

第 4 章

实现存储空间和重复数据删除 (Implementing Storage Spaces and Data Deduplication)

目录：

实现存储空间和重复数据删除	1
(Implementing Storage Spaces and Data Deduplication)	1
单元概述 (Module Overview)	1
实现存储空间 (Implementing Storage Spaces)	2
管理存储空间 (Managing Storage Spaces)	13
实验 A: 实现存储空间 (Implementing Storage Spaces)	23
实现重复数据删除 (Implementing Data Deduplication)	28
实验 B: 实施重复数据删除 (Implementing Data Deduplication)	45
单元复习和作业 (Module Review and Takeaways)	48

单元概述 (Module Overview)

Windows Server 2016 操作系统引入了许多存储技术和对现有存储技术的改进。您可以使用存储空间 (Storage Space)，Windows Server 2016 的一项功能，它根据从操作系统抽象物理存储的存储池和虚拟磁盘来虚拟化存储和配置存储。重复数据删除 (Data Deduplication) 是一项功能，您可以使用它来查找和删除重复数据，同时保持数据的完整性。本单元介绍如何在 Windows Server 存储体系结构中使用这两个新功能。

目标 (Objectives)

完成本单元后，您将能够：

- 在企业存储需求的上下文中描述和实现存储空间功能。
- 管理和维护存储空间。
- 描述和实现重复数据删除。

第 1 课 实现存储空间 (Implementing Storage Spaces)

在服务器上管理直接连接存储 (DAS) 对于管理员来说是一项繁琐的任务。为了克服这个问题,许多组织使用将磁盘组合在一起的存储区域网络 (SAN)。然而, SAN 是昂贵的,因为它们需要特殊的配置,有时是特殊的硬件。为了帮助克服这些存储问题,您可以使用存储空间将磁盘集中在一起。然后,存储空间作为可以跨越池中的多个物理磁盘的单个磁盘呈现给操作系统。本课介绍如何实现存储空间。

课程目标 (Lesson Objectives)

完成本课后,您将能够:

- 实现存储空间作为企业存储解决方案。
- 描述存储空间功能及其组件。
- 描述存储空间的功能,包括存储布局,驱动器分配 (drive allocation,) 和配置方案 (如精简配置, thin provisioning)。
- 描述 Windows Server 2016 中对存储空间功能的更改。
- 描述存储空间的常见使用场景,并权衡其优点和局限性。
- 比较存储空间与其他存储解决方案。

企业存储需求 (Enterprise storage needs)

在大多数组织中,有关存储需求的讨论可能很紧张。这通常是因为存储成本是许多信息技术 (IT) 预算的主要项目。尽管单个存储单元的成本逐渐降低,但组织生产的数据量仍在快速增长,因此存储的总体成本持续增长。

因此,许多组织正在调查存储解决方案,为现有解决方案提供成本效益的替代方案,而不会牺牲性能。存储规划期间组织的典型需求是如何降低提供基础架构即服务 (infrastructure as a service , IaaS) 存储服务的成本和工作量。在规划存储解决方案时,您需要评估存储解决方案的规模。如果您的存储解决方案不能很好的扩展,它的成本会更高。此外,您应考虑部署廉价的网络和存储环境。这是通过部署行业标准的服务器,网络和存储基础架构构建高度可用、可扩展的软件定义的存储 (software-defined storage)。

最后,在规划如何降低提供 IaaS 存储服务的成本时,应考虑使用分解的计算和存储部署。虽然许多融合计算/存储解决方案提供了更简单的管理功能,但他们还需要同时缩放两个组件。换句话说,您可能必须在扩展存储时添加与以前的硬件相同的计算能力。为了降低提供 IaaS 存储服务的成本,在规划存储解决方案时,应考虑独立管理和独立扩展。

虽然您的要求可能决定在存储规划期间要考虑的高级功能,但评估存储解决方案时的主要驱动因素通常是容量,性能,成本和弹性。尽管您可以分别对这些驱动因素中的每一个进行冗长的讨论,但您的存储解决方案需要采用平衡的存储部署方法。

在规划平衡存储部署方法以满足存储需求时,您需要根据成本来评估容量和性能要求。为了提高成本效率,您的存储环境应使用固态硬盘 (SSD) 用于高活动数据 (更高的性能), 硬盘驱动器 (HDD) 用于偶尔访问的数据 (成本更高的容量)。


