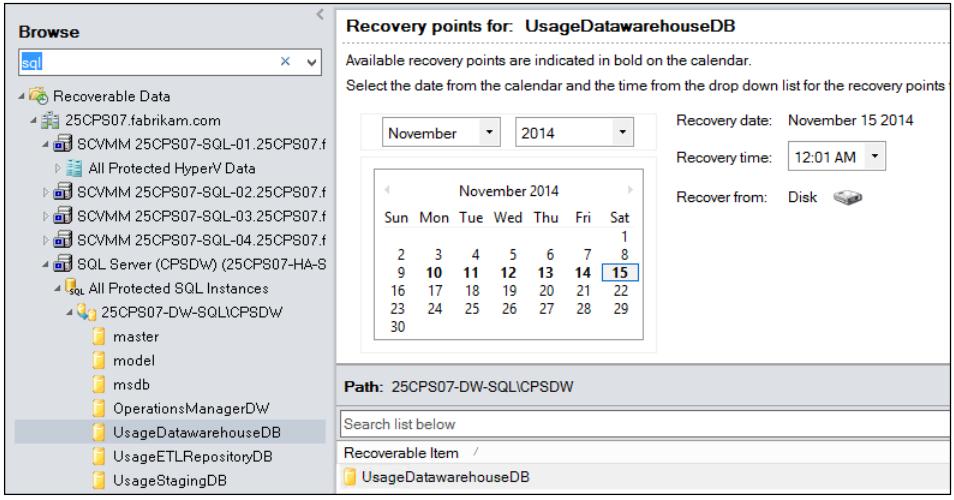
www.dbooks.org

8. On the Summary page, review your settings, and then click Recover.
9. After the recovery completes, click Close to close the Recovery Wizard.
10. Start the VM that you just recovered.
11. After you recover the VM, you must synchronize it as follows (the protection status of this VM will show as Replica Inconsistent until it is synchronized):
    a. In the DPM administrator console, open the Protection workspace.
    b. Locate and then right-click the recovered VM. Click Perform Consistency Check.
    c. In the Microsoft System Center 2012 R2 Data Protection Manager dialog box, click Yes to perform the consistency check.

## Recovering a database to its original location

As part of a recovery process, recovering databases by using DPM is an important task. While the procedure that follows provides the general steps for database recovery, when you recover a CPS feature, you must follow the feature's specific recovery steps as described in the "Recovering from management cluster component failures" section in the CPS Admin Guide.

1. In the DPM administrator console, click Recovery in the navigation bar.
2. In the Recoverable Data area, expand SQL Server, All Protected SQL Instances, expand the instance of SQL Server that hosts the database that you want to recover, and then select the appropriate database, as shown in the following image.




3. In the Recovery Points pane, select a recovery point by clicking any date and time in the calendar to see available recovery points. Dates that are shown in bold have active recovery points. To minimize data loss, it is important to choose to recover from the latest possible recovery point.
4. Click Recover in the Actions pane to launch the Recovery Wizard.

5. In the Recovery Wizard, on the Select Recovery Type page, select Recover To The Original Instance Of SQL Server (Overwrite Database), and then click Next.
6. On the Specify Database State page, keep the default setting of the Leave Database Operational option.
7. On the Specify Recovery Options page, leave the Network Bandwidth Usage Throttling and SAN Recovery settings without any changes.
8. On the Summary page, review your settings, and then click Recover.
9. After the recovery completes, click Close to close the Recovery Wizard.
10. After you recover a database, it must be synchronized by DPM. The protection status of this database will be Replica Inconsistent until you synchronize it as follows:
    a. In the DPM administrator console, click Protection in the navigation bar.
    b. Right-click the recovered database, and then click Perform Consistency Check.
    c. In the Microsoft System Center 2012 R2 Data Protection Manager dialog box, click Yes to perform the consistency check.

## Using the CPS database consistency runbooks

In case of a database restore for a feature, there may be inconsistencies between various databases that store information about Windows Azure Pack plans, subscriptions and add-ons, and service quotas and limits. CPS includes runbooks that you can use to detect and, in some cases, to automatically recover from data inconsistencies between the databases of the features that are deployed as part of CPS. The data consistency runbooks are automatically imported during the CPS installation process and are available in SMA through the Windows Azure Pack management portal for administrators.

You can run the master runbook Invoke-DataConsistency to detect and recover from inconsistencies between features' state when there is live traffic on the system. However, you must first run a script to reset the passwords across the system. This requires downtime of the whole stamp.

Use the CPS data consistency runbooks for detection and recovery after you restore a management feature database from backup. For example, you may have restored a database to fix data corruption.

Table 7-2 summarizes the data consistency checks that the detection runbooks can perform and the features that are involved for each check. For more details, refer to the CPS Admin Guide.



www.dbooks.org

**TABLE 7-2** Data consistency checks for management cluster features

| DATA CONSISTENCY CHECKS | FEATURES |
|---|---|
| VM Cloud resource provider objects | Windows Azure Pack<br>SPF<br>VMM |
| VM Cloud usage (billing impact) | Windows Azure Pack (Service Management API, Usage)<br>SPF<br>Operations Manager<br>VMM |
| VM Cloud usage analytics | Service Reporting<br>Windows Azure Pack<br>SPF<br>Operations Manager<br>VMM |
| SQL Server admin/tenant/usage | Windows Azure Pack (Service Management API, SQL Server resource provider) |
| MySQL admin/tenant/usage | Windows Azure Pack (Service Management API, MySQL resource provider) |

# Recovering from failures of management cluster features

This section describes how to recover from data failures of certain management cluster features in the CPS environment and lists the steps that are specific to database recovery, as well as the steps that are specific to VM recovery. Follow the respective steps for each feature listed in the CPS Admin Guide for database or VM recovery as needed.

**IMPORTANT** **Perform all procedures in this section by using an account that is a member of the <Prefix>-Ops-Admins group.**

## Recovering Operations Manager

Microsoft System Center Operations Manager plays a key role in monitoring the entire CPS environment. If you have exhausted all options when trying to recover from Operations Manager failures, you can recover Operations Manager data to restore functionality.

To recover the Operations Manager database, complete the following steps:

1. From the Microsoft System Center Virtual Machine Manager (VMM) console, shut down all Operations Manager VMs that are located on the management cluster.

2. Use the steps in the section "Recovering a database to its original location" to recover the following Operations Manager databases:

| SQL SERVER INSTANCE | DATABASE NAME |
|---|---|
| SCOMDB | OperationsManager |
| CPSDW | OperationsManagerDW |

   To minimize data loss, be sure to select the latest recovery point.

3. For each SQL Server instance, enable Service Broker by doing the following:
   a. On the Console VM, open SQL Server Management Studio.
   b. Connect to the following SQL Server cluster and instance:

| SQL SERVER CLUSTER NAME | SQL SERVER INSTANCE |
|---|---|
| <Prefix>-OM-SQL | SCOMDB |
| <Prefix>-DW-SQL | CPSDW |

   c. Connected to one of the instances that is specified in the table, run the appropriate command for that instance. For example, for the SCOMDB instance, run the following command:

```
ALTER DATABASE OperationsManager SET ENABLE_BROKER
```

      For the CPSDW instance, run the following command:

```
ALTER DATABASE OperationsManagerDW SET ENABLE_BROKER
```

4. Start the Operations Manager VMs (<Prefix>-OM-01, -OM-02 and -OM-03).
5. Detect and repair any data consistency issues by following the required steps in the “How to use data consistency runbooks” section in the CPS Admin Guide.

To recover the Operations Manager VMs, complete the following steps:

1. From the VMM console, shut down all Operations Manager VMs in the management cluster.
2. Use the steps in the "Recovering VMs to their original location" section to recover the three Operations Manager VMs. To minimize the data loss, select the latest recovery point.
3. Restart the Operations Manager VMs.
4. Detect and repair any data consistency issues by following the required steps in the “How to use data consistency runbooks” section in the CPS Admin Guide.



www.dbooks.org

## Recovering Virtual Machine Manager

VMM plays a key role in managing the hosts and VMs in the CPS environment. If you have exhausted all options to try to recover from application failure, you can use DPM to recover the VMM database to an older point in time.

To recover the VMM database, complete the following steps:

1. From the console VM, open Failover Cluster Manager.
2. Connect to the management cluster.
3. Shut down the two VMM VMs (<Prefix>-VMM-01, -VMM-02) that are located on the management cluster.
4. Use the steps in the section "Recovering a database to its original location" to recover the VMM database (called VirtualManagerDB in SCSHAREDDB SQL Server instance). To minimize data loss, be sure to select the latest recovery point.
5. Open Failover Cluster Manager, and connect to the management cluster.
6. Start the VMM VMs.
7. In the VMM console, verify that the content in the Fabric workspace is updated.
8. Detect and repair any data consistency issues by following the required steps in the "How to use data consistency runbooks" section in the CPS Admin Guide.

To recover the VMM VMs, complete the following steps:

1. From the console VM, open Failover Cluster Manager.
2. Connect to the management cluster.
3. Shut down the two VMM VMs (<Prefix>-VMM-01, -VMM-02) that are located on the management cluster.
4. Use the steps in the "Recovering VMs to their original location" section to recover the VMM VMs. To minimize data loss, be sure to select the latest recovery point.
5. In Failover Cluster Manager, connect to the management cluster, and then click Roles. In the Roles pane, right-click each VMM VM, and then click Start.
6. In Failover Cluster Manager, connect to the VMM guest cluster <Prefix>-HA-VMM, and then click Roles. If the <Prefix>-HA-VMM clustered role is not running, right-click the role, and then click Start Role.
7. Detect and repair any data consistency issues by following the required steps in the "How to use data consistency runbooks" section in the CPS Admin Guide.

## Recovering infrastructure VMs

There are three infrastructure VMs (for Active Directory, DNS, and DHCP) in the management cluster. All of them are backed up locally by using Windows Server Backup.

If one or more of the Active Directory/DNS/DHCP instances fails because of corruption or deletion of critical directories, you can use bare metal recovery to recover the instance. The

procedure for using bare metal recovery to recover a single instance is described in this section. If there are multiple instance failures, you must repeat this procedure sequentially for all failed instances.

1. From the VMM console, connect to the failed domain controller VM. Boot the Active Directory/DNS/DHCP server into Windows Recovery Environment (WinRE). The server automatically boots into WinRE if it fails to boot into normal mode twice. If the server boots normally, run the following commands at a command prompt to restart in WinRE mode:

   ```
   reagent /boottore
   shutdown /r /t 0
   ```

2. In WinRE mode, click Troubleshoot.
3. On the Advanced Options screen, click System Image Recovery.
4. Select the Administrator account on the System Image Recovery screen.
5. Type the password on the next screen.
6. In the Re-image Your Computer Wizard, you can see the latest available system image for recovery. If you want to recover to an older point in time, click Select A System Image, and choose the desired point in time. Click Next.
7. Click Next on the Choose Additional Restore Options page.
8. Click Finish to complete the Re-image Your Computer Wizard. The following screens display the progress of the recovery as all volumes are restored.
9. In the dialog box that is displayed, click Restart to restart the computer.
10. After recovery completes, schedule full server backups by using the Windows Server Backup tool, as described in Configure Automatic Backups to a Volume at *http://technet.microsoft.com//library/dd851674.aspx*. You can do this as follows:
    a. On the Select Backup Configuration page, click Full Server (recommended).
    b. On the Specify Backup Time page, click Once A Day, and then select 12:00 AM as the backup time.
    c. On the Specify Destination Type page, click Back Up To A Volume.
    d. On the Select Destination Volume page, select Local Disk (F:) as the destination volume.

## Protecting tenant VMs

Customers purchase CPS to get Azure-consistent cloud-in-a-box and provide infrastructure as a service (IaaS). This allows tenants to create VMs on an as-needed basis using the service provider's portal. These providers know that providing just the infrastructure is a race-to-the-bottom proposition and that they can make money only by providing value-added services



www.dbooks.org

(like backup and disaster recovery as a service). These service providers offer daily backup SLAs to their customers and any violation leads to financial penalties. This very important customer need to back up reliably at scale (up to 2,000 VMs in a rack or 8,000 VMs in a 4-rack stamp) was delivered in the tenant VM protection feature built into CPS. In addition, backup storage was deduplicated by default to reduce the overall costs for the customers.

This section covers the DPM server configuration for tenant backup, how to add new VMs that are added on daily basis, and how to recover these tenant VMs in case of a disaster.

## Using DPM servers for tenant backup

By default, the CPS installation process provisions eight tenant DPM servers per rack that can be used for tenant backup. These servers are deployed to the compute clusters and use the naming convention DPM-TenantVM-0# (-01 through -08 on the first rack, -09 through 16 on the second rack, and so on). All of these DPM servers are pre-configured and ready to protect.

To provide spindle isolation and to keep backups on a separate pool, one storage pool on each rack is assigned for backup. This backup pool is configured as dual parity to maintain N+2 redundancy and has total usable capacity of 115.2 TB. Each tenant backup DPM server is provisioned with 20 TB of allocated disk space (20 VHDs of 1 TB each) that is provisioned on 15.4 TB of physical space. This difference between allocated and physical disk space is addressed by data deduplication that runs on the backup pool.

## Adding tenant VMs to backup

As the new VMs are deployed on the CPS stamp, customers can run a runbook (called Protect-TenantVMs) to protect new tenant VMs that were just created. All VMs are configured to protect once daily with a retention period of one week. Test VMs that do not need DPM protection can be excluded by specifying an exclusion VM list using a runbook (called Add-DPMExclusionItems).

You must run the Protect-TenantVMs runbook to manage tenant VM protection. This runbook adds up to 75 newly created VMs to a protection group in DPM. You should run this runbook manually or through a scheduled task once each day. After a tenant VM is added to a protection group, by default, the tenant VM is configured for daily backup, with a retention period of seven days. This runbook is designed to protect 75 new VMs per run per day to ensure enough time to complete tenant VM backups in the backup window and enough time for the deduplication process to complete. If more than 75 new VMs were created (on one rack) and you need to add them to a protection group on the same day, you can run this runbook more than once to protect the additional VMs.

The data deduplication process reduces backup storage usage. There is a default schedule for data deduplication and for tenant backups. Table 7-3 shows the default schedule.

**TABLE 7-3** Default schedules for deduplication and backup

| WINDOW TYPE | DEFAULT SCHEDULE |
|---|---|
| Deduplication | 6:00 AM to 10:00 PM |
| Backup | 10:00 PM to 6:00 AM |

You should plan to run the Protect-TenantVMs runbook so that it does not interfere with the backup window. Therefore, run it any time between 6:00 AM and 6:00 PM local time (at least three to four hours before the backup window starts).

If you need to prevent protection of some VMs, you can run the Add-DPMExclusionItems runbook and specify VM names (wildcard characters are supported) that should be excluded during VM protection.

## Recovering tenant VMs

All tenant VMs are deployed with a single parent VHD. DPM's original location recovery workflow will not work for tenant VMs. Complete the following steps to recover tenant VMs:

1. In the VMM console, determine the name of the host on which the VM that you want to recover is located by doing the following:
   a. In the VMs And Services workspace, expand All Hosts, and then click Compute Clusters.
   b. In the VMs pane, type the name of the VM.
   c. Note the value in the Host column that is associated with the VM.
   d. Note which compute cluster the host is a member of. (Under Compute Clusters, click each cluster to view the members.)
   e. Right-click the VM, and then click Properties. Click the Hardware Configuration tab. Under Bus Configuration, the VHDs that are attached to the VM are listed. Click the operating system VHD (typically the first one under IDE Devices) to see if there is a VHD chain. Note the value in the Fully Qualified Path To Parent Virtual Hard Disk box (for example, copy and save it to Notepad). If the VM properties are corrupted and you cannot access them, you can skip this step.
2. In the VMM console, find a tenant share that has enough available capacity to store the recovered VM by doing the following:
   a. In the Fabric workspace, expand Storage, and then click File Servers.
   b. In the File Servers, File Shares pane, expand the file server that is in the same rack as the compute cluster where the Hyper-V host that you identified in step 1c resides.
   c. Use the Available Capacity column to find a TenantShare with enough free space. (This procedure uses the example share \\<Prefix>-FS-02.contoso.com\TenantShare14.)



www.dbooks.org

3. On the Console VM, open Failover Cluster Manager, and connect to the compute cluster on which the host that you identified in step 1c is a member of.
4. Under the cluster name, click Roles.
5. In the Roles pane, find the cluster resource name of the VM that you want to recover. The name will be in the format SCVMM VMName Resources.
6. On the Console VM, open Windows PowerShell, and run the following commands to delete the VM. Press Enter after each command. Note that the Hyper-V host is the host on which the VM that you want to recover is located.

```
Stop-VM -ComputerName HyperVHostName -Name VMName
Remove-VM -ComputerName HyperVHostName -Name VMName
```

7. Create a symbolic link to the tenant share that you identified in step 2 by first running the following command:

```
Enter-PSSession -ComputerName HyperVHostName
```

   In the remote session, run the following commands:

```
cd c:\
cmd /c "mklink /d DirectoryName \\SharePath"
exit
```

8. On the Console VM, find the DPM server that backs up the VM that you want to recover. To do this, complete the following steps:
   a. Open the Operations console.
   b. In the Monitoring workspace, expand System Center 2012 R2 Data Protection Manager, select State Views, and then click Protected Servers.
   c. In the Look For box, enter the cluster resource name of the VM.
   d. In the DPM server column, note the name of the DPM server that backs up the VM.

   You can also do this by running the following Windows PowerShell command from the Operations Manager Shell:

```
Get-SCOMClassInstance | where {$_.DisplayName -like '*clusterresourcename*'} |
foreach { $_.'[Microsoft.SystemCenter.DataProtectionManager.
```

9. Open the DPM administrator console, and connect to the DPM server that you identified in step 8. Find and note the name of the protection group that the VM that you want to recover was added to.

10. On the Console VM, recover the VM by running the following Windows PowerShell commands as an elevated user. Press Enter after each command. Note that DPM-TenantVM-0# is the name of the DPM server that you identified in step 8, ProtectionGroupName is the protection group that the VM is a member of, VMName is the NetBIOS name of the VM that you want to recover, and SymbolicLinkOnHyperVHost is the symbolic link that you created earlier, for example c:\test1.

```
$pg = Get-DPMPRotectionGroup -DPMServerName DPM-TenantVM-0# | where {$_.Name -eq
"ProtectionGroupName"}
$ds = Get-DPMDatasource -ProtectionGroup $pg | where {$_.Computer -eq "VMName"}
Get-DPMRecoveryPoint -Datasource $ds | select Name, BackupTime      ## this is
used for display only
$rps = Get-DPMRecoveryPoint -Datasource $ds
$rpo = New-DPMRecoveryOption -HyperVDatasource -TargetServer HyperVHostName -
RecoveryLocation AlternateHyperVServer -RecoveryType Recover -TargetLocation
<SymbolicLinkOnHyperVHost>
$rp = $rps[$rps.Length - 1] ## Value of -  1 indicates the latest recover point.
A value of - 2 would be the recovery point before that.
$ri = Get-DPMRecoverableItem $rp -BrowseType Child
Recover-RecoverableItem -RecoverableItem $rp -RecoveryOption $rpo
```

11. On the Hyper-V host on which the VM is located, open Windows PowerShell as an elevated user.
12. Perform a storage migration by running the following command where SOFSShare is the share that you identified in step 2.

```
Move-VMStorage -ComputerName HyperVHostName -VMName VMName -
DestinationStoragePath SOFSShare
```

13. Re-parent the VM to its original parent that you identified in step 1e by running the following command. (You can skip this step and continue to step 15 if the original VM configuration was corrupted and you could not get this property value in step 1e.)

```
Get-VMHardDiskDrive VMName | Get-VHD | where {$_.parentPath -ne $null} | Set-VHD
-ParentPath "\\SharePathofParentVHD"
```

14. Delete the "local" parent VHD (that was just recovered).
15. Delete the symbolic link. To do this, open a Windows PowerShell session as an elevated user, and then run the following commands. (Press Enter after each command.)

```
Enter-PSSession -ComputerName HyperVHostName
del DirectoryName
exit
```



www.dbooks.org

16. From a Console VM, run the following Windows PowerShell commands to configure the VM as highly available. Press Enter after each command. (You must connect the VM to its original cluster resource role.) Note that in the following commands, VMClusterResourceName is the cluster resource name for the VM (for example "SCVMM VMName Resources"), ComputeClusterName is the compute cluster name on which the Hyper-V host resides, and VMConfigLocation is the location that is identified in the Get-VM command that you run in this procedure.

```
Get-VM –Name VMName | Select VMId, ConfigurationLocation
$res = Get-ClusterResource -Name "VMClusterResourceName" -Cluster
ComputeClusterName
Set-ClusterParameter -InputObject $res -Name VMId -Value <VMId> -Cluster
ComputeClusterName
Set-ClusterParameter -InputObject $res -Name VmStoreRootPath -Value
"VMConfigLocation" -Cluster ComputeClusterName
```

# Monitoring backups

DPM automatically backs up data according to the backup settings. At times, DPM backups might fail. For example, a communication failure may occur with the SQL Server servers, they might not be running, or there might be other connection issues.