• 在您的存储规划中,您应评估您的存储解决方案是否需要支持以下功能:

- 镜像/奇偶校验支持
- 数据条带
- 机箱感知
- 存储分层
- 存储复制
- 重复数据删除
- 数据加密
- 性能分析

如果您只部署 HDD, 您的预算约束将阻止您满足您的性能要求; 这是因为 HDD 提供更高的容量, 但是具有更低的性能。同样, 如果您只部署 SSD, 您的预算限制将阻止您满足您的容量需求; 这是因为 SSD 提供更高的性能, 但容量更低。因此, 您的平衡存储部署方法将很可能包括 HDD 和 SSD 的混合, 以适当的成本实现最佳的性能和容量。

包括在您的存储规划中, 您应该考虑您的存储解决方案是否需要支持大多数存储产品的常见功能, 例如

- 镜像/奇偶校验支持 (Mirror/parity support)
- 数据剥离 (Data stripping)
- 机箱感知 (Enclosure awareness)
- 存储分层 (Storage tiering)
- 存储复制 (Storage replication)
- 重复数据删除 (Data deduplication)
- 数据加密 (Data encryption)
- 性能分析 (Performance analysis)

 **注意:** 此列表仅用于提供建议, 并不是大多数存储产品的常见功能的详尽列表。您的组织的存储需求可能不同。

数据量的增长, 不断增加的存储成本以及确保数据的高可用性的需要是 IT 部门解决的难题。Windows Server 2016 提供了许多存储功能, 目的是解决存储管理的这些重要方面的挑战。

问题: 在规划企业存储战略时, 应考虑哪些因素?

问题: 您的组织使用哪些存储技术?

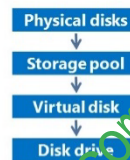
什么是存储空间 (What are Storage Spaces)?


存储空间 (Storage Spaces) 是 Windows Server 2016 和 Windows 10 中内置的存储虚拟化功能。

存储空间功能由两个组件组成:

- 存储池 (Storage pools)。存储池是聚合到单个逻辑磁盘中的物理磁盘的集合, 允许您将多个物理磁盘作为单个磁盘进行管理。您可以使用存储空间将任何类型和大小的物理磁盘添加到存储池。
- 存储空间 (Storage spaces)。存储空间是从存储池中的可用空间创建的虚拟磁盘。存储空间具有诸如弹性级别 (resiliency level), 存储层 (storage tiers), 固定配置 (fixed provisioning) 和精确管理控制 (precise administrative control) 的属性。存储空间的主要优点是您不再需要管理单个磁盘。相反, 您可以将它们作为一个单元进行管理。虚拟磁盘相当于 SAN 上的逻辑单元号 (Logical unit number, LUN)。

- 使用存储空间:
 - 将任何类型和大小的物理磁盘添加到存储池
 - 从池中创建高可用性虚拟磁盘
- 要创建虚拟磁盘, 您需要:
 - 一个或多个物理磁盘
 - 包含磁盘的存储池
 - 使用来自存储池的磁盘创建的虚拟磁盘 (或存储空间)
 - 基于虚拟驱动器的磁盘驱动器



 **注意:** 使用存储空间功能创建的虚拟磁盘与具有 .vhd 和 .vhdx 文件扩展名的虚拟硬盘文件不同。

要创建虚拟磁盘，您需要具备以下条件：

- 物理磁盘。物理磁盘是诸如串行高级技术附件（Serial Advanced Technology Attachment，SATA）或串行连接 SCSI（serial-attached SCSI，SAS）磁盘的磁盘。如果要将物理磁盘添加到存储池，磁盘必须符合以下要求：
 - 创建存储池需要一个物理磁盘。
 - 需要至少两个物理磁盘才能创建弹性的镜像虚拟磁盘。
 - 需要至少三个物理磁盘才能通过奇偶校验（parity）创建具有弹性的虚拟磁盘。
 - 三向镜像（three-way mirroring）需要至少五个物理磁盘。
 - 磁盘必须为空白且未格式化，这意味着磁盘上不会存在卷。
 - 磁盘可以使用各种总线接口连接，包括 SAS