Regularly monitor DPM. CPS installs DPM, the DPM Central Console, and the management pack that is integrated with Operations Manager. You can monitor all failure alerts in Operations Manager. To do this, complete the following steps:

1. Log on to one of the Console VMs by using an account that is a member of the <Prefix>-Ops-Admins group.
2. Open the Operations console.
3. In the Monitoring workspace, expand System Center 2012 R2 Data Protection Manager to see the Alert and State views and any associated alerts.

The DPM Central Console is integrated with Operations Manager (and does not show up as a separate interface). It enables you (through the System Center 2012 R2 Data Protection Manager node) to monitor all management cluster and tenant DPM servers and to take actions in response to alerts.

> ***See also*** *For more information about managing multiple DPM servers with Central Console, see http://technet.microsoft.com/library/jj860391.aspx.*

Recovering a DPM server includes leveraging DPM database backups. Each DPM server is scheduled to back up the database every four hours by using the Windows DPMDBBackup scheduler job.

Because this backup job can fail for many reasons, you should regularly monitor the status of this job on all DPM servers. The most common possible failures are summarized in Table 7-4.

**TABLE 7-4** Common backup failures and their resolutions

| ISSUE | SYMPTOM | RESOLUTION |
|---|---|---|
| The request to SQL Server to export and copy the DPM database fails. | Event Viewer displays a SQL Server database error. | Fix the SQL Server error, and then run the DPMDBBackup job again. |
| Drive runs out of space. | Event Viewer displays a Disk Full error. | Delete the oldest file (and make a copy if necessary). Then, run the DPMDBBackup job again. |

## Case study: A real-world CPS customer

This chapter concludes with a story from a real-world customer, NTTX Select, a hosting service provider offering customized, managed Microsoft-based IT solutions for businesses throughout the United Kingdom. They use Microsoft Cloud Platform System to create IaaS offerings in days (instead of months), directed at government and highly regulated industries that required UK government certified cloud solutions. They are able to better scale and protect workloads and focus staff on new services rather than infrastructure configuration and coordination. They use a two-rack CPS stamp in which one DPM server protects the management cluster and eight DPM servers protect their tenant VMs.

NTTX Select's customers expect them to maintain their critical workloads and to scale and protect those workloads. With CPS, NTTX can better fulfill those expectations and grow customer trust. "CPS will soon be a known commodity in the industry, and when customers hear that we are using it, they will know that they are getting the highest possible quality in terms of availability, performance, and scale in a Microsoft environment," says Philip Moss, Chief Executive Office and Chief Technology Officer of NTTX Select.

> ***See also*** *You can read the published case study at https://customers.microsoft.com/Pages/Download.aspx?id=10642.*



www.dbooks.org

CHAPTER 8

# Optimizing backup storage

This chapter begins by describing the implication of exponential growth in backup data on the cost of maintaining backups. It then explores the options available to users of Microsoft System Center Data Protection Manager (DPM) and Azure Backup to reduce the amount of storage consumed. Finally, the chapter briefly delves into setting up deduplication for DPM.

## Exponential growth in backup storage

Backup storage is growing at an exponential rate, both in terms of the size of the data being protected and the duration for which the backup data must be retained. A simplistic formula encapsulates this notion:

*backup storage requirement = InitialSize + (InitialSize × DailyChurn% × RetentionPeriodInDays)*

The formula above assumes that backups are taken daily and retained for a fixed period of time (in days) and then removed. The amount of data that is generated for each backup is captured as a percentage of the initial data using the variable DailyChurn%.

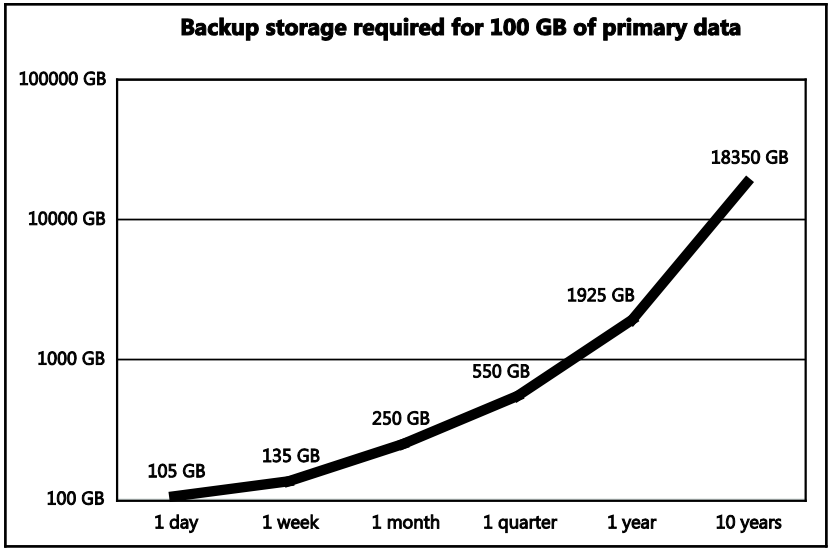



**FIGURE 8-1** Graph showing the exponential storage requirements for retaining daily backup points at increasingly longer retention periods

To illustrate, apply this formula to the backup of a single 100-GB virtual machine. Assuming a daily churn rate of 5 percent and a one-week retention period, the total backup storage requirement is 135 GB. If the retention increases from one week to one month, then the backup requirement increases to 250 GB, and with a three-month retention requirement, the backup storage requirement is 550 GB. Add to this a few "full copies" of the data for resiliency and faster restore and you quickly start running into a management and cost nightmare for protecting even small amounts of primary data.

With a market rate of approximately 27 cents per GB of storage (based on $280 as the cost of 1 TB of SAS storage), the cost of primary storage for 100 GB of data is $27, while the cost of backing up the 100 GB of data with one month of retention is $67.50! Thus, it makes great sense for administrators to focus on reducing the cost per gigabyte of storing backup data.

# Containing the cost of backup storage

Administrators use a combination of hardware and software to reduce the cost per gigabyte of storing backup data. Typical strategies adopted include the following:

- Using cheaper media, such as tape, to store backup copies
- Reducing the cost of backup copies stored, for example, by moving from double- to single-drive redundancy or by increasing the total drive count in a given RAID array configuration.
- Reducing the performance of backup storage media, by using SATA disks instead of SAS disks for example
- Compressing stored backup data
- Deduplicating stored backup data

## Software approaches to reducing stored backup data

The two key software approaches that yield storage savings are compression and deduplication. The fundamental difference between these two is the scale and size of data on which they operate.

Deduplication is essentially "removing duplicates" in the stored data. Backup data has two parts—the initial data and the daily churn. While the initial data is expected to have some amount of duplicate data, there also tends to be significant deduplication of data in the daily churn. Of course, this depends on the workloads and datasets being run. On the whole, it is reasonable to expect between 20 and 60 percent savings across a wide variety of configurations. A great example of the value of deduplication is the savings achieved when applied to the backup of virtual machines. Each virtual machine consists of at least 10 GB of the operating system binary files. These files change very little from virtual machine to virtual machine. If 1,000 virtual machines are backed up, then the total data size occupied by the operating system binary files is approximately 10 TB! Virtual machines typically show the



www.dbooks.org

largest deduplication savings rates—to the tune or 90 percent or more—and most of it comes from deduplication of the initial backup data. This reduction is enormous and can potentially save thousands of dollars in storage costs alone. For this reason, any enterprise-grade backup solution needs deduplication capabilities.

Compression works at a much smaller scale. Chunks of data can be reduced if some commonalities are found within the chunk. These chunks of data could be as small as 4 KB or as large as 2 MB. This is the key difference between compression and deduplication. Deduplication covers the whole backup data to find common chunks, while compression takes each chunk and attempts to reduce the data.

## Deduplication

Deduplication is an established technology in the backup domain for reducing storage costs. However, the additional processing and memory required for deduplication is the price that must be paid in exchange, and a balance between the compute overhead and storage savings must be established.

The authoritative storehouse of the original data is called the *chunk store*. With deduplication enabled, all backup recovery points would be references to data lying in the chunk store, plus any additional unduplicated data.

Two key questions determine the approach to deduplication processing that is provided by the backup product:

1. Where is the deduplication processing done (or) who incurs the compute overhead of deduplication?
2. Where is the deduplicated chunk store stored?

Based on the answers to these two questions, a few distinct types of deduplication can be specifically applied to backup data. Two sub-types of deduplication are based on where the deduplication processing is done:

- Source deduplication
- Target deduplication

## Source deduplication

In source deduplication, the deduplication processing is done at the primary location (source). The overhead of deduplication is incurred by the production server, and the deduplication happens before any data transfer to the backup location (target).

If a given block of data has already been backed up previously and exists in the chunk store, then it is not copied again over the network to the backup location. Instead, the backup recovery point contains a reference to the existing data in the chunk store.

If a given block of data has not been backed up previously then it is copied to the target side and inserted in the chunk store. This is done "inline" as a part of the backup workflow.

Inherently, source deduplication also provides the benefit of network deduplication.

### Target deduplication

In target deduplication, the deduplication processing is done at the backup location (target). The overhead of deduplication is incurred by the backup infrastructure, and the deduplication happens after the data is transferred between the primary location and the backup location. After the data arrives on the target, it can be processed post-facto as a batch job after the backup job has completed or "inline" as a part of the backup workflow.

Target deduplication does not provide any network-related benefits. The chunk store always resides on the target side in target deduplication.

## Azure Backup

Azure Backup uses the Azure Storage subsystem in its raw form, and the Azure Storage fabric is highly optimal in storing data within the recovery points of a data source. Instead, Azure Backup adds value to this already optimized storage stack by compressing the data written to Azure Storage. This storage could be further optimized in the future by adding deduplication across data sources.

## Data Protection Manager

The deduplication style that DPM uses is essentially target deduplication. DPM uses Data Deduplication, a feature of Windows Server 2012 and later, to achieve a high level of deduplication savings.

Both DPM and Azure Backup use some variation of target deduplication. This means that the product workload is unaffected by the deduplication operation, and the customer need not account for the impact of a CPU-intensive operation like deduplication on the production server. This keeps the production servers stable while gaining the benefit of deduplication on the backup storage. However, target deduplication does not optimize the network throughput. The network transfer is not deduplication-optimized either between the primary location and the DPM server or between the DPM server and Azure Backup. The data transfer with Azure Backup is optimized to some extent by the compression that is applied to the data before it is sent to Azure. This helps in sending data over bandwidth-constrained Internet pipes, but doesn't achieve the level of savings that can be achieved with WAN optimizers.

**TIP** **DPM customers sending data over the WAN between the primary location and the DPM server can use a WAN optimizer to reduce network bandwidth costs.**

# Using deduplication with DPM

DPM is designed to work with Windows Server Deduplication in a very specific configuration. More details about the setup and instructions for enabling deduplication are available in the whitepaper at *https://technet.microsoft.com/en-us/library/dn891438.aspx*. This section focuses on the high-level architecture and the reasons for some of the constraints on the deployment.



www.dbooks.org

## High-level deployment architecture and constraints

In a deduplication-enabled DPM deployment, the DPM VHDs must be placed on a separate file server system volume—typically a storage volume for the shared folder. Figure 8-2 shows a backup deployment with deduplication enabled.

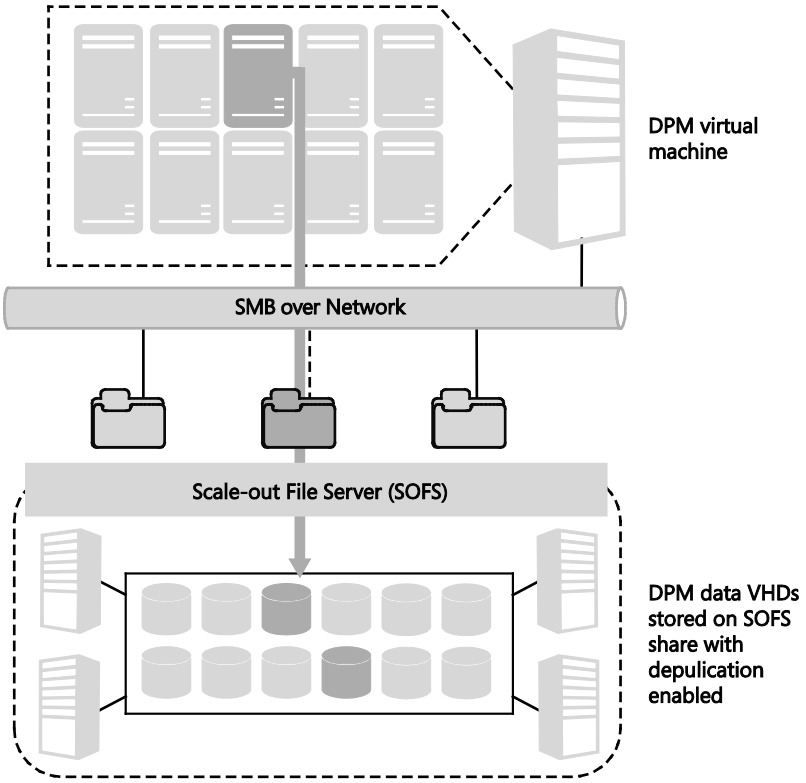



**FIGURE 8-2** Architecture diagram showing the key elements of a backup deployment with deduplication enabled

## Understanding the backup-deduplication software stack

Windows Server Deduplication processes data as files. Hyper-V translates the data processing performed by DPM into file changes—that is, in-guest writes become VHD writes, which then become file writes. These VHD files are then available for deduplication processing. This layered separation of VHD I/O operations and deduplication I/O operations are primarily about ensuring that the CPU and I/O used by deduplication does not cause problems with the performance of the workload, which in this case is DPM. There are two main restrictions—the SOFS volume size and the file size—for which the details can be found at *http://blogs.technet.com/b/filecab/archive/2014/12/04/sizing-volumes-for-data-deduplication-in-windows-server.aspx* .

It is easy to scale up the SOFS compute if needed without impacting any other part of your infrastructure. By leveraging Windows Server Deduplication, customers are using a capability that they have already paid for with Windows Server licenses. The deduplication capability is essentially available for free.

The deduplication technology used is variable-chunking with potentially large chunk sizes. This allows the deduplication to yield excellent results when removing duplicates. The chunk store is contained within the SOFS volume that has been enabled for deduplication, so the deduplication benefits are restricted to the data in the volume. This has implications on where the DPM VHDs are placed and how much overall savings can be achieved on the backup data.

The TechNet whitepaper referenced previously includes other nuances as well:

- Ensuring that backup and deduplication operations don't overlap
- Provisioning IOPS for both backup and deduplication operations, especially if they will overlap
- Setting up the deduplication job
- Software version prerequisites

## Deduplication benefits: A real-world scenario

With the release of for DPM 2012 R2 Update Rollup 4 (UR4) and Windows Server 2012 R2 with the November 2014 update rollup, customers could start using deduplication with their DPM setup. One such customer is a leading hosting provider in Europe and the United States, providing hosted solutions based on Microsoft Dynamics.

The production environment of this hosting provider consisted of a Hyper-V compute cluster with 61 virtual machines using a total of 3 TB of production storage. The workload running in these virtual machines was a combination of Microsoft Dynamics and Microsoft SQL Server.

Backup was done using DPM 2012 R2 with mandatory update rollup bits installed. The backup of the virtual machines was done daily, with a backup window between the off-peak hours of 10 PM and 6 AM. In compliance with the recommended architecture, DPM was installed in a virtual machine, and the DPM storage was placed on an SOFS volume with deduplication enabled. The deduplication jobs then ran on the SOFS volume after 6 AM when the backup jobs had finished.

Table 8-1 shows the deduplication results in this real-world scenario.

**TABLE 8-1** Backup storage consumption before and after enabling deduplication

| | |
|---|---|
| Backup storage before deduplication | 4.77 TB |
| Backup storage after deduplication | 1.21 TB |
| Deduplication savings | 74% |



www.dbooks.org

CHAPTER 9

# Integration with System Center

Each new release of Microsoft System Center Data Protection Manager (DPM) has extended the application's capabilities in core backup scenarios as well as in the management plane. Since DPM 2012, administrators have been able to manage multiple DPM servers under a single console known as the DPM Central Console instead of using Remote Desktop Protocol (RDP) to remotely connect to each of their DPM servers. By using the Central Console, administrators can examine alerts consolidated from DPM servers, take action on alerts as suggested by the console, resume backups, and perform other tasks, all without needing to separately connect to each DPM server. This chapter covers DPM monitoring and management scenarios, including reporting, alerting, and automatic client protection using other System Center products such as Microsoft System Center Operations Manager and Microsoft System Center Virtual Machine Manager.

## Management and monitoring scenarios and challenges

Most administrators who deploy DPM have more than one DPM server in their production environments. They need a centralized management and monitoring solution (for example, for reporting, monitoring, and alerting) that provides a single view across all DPM servers. Consider the importance of consolidated alerting for example. If there is a network outage, failed backups can inundate the management console with alerts, and addressing these alerts can be difficult to manage. Consolidated alerts in the management console for repetitions and for alerts with the same root cause are useful in this situation. Service level agreement (SLA)-based alerting, where alerts are generated when an SLA is broken, are also useful. To resolve all of these alerts, most administrators prefer to take action from a single console.

Consolidated reporting is also helpful to administrators. Prior to DPM 2012 R2 UR5, there were only six standard reports that administrators could review per DPM server. But most organizations prefer an aggregated view across DPM servers and the flexibility to create their own reports since one size does not fit all. For example, the CIO of a service provider might be interested in knowing if the company is able to meet the SLA for customers. In addition, the CIO might prefer particular colors and a type of chart for reports.

## Enterprise reporting capabilities

The new, enhanced backup reporting framework offers aggregation, scalability, and customizability, and enables centralized report generation from a single pane in Operations Manager. Pre-defined views are created and documented in the Operations Manager data warehouse, which enables the Microsoft partner community to build value on top of the reporting framework. This framework was created with four design points in mind:

- **Aggregation** The Reporting management pack enables and supports collecting reporting data from all DPM servers deployed on a network. For example, if a domain deployment contains 10 DPM servers, each containing 300 data sources configured into a separate Protection Group (PG), the Reporting management pack enables Operations Manager to monitor and report data on all 3,000 data sources (10 servers x 300 data sources).
- **Scalability** The Reporting management pack enables Operations Manager to scale up to any number of DPM sources and any number of data sources. For example, if a deployment environment scales up from 10 DPM servers with 300 data sources on each server to 20 DPM servers as the organization grows, there is no need to gather additional reports from the new DPM servers. The Reporting management pack enables Operations Manager to fetch data from all DPM servers in the environment automatically and to collate and store all of the data in the data warehouse.
- **Customizability** The Reporting management pack exposes several Operations Manager data warehouse views. These views can be used to query the Operations Manager data warehouse and produce extensive reports using any framework, scripting, or programming language of your choice.
- **Flexibility** Customers have the flexibility to choose the reporting tool of their choice, such as SQL Server Reporting Services (SSRS) or Crystal Reports.

Figure 9-1 shows a sample Executive Summary report displayed in the Operations Manager Reporting section. This example shows the power of the reporting platform. Notice the deduplication savings here even though deduplication is enabled on file servers and DPM running on a virtual machine has no knowledge of it. Management packs contain the Reporting pack and the File Server Deduplication pack to enable this report.



www.dbooks.org

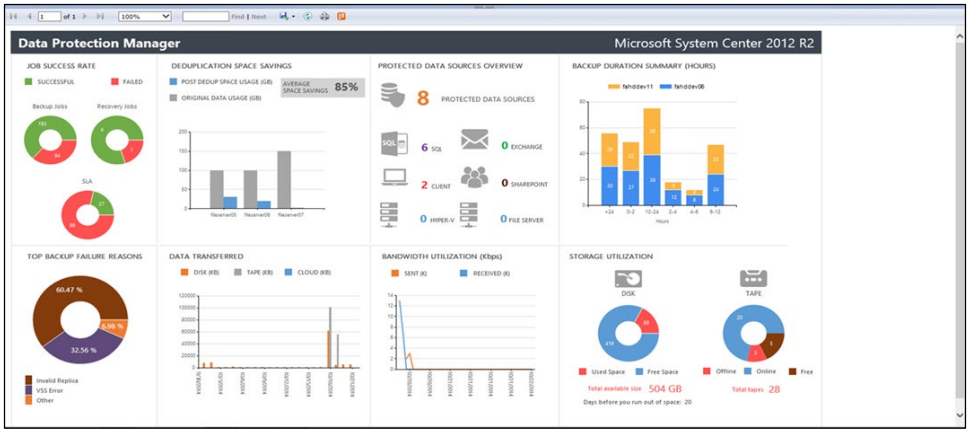



**FIGURE 9-1** CIO Report

**IMPORTANT** **To use the enterprise reporting capabilities, the server must be running at least DPM 2012 R2 UR5. In addition, Operations Manager 2012 R2 is required with the data warehouse up and running.**

***See also*** *For more information about the salient features of the recently released DPM Management Reporting, Discovery, and Monitoring management packs, see http://blogs.technet.com/b/dpm/archive/2015/04/16/announcing-centralized-and-customizable-backup-reports-using-data-protection-manager.aspx.*

# Management and monitoring solutions

As mentioned previously, the DPM Central Console is actually an Operations Manager console that allows an administrator to combine the health monitoring capabilities of Operations Manager with DPM administrative tasks and to administer multiple DPM servers from a single location. This is accomplished by means of a set of views and tasks in the Operations Manager console that provide most of the functionality that is available in the DPM Administrator Console. The reason for saying "most of the functionality" is that the Central Console shows only the objects for which the alert is generated as opposed to showing all objects.

**NOTE** **The Central Console cannot be installed on DPM servers; only the Operations Manager agent should be installed on DPM servers. The Central Console should be installed on the Operations Manager server.**

***See also*** *For more information about Central Console features and the various tasks that can be performed with the Central Console, see https://technet.microsoft.com/en-us/library/jj860391.aspx.*

To install the Central Console from DPM 2012 R2 UR5, you first need to import the following four DPM management packs into your Operations Manager server:

- Microsoft.SystemCenter.DataProtectionManager.2012.Discovery.mp
- Microsoft.SystemCenter.DataProtectionManager.2012.Library.mp
- Microsoft.SystemCenter.DataProtectionManager.2012.Reporting.mp
- Microsoft.SystemCenter.DataProtectionManager.DedupReporter.mp

After importing the required management packs, launch the Setup screen of Operations Manager and select both Install Central Console Server and Client Side Components To Monitor DPM Servers, and then use the scoped DPM Administrator Console.

> ***See also*** *For more details about Central Console installation instructions, see https://technet.microsoft.com/en-us/library/hh758189.aspx.*

## SLA-based alerts

One new feature introduced in DPM 2012 UR5 is alerts in DPM and the Operations Manager console when SLAs are not met (for example, at least one backup a day for all data sources). For any protection group that misses the configured backup SLA, DPM raises an alert that is visible in Operations Manager as well as the DPM server. You can configure SLA requirements on your DPM server by using the following Windows PowerShell commands:

```
Set-DPMProtectionGroupSLA –ProtectionGroup $pg –SLAInHours $sla
Set-DPMProtectionGroupSLA –ProtectionGroupId $pgId –SLAInHours $sla
```

> **NOTE** **Valid values for SLAInHours are 2, 4, 6, 8, 12, 24, 48, 72, and so on (any number divisible by 24 or a multiple of 24 up to 90 days). The most common number is 24 hours since most customers prefer a daily backup SLA.**

To accomplish this, DPM checks nightly for all the protection groups that have the SLA configured. If it determines that a protection group missed the SLA, it raises an alert. For example, if the administrator configured a backup SLA of 24 hours for a protection group with two backups per day and no recovery point was created in the past 24 hours, a missed SLA job would run and display an SLA missed alert for the protection group, as shown in Figure 9-2.



www.dbooks.org

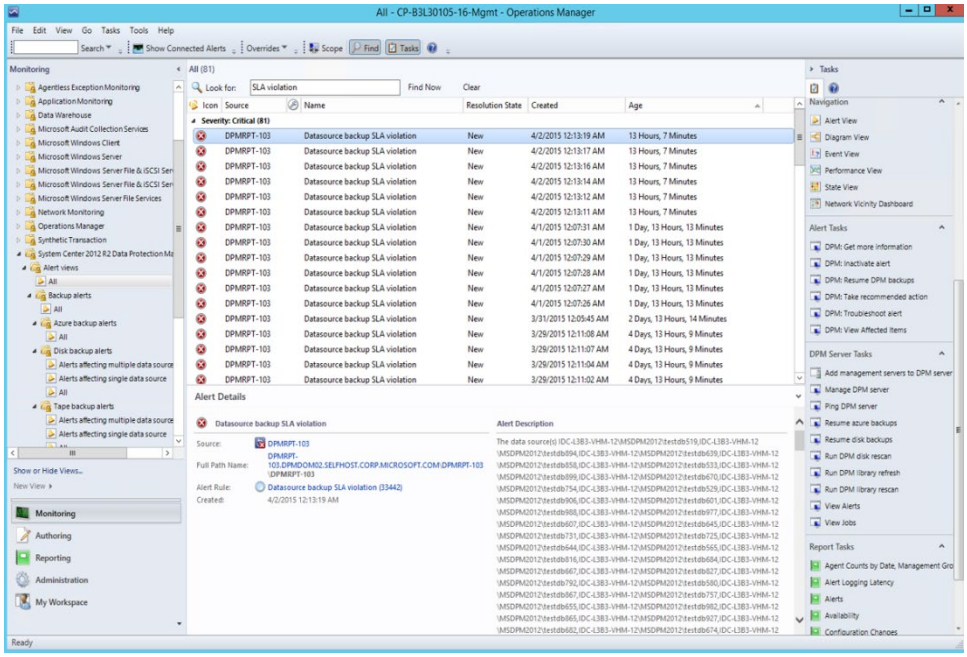



**FIGURE 9-2** Operations Manager console showing SLA violations with information on the DPM server name, alert description, creation date, and so on

These alerts cannot be auto-resolved because if a data source has missed its SLA, you cannot travel back in time to create a recovery point for it. The only way to resolve an alert like this is to manually inactivate the alert by one of the following methods:

- Click DPM: Inactivate Alert under Alert Tasks in the Tasks pane as shown on the right side of Figure 9-2.
- Click the Inactivate button on the toolbar as shown in Figure 9-3.
- Click the Inactivate link in the Resolution section of the Details for the alert as shown in Figure 9-3.

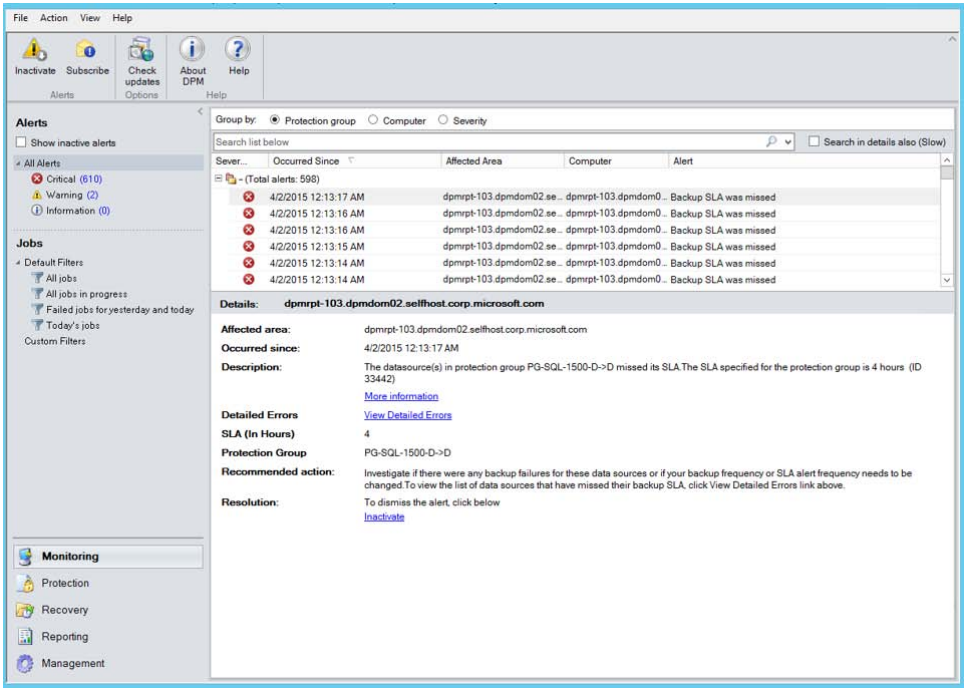



**FIGURE 9-3** Scoped DPM console showing SLA violations with information on the alert description, PG name, SLA, and the recommended action

You can also remove SLA alerts by running the following Windows PowerShell command for each protection group:

```
Set-DPMProtectionGroupSLA -ProtectionGroup $pg -SLAInHours
```

**IMPORTANT** SLA miss alerts will not be resolved automatically on the next successful recovery point. Administrators must manually inactivate the alerts.

## Client Auto Deployment

In large organizations with many desktop and laptop computers, the process of deploying DPM agents can be cumbersome. To automatically deploy the DPM protection agent on laptop and desktop computers, administrators can use the Client Auto Deployment management pack (ClientAD.msi). In an environment where Operations Manager is being used, ClientAD.msi is installed on the Operations Manager server and the Operations Manager agents are deployed on the DPM servers. As soon as a new laptop or desktop computer joins the network, Active Directory Domain Services notifies System Center Configuration Manager, which then pushes the agents to the client computers.



www.dbooks.org

ClientAD.msi ships with the DPM installation media, but you should always obtain and use the latest version. In addition to providing the elements for the Central Console, ClientAD.msi discovers all DPM servers and protected resources in the environment and measures their health. Complete the following steps to auto protect client machines:

1. Download and install the latest version of ClientAD.msi from *http://go.microsoft.com/fwlink/?LinkId=207880* and install it on an Operations Manager server.
2. Install the Operations Manager monitoring agents on the DPM servers that protect the client computers. This ensures that the DPM servers are discoverable by the Operations Manager. Operations Manager runs discovery once every 24 hours, but this property is configurable.
3. Select Administrator, then Agent Managed, then Discovery Wizard, and then Discover DPM Server And Install.
4. Import the Client Auto Deployment management pack into the Operations Manager server.
5. Add the domain name in the configuration file DomainsForAutoDeployment.xml in the following folder:

   C:\Program Files\Microsoft Data Protection Manager\Auto Deployment\Config
6. Add the exclusion list of DPM servers in exclusions.txt located in the same folder.
7. Select Monitoring, then DPM Client Auto Deployment, and then Auto Deployment DPM Server State. You will see all the DPM servers on which Operations Manager monitoring agent is installed.
8. Select a DPM server and include it for auto client deployment.
9. On the Monitoring tab in Operations Manager, select DPM Client Auto Deployment and then Auto Deployment DPM Server State.
10. Select the DPM server, and under DPM Server Task, select Include For Auto Deployment.
11. Right-click the DPM server to check the properties and ensure Is DPM Included is set to True. Available Capacity is set to 1,000 if no client is protected from the DPM server. For example, if 200 clients are already protected by this DPM server, available capacity should be set to 800 (1,000 – 200).

After Operations Manager finds at least one DPM server, it creates a list of all the client computers in the protected domains on DPM servers that have auto-protection enabled. Operations Manager then compares the list of computers against the list of protected and excluded computers and creates a list of the computers that need protection but aren't part of a protection group. Using this list, Operations Manager creates protection groups, assigning a maximum of 100 laptop computers per DPM server per day. When a computer connects to the network, Configuration Manager installs the DPM agent on the computer. After the computer has been added to a protection group, DPM triggers backups based on the protection group settings.

**NOTE** **The current version of the Client Auto Deployment management pack is in Preview and supports only Windows XP, Windows Vista, and Windows 7 clients. Support for Windows 8, Windows 8.1, and Windows 10 clients will be announced in a later release of DPM.**

***See also*** *More details concerning the Client Auto Deployment management pack can be found at https://technet.microsoft.com/en-us/library/hh757976.aspx.*



www.dbooks.org

CHAPTER 10

# Integration with Azure Backup

As enterprises look to offload their infrastructure to the public cloud and reduce capital expenditure (CAPEX), backup is one of the workloads that has the least friction moving to cloud. Organizations use traditional backup software and leverage on-premises, disk-based storage for their short-term backup retention. In addition to the disk-based storage, organizations use tape for long-term retention. The overall storage CAPEX for organizations increases as data grows, and administrators are looking for lower cost alternatives without requiring upfront investment in storage for backup purposes. Azure Backup is a good choice when it comes to maintaining backups for organizations large or small. Azure Backup provides an easy-to-use interface to back up Windows-based servers from on-premises to Azure. By integrating System Center Data Protection Manager 2012 R2 (DPM) with Azure Backup, large organizations can embrace Azure Backup as a viable, long-term retention target. This chapter describes how Azure Backup and DPM provide a compelling, hybrid cloud backup solution for organizations.

## Advantages of Azure Backup

Azure Backup makes a great case for moving on-premises tape and disk infrastructure to the cloud. As with all cloud solutions, it is cost effective, with a pay-as-you-go model and no upfront costs. But unlike other cloud-connect strategies, Azure Backup is built as a cloud-first software as a service.

This model has several advantages. The service comes with 99.9 percent availability time. As users create a backup vault to store data, the data is stored in geo-replicated storage, protecting it from disasters. Even if there is an outage of one of the Azure datacenters, the data is accessible.

But it is not sufficient for the data to be geo-redundant; the service that enables access to data should also be geo-redundant. Azure Backup is available in two or more regions per geography and has a built-in business continuity plan so that even when the primary Azure datacenter experiences an outage, the service fails over to a new datacenter. Therefore, regardless of whether the organization loses on-premises data or whether the Azure datacenter has an outage, both the data and the backup service are available for customers to

retrieve their data. If Azure fails over to a secondary data center, customers are able to browse all the recovery points associated with backup, pick any recovery point, and perform a restore, as well as continue backing up data to the service post failover.

With Azure Backup, backed up data is always encrypted on both the wire and at rest on Azure such that it is always secure before it leaves the on-premises datacenter. The Azure Backup service also maintains backup metadata that enables customers to restore data anywhere from Azure to an alternate Windows-based or DPM server.

> ***See also*** *For more information about Azure Backup and a list of frequently asked questions, see http://azure.microsoft.com/en-us/documentation/articles/backup-azure-backup-faq/.*

# Backup scenarios

This section examines some of the different backup scenarios that can be implemented using Azure Backup.

## Tape replacement

Many organizations store their backup data on-premises on disk media and store their long-term retention data on tapes. They invest in significant tape infrastructure to meet their compliance requirement. Besides the cost of tape infrastructure, tapes require manual intervention to replace the older ones. Tapes must be labelled correctly and risk potential errors and data loss if they are mishandled. To store tape data offsite, organizations must arrange tape pick-up on a daily or weekly basis. In addition, to recover offsite data, organizations must request all the relevant tapes and restore the data.

Cloud-based backup inherently addresses all of the preceding issues. As long as organizations have network connectivity to the cloud provider, backup to the cloud saves money. Organizations can leverage the pay-as-you-go model with the cloud rather than investing in upfront costs for tape storage.

Beyond the cost savings, there are other inherent advantages to storing backup data in the cloud:

- Data can be retrieved even if there is a disaster on-premises.
- Restore times are cut down and there is no need to wait for the tape delivery from offsite.
- There is no need to restore all data to retrieve a single item.

Azure Backup is clearly a good solution to address the tape replacement scenario because it provides a competitive tape replacement strategy for businesses.



www.dbooks.org

## Branch office backup

Branch offices typically have fewer machines and smaller infrastructure than a large datacenter. However, the data generated in branch offices is often critical for the business. Some organizations back up this data locally in the branch office, which means they need to purchase additional storage for each branch in addition to managing the complexity of the storage and backup infrastructure in each branch.

More than a handful of branch offices increases the management complexity multi-fold. In this case, some organizations back up their branch office data to the main office. But this again means that the main office must purchase the storage necessary to support all the workloads that are backed up from each of the branches.

With cloud-based backup, however, organizations can eliminate their local storage and back up data directly to the cloud. Azure Backup enables businesses to back up their Windows-based servers directly to the cloud, thereby eliminating local storage at each of the branch offices.

## Windows client backup

With Azure Backup, organizations can back up files and folders on their Windows-based desktops and laptop computers directly to the cloud with an entirely self-service model where the IT administrator needs to take minimal or no action on behalf of the user. Since the data is encrypted when it leaves the computer, data is always secure. Individuals in a small organization can either share the same vault or each user can have a dedicated vault or subscription, depending on the sharing needs among the individuals in the organization.

## Protection of Microsoft Azure assets

With an ever-increasing number of enterprises and small businesses moving their workloads to the cloud, organizations need a simple mechanism to ensure that data created in the cloud is also backed up, just as is done on-premises. Microsoft Azure inherently provides high availability and redundancy of storage with the guarantee that if there is a storage or computer outage, the application can continue to run using redundant storage or computers. With the support for backup of Azure IaaS virtual machines (VMs), however, (which is currently in preview at the time of this writing) organizations can get the benefit of additional protection since they can protect their workload data from software corruption or data loss scenarios as well. In addition, organizations can always test their backups by performing a restore of data periodically.

> ***See also*** *For more information on the support for backup of Azure IaaS VMs, see http://azure.microsoft.com/blog/2015/03/26/azure-backup-announcing-support-for-backup-of-azure-iaas-vms/.*

## Getting started with Azure Backup

To configure your on-premises backup to Azure, all you need is an Azure subscription. With a subscription, you can sign in to the Azure Portal at *https://manage.windowsazure.com* and download the agent and the vault credentials from Azure. The downloaded Microsoft Azure Recovery Services agent then needs to be installed on the Windows-based Server or the DPM server that is sending data to Azure.

After downloading the agent and vault credentials, perform the following steps to complete protection configuration:

1. Install Azure Recovery Services agents on the DPM server.
2. Click Register at the top left corner of the screen to launch the Register Server Wizard (see Figure 10-1), which can be used to register the DPM server to a specific vault using the vault credentials.

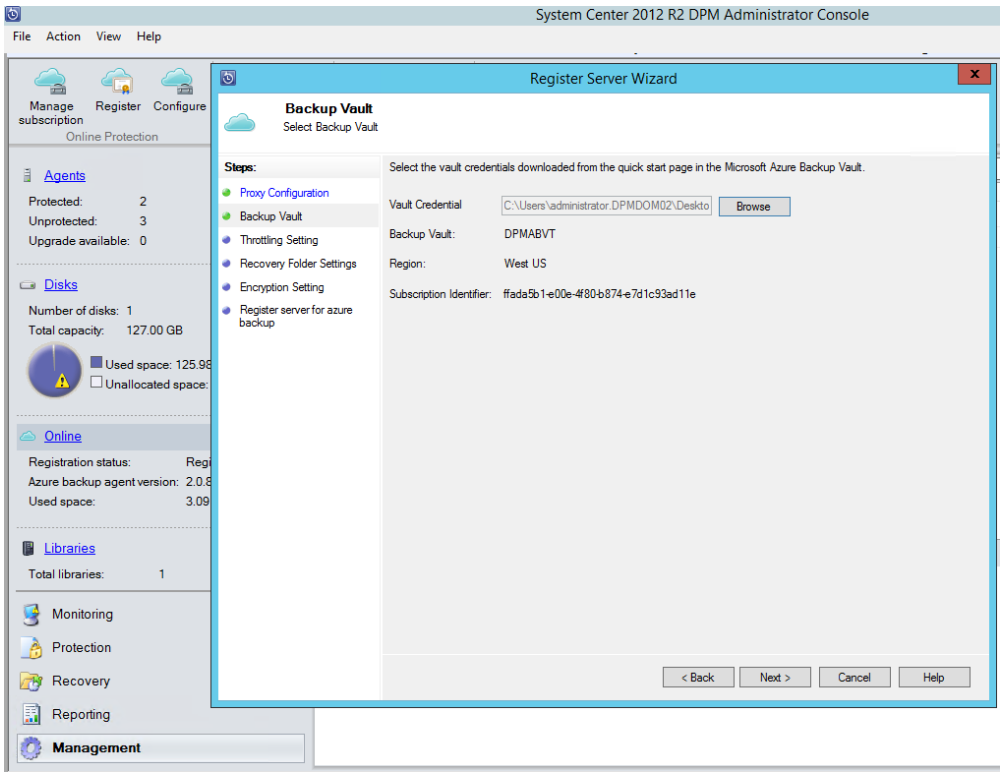



**FIGURE 10-1** The Register Server Wizard

3. Specify the bandwidth throttling settings, recovery staging area, and encryption settings.



www.dbooks.org

The recovery staging area is used as a scratch space area for data to be compressed and encrypted before it is sent to Azure. This staging area is also used to download the backup recovery points from Azure before restoring it to the target production server.

The encryption passphrase is something that only the administrator has access to, and it is used by a Windows-based or DPM server to encrypt data before it is sent to Azure. Azure Backup service does not store or use this information in any way.

# Azure Backup capabilities

The latest service releases on Azure Backup and DPM offer three key capabilities:

- Expanded workload support, including support for Microsoft Exchange, Microsoft SharePoint, and Windows client operating systems
- Support for long-term retention of backup with rich retention policies
- Support for offline seeding of initial replica to Azure

## Expanded workload support

Just as Microsoft SQL Server Hyper-V VMs and file/folders have been supported for backup to the cloud, now Exchange servers, SharePoint servers, and Windows client systems are eligible for backup from on-premises environments to Microsoft Azure.

The backup and restore experiences for these scenarios are consistent with backup and restore of data to on-premises disk media with just one exception: for SharePoint farms, the granularity of restore is a content database. This is unlike restore from disk replica where customers can restore individual items from their SharePoint content databases. Cloud replica can be used if there is a disaster at the primary datacenter or if the local disk replica copy is broken. In this case, after the content database is downloaded, it can be attached to an existing on-premises SharePoint farm to allow retrieval of the items.

## Long-term retention

As mentioned previously, you can use Azure Backup as a tape replacement strategy for your organization. Integral to this strategy is the ability to store data for many years. With Azure Backup, you can technically store data for up to 99 years, but in practice, most businesses need a solution that retains data for about 7 to 10 years. You can choose the backup frequency to be daily, weekly, monthly, or yearly, and the retention policy is based on the backup frequency. Figure 10-2 shows how you can configure the backup frequency from DPM to Azure Backup, while Figure 10-3 illustrates the retention policy for Azure Backup.

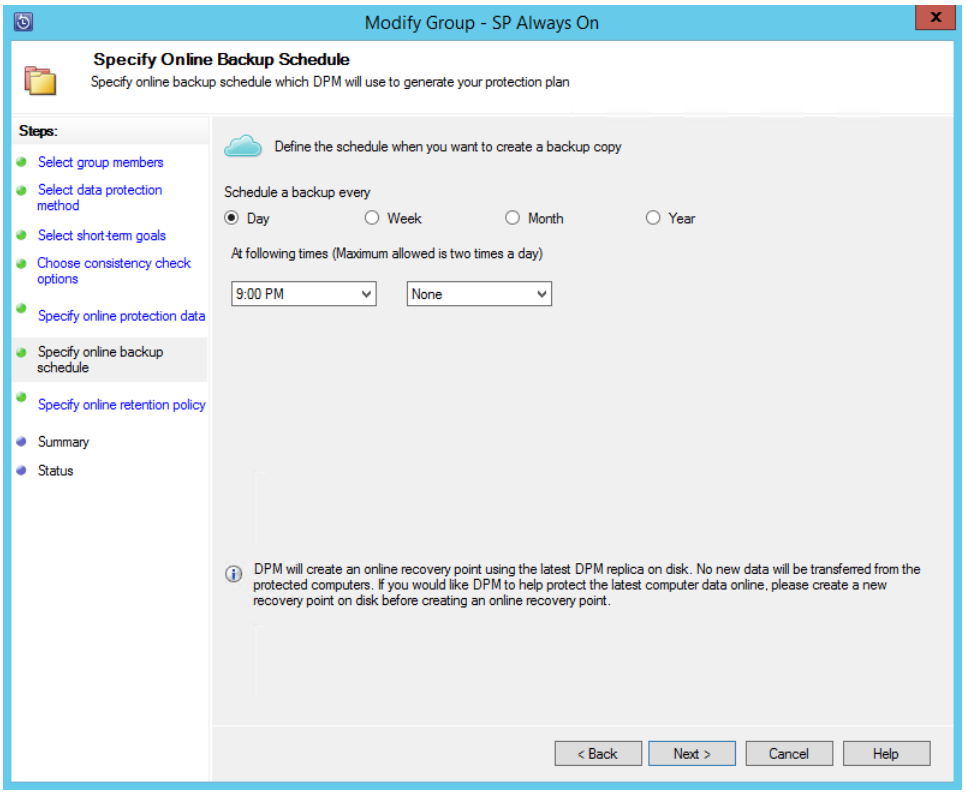



**FIGURE 10-2** Configuring the backup frequency from DPM to Azure Backup



www.dbooks.org

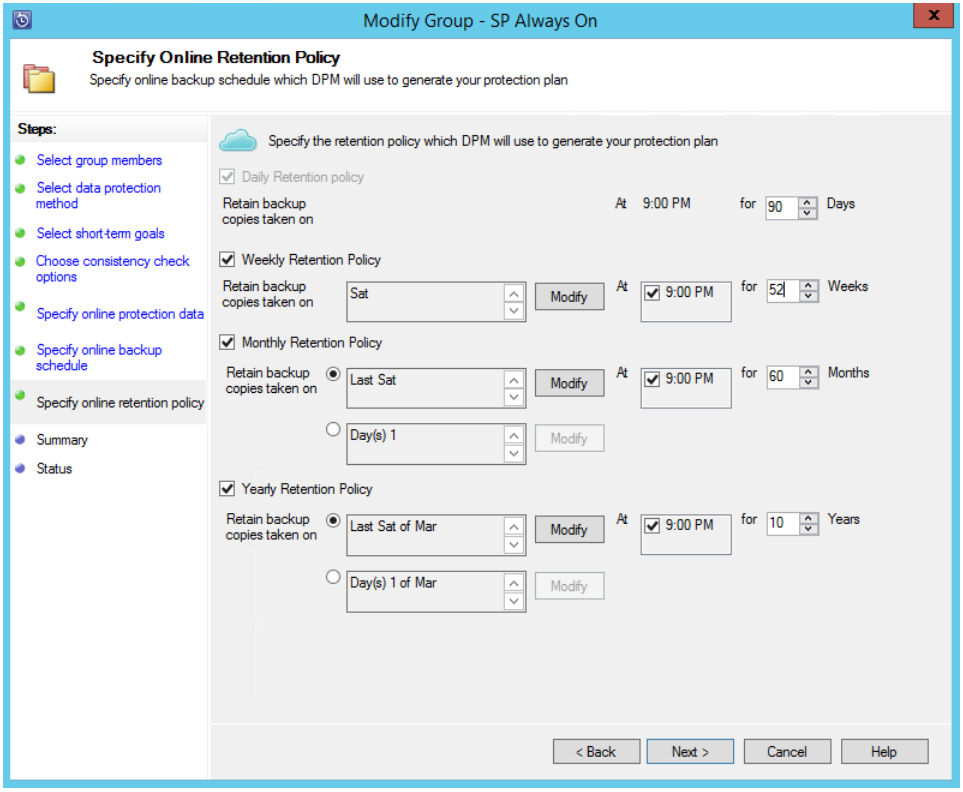



**FIGURE 10-3** Configuring a retention policy for Azure Backup

***See also*** *For detailed information about the steps for configuring Azure Backup long-term retention policies. see http://azure.microsoft.com/en-gb/documentation/articles/backup-azure-backup-cloud-as-tape/.*

## Offline seeding of initial replica

If you cannot use your network for sending an initial copy of data to Azure Backup, you can use the Azure Import service instead. With this approach, you need to specify the offline backup method when creating a new protection group in DPM. You also need to specify your Microsoft Azure subscription, import job name, storage account, and the container where you will place the data that will be received by Microsoft Azure. When the backup starts, the data is copied to the set of specified removable SATA disks, and when the backup is complete, your disks can be shipped to Microsoft Azure. An Azure import job is then created and tracked by the Azure Backup client for the arrival of contents to Azure. When the data arrives in Microsoft Azure, the initial backup process completes by copying the data to the Azure backup vault.

> ***See also*** *For more details concerning the export and import steps, see http://azure.microsoft.com/en-gb/documentation/articles/backup-azure-backup-import-export/.*



www.dbooks.org

CHAPTER 11

# Protecting Azure IaaS virtual machines

Organizations build business continuity procedures to encompass all critical infrastructure. Backup is one such business continuity procedure, and organizations can choose from an extensive set of vendors who can cater to their on-premises backup needs. However, as organizations move more of their infrastructure pieces into the public cloud and treat the public cloud as an extension of their on-premises infrastructure, backup procedures are extended to include public cloud elements as well. Microsoft Azure is no different.

Chapter 4, "Protecting Azure IaaS workloads," covered how to back up workloads running in Azure by using Microsoft System Center Data Protection Manager (DPM) running in an Azure virtual machine (VM). This chapter covers the backup of the infrastructure element: the Azure VM itself. The chapter starts by exploring the motivations that drive VM backup in Azure. It then describes the differences between backup of VMs on-premises and in Azure and how VM backup actually works in Azure. Finally, the chapter examines some considerations for planning for VM backup.

## Why back up Azure VMs?

The reasons for backing up VMs on-premises and in Azure are similar. Different customer segments derive different value from the solution, but here is a list of commonly encountered reasons for backup of VMs:

- **Reduced restore time during disasters** The VM contains all the information needed to get the application up and running. This includes the operating system, the application software, the configuration settings, and the data. It would typically take more time to piece together all of these at the time of restore, so restoring the complete VM is much faster.
- **Greater suitability for long-term retention** The VM includes the user data and the associated application software to work with it. Thus, the restores done from very old backup points are more likely to achieve any data retrieval goals.
- **Easier management** The VM acts as an encapsulating entity for the data. Rather than manage tens to hundreds of smaller data entities, the VM makes it simpler for the backup administrator. This makes patch rollback scenarios easy to address.

## Tradeoffs with VM backup

There are several tradeoffs with backup of Azure VMs, and the primary one is granularity. Restore of more granular data sources, like a Microsoft SQL Server database, is more challenging with VM backup. The granularity of restore is typically that of the full VM, although in some cases the granularity could be a virtual disk. This impacts the restore time and user experience in item-level recovery scenarios.

Another tradeoff concerns the backup schedule and frequency of backup. A good example of this is SQL Server running on an Azure VM. The SQL Server backup frequency can be as low as 15 minutes, while VM backup is typically once a day. Attempting VM backup every 15 minutes would have an adverse impact on the SQL Server workload.

Because of the above considerations, it's best to have a backup strategy that is a combination of granular workload-level backup and Azure VM backup.

## Azure Backup vs. on-premises backup

There are similarities in the backup requirements for both Azure and on-premises VMs:

- **Application consistency** Windows workloads use the Volume Snapshot Service (VSS) to ensure that data is written to the disk correctly before the backup is taken. Since VSS is absent in Linux, only file-system consistency can be expected from Linux VMs.
- **No shutdown of running workloads** Production workloads need to have a high uptime, and to shut down the workload or VMs for backup is not a realistic option.

The fundamental difference between VM backup in Azure and on-premises is the access to the host fabric. With on-premises backup, the backup agent is deployed on the host and the backup of the VMs happens with the involvement of the hypervisor installed on the host.

With Azure VM backup, there is no access to the host fabric or the hypervisor of the Azure cloud. However, Azure does provide APIs to access and manipulate individual objects that are virtualized. For example, the Azure Storage APIs provide extensive capabilities around blob creation, blob deletion, reading and writing data, and blob snapshots. Similar APIs are provided for VM operations. Instead of relying on the hypervisor, the backup solution needs to rely on these public APIs to facilitate the backup of Azure VMs, along with ensuring consistency and uptime of the workload.



www.dbooks.org

## How VM backup in Azure works

The internal process by which VM backup in Azure takes place consists of the following steps:

1. Freeze disk writes from the workload using VSS within the guest.
2. Use the Azure Storage APIs to take blob snapshots of the virtual disks attached to the VM.
3. Complete the VSS process and release pending writes from the workload.
4. Copy the data from the blob snapshot (data transfer).

Figure 11-1 illustrates the interaction between the Azure Backup service and the Azure VM during backup. The sections that follow go into further details concerning the backup process.

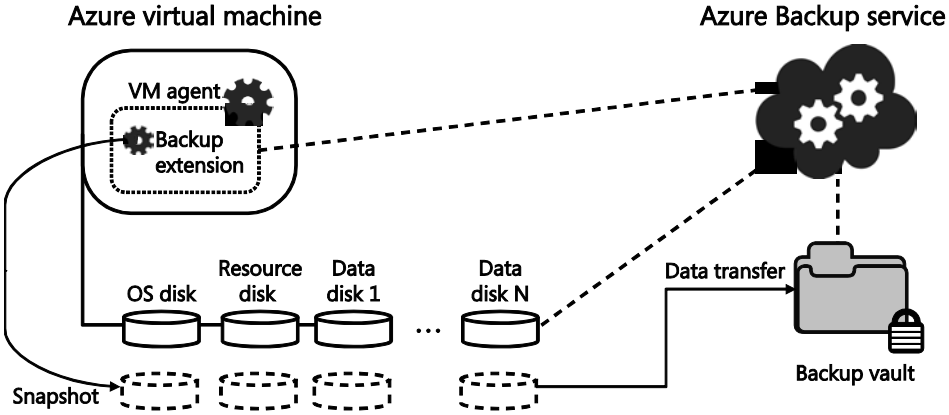



**FIGURE 11-1** Architecture diagram showing the interaction between the Azure Backup service and the Azure VM during backup

## The backup extension

Steps 1 and 2 in the backup process work on two different layers of the software stack. Specifically, step 1 is within the guest and step 2 is outside the guest, dealing with the Azure Storage APIs. Coordination is needed between these two steps to take the snapshot at the right moment, and this is done using the backup extension. The backup extension is software that plugs into the VM Agent infrastructure, and having the VM agent installed is a prerequisite for Azure VM backup to work correctly.

The Azure Backup service, which is currently in Preview at the time of this writing, is the controlling entity that triggers the backup based on the backup policy and schedule that has been set by the user. The backup trigger is essentially a command to the backup extension to start the VSS process inside the guest. When the data has been flushed to the disk and the remaining disk writes are frozen, the backup extension then immediately takes a snapshot of all the virtual disks associated with the VM using the Azure Storage blob APIs. After the

snapshots have been taken, the backup extension releases the pending disk writes and indicates to the Azure Backup service that the snapshotting job is done. Thus, the backup extension plays a key role in the overall backup workflow.

> ***See also*** *For more information on the backup extension, see http://azure.microsoft.com /blog/2014/04/11/vm-agent-and-extensions-part-1/. And for information about the VM Agent infrastructure, see https://msdn.microsoft.com/en-us/library/azure/dn832621.aspx.*

## Data transfer

With the snapshot taken, the remaining task is to copy the data away from the workload to a secure backup location. With Azure Backup, the data from the blob snapshot is copied to the Azure Backup vault. Only the data that has been changed since the last backup will be copied, thereby enabling incremental recovery points and ensuring that the backup data is stored as efficiently as possible.

The data can be transferred and stored anywhere, based on the user preference, for example:

- **A local copy in the same region or even the same storage account** Typically, this is achieved using the Copy Blob API and is done quickly. The downside to this approach is that the Copy Blob API creates full copies of the data, which can quickly add up on the monthly storage bill.
- **A local copy in the Azure Backup vault** This ensures that unlike the Copy Blob APIs, the data is stored as incremental recovery points when compared to the existing recovery points. This is more storage efficient and has a lower impact on the monthly storage bill.
- **A local copy and a remote copy using GRS** By ensuring that the backup data is replicated to an alternate datacenter, organizations can recover data in case of any major disaster that brings down the local Azure datacenter.
- **A remote copy in Azure, with or without the Azure Backup vault** This option is also known as cross-region backup where the user can effectively choose any region as the place to store the backup data. However, the downside to this approach is the egress costs incurred during backup.
- **A copy of the data in the user's on-premises datacenter** This option is similar to cross-region backup, but the second region is not an Azure datacenter. As with the other option, the downside in this approach is the egress cost incurred during backup. The major value of this option is that it insulates organizations against any major worldwide cloud outages.

> ***See also*** *For more information on the Copy Blob API, see https://msdn.microsoft.com /en-us/library/azure/dd894037.aspx.*



www.dbooks.org

## Learn more

There are other considerations to keep in mind while planning for backup of Azure VMs:

- Backup schedule
- Retention policy
- IOPS characteristics of the storage account in which the virtual hard disks are housed
- Backup of VMs on premium storage

To learn more about using Azure Backup for the backup of Azure VMs, read the online documentation that covers these aspects in much greater detail at *http://azure.microsoft.com/en-us/documentation/articles/backup-azure-vms*.

www.dbooks.org

# About the authors



**SHREESH DUBEY** is Principal Group Program Manager, Enterprise Backup, for the Cloud & Enterprise Group at Microsoft Corporation. He has 27 years of experience in the IT industry (at Intel and Microsoft) with diverse technology experience in datacenter infrastructure, storage/replication, data protection, JavaScript/AJAX, developer tools, embedded systems, and EDA tools. He holds two patents in storage replication and an MSEE from the University of Texas, Austin. He currently lives in Hyderabad with his wife and two daughters, an eleventh grader and a senior at the University of California, Berkeley. He enjoys downhill skiing, all outdoor activity, cooking, and insanely spicy food.



**VIJAY TANDRA SISTLA** is Lead Program Manager for the Cloud & Enterprise Group at Microsoft Corporation. Vijay has 16 years of combined experience in product development, quality, and customer support in cloud, business continuity, disaster recovery, and synchronization technology areas. He holds three patents in business continuity and SQL Server technologies, and holds a Master of Mechanical Engineering degree from the University of Texas, Arlington. He is also passionate about travel and food.



**SHIVAM GARG** is Principal Manager – Program Management, Enterprise Backup, for the Cloud & Enterprise Group at Microsoft Corporation. Shivam has 16 years of experience in the IT industry with diverse technology experience in the data protection, replication, search relevance, local and mobile search, and supply chain domains. He holds an MBA from IIM Bangalore and a bachelor's degree from IIT Kanpur, India. He currently lives in Hyderabad, known as the City of Pearls, with his wife, Jyoti, and two kids, a fifth grader and a kindergartner. He loves water sports (snorkeling, rafting, jet skiing) and reading spiritual books.



**AASHISH RAMDAS** is Program Manager II, Enterprise Backup, for the Cloud & Enterprise Group at Microsoft Corporation. He has 9 years of experience in the IT industry with a balance of business and engineering functions. He started in the telecommunications domain but now holds backup/replication quite dear. He received an MBA from the Indian School of Business, Hyderabad, and a BTech in CSE from NIT Trichy, India. He is passionate about statistics and data visualization, geopolitics, pricing and COGS optimization, spicy chicken biryani, and badminton—depending on the time of the day.

# About the series editor



**MITCH TULLOCH** is a well-known expert on Windows Server administration and cloud computing technologies. He has published hundreds of articles on a wide variety of technology sites and has written, contributed to or been series editor for over 50 books. Mitch is one of the most popular authors at Microsoft Press—the almost two dozen ebooks on Windows Server and System Center he either wrote or was Series Editor on have been downloaded more than 2.5 million times! For a complete list of free ebooks from Microsoft Press, visit the Microsoft Virtual Academy at *http://www.microsoftvirtualacademy.com/ebooks*.

Mitch has repeatedly received Microsoft's Most Valuable Professional (MVP) award for his outstanding contributions to supporting the global IT community. He is a ten-time MVP in the technology area of Windows Server Software Packaging, Deployment & Servicing. You can find his MVP Profile page at *http://mvp.microsoft.com/en-us/mvp/Mitch%20Tulloch-21182*.

Mitch is also Senior Editor of WServerNews, a weekly newsletter focused on system admin and security issues for the Windows Server platform. With almost 100,000 IT pro subscribers worldwide, WServerNews is the most popular Windows Server–focused newsletter in the world. Visit *http://www.wservernews.com* and subscribe to WServerNews today!

Mitch also runs an IT content development business based in Winnipeg, Canada, that produces white papers and other collateral for the business decision maker (BDM) and technical decision maker (TDM) audiences. His published content ranges from white papers about Microsoft cloud technologies to reviews of third-party products designed for the Windows Server platform. Before starting his own business in 1998, Mitch worked as a Microsoft Certified Trainer (MCT) for Productivity Point.

For more information about Mitch, visit his website at *http://www.mtit.com*. You can also follow Mitch on Twitter @mitchtulloch.

www.dbooks.org




From technical overviews to drilldowns on special topics, get *free* ebooks from Microsoft Press at:

**www.microsoftvirtualacademy.com/ebooks**

Download your free ebooks in PDF, EPUB, and/or Mobi for Kindle formats.

Look for other great resources at Microsoft Virtual Academy, where you can learn new skills and help advance your career with free Microsoft training delivered by experts.

Microsoft Press

# Now that you've read the book...

## Tell us what you think!

Was it useful?
Did it teach you what you wanted to learn?
Was there room for improvement?

**Let us know at http://aka.ms/tellpress**

Your feedback goes directly to the staff at Microsoft Press, and we read every one of your responses. Thanks in advance!



Microsoft

www.dbooks.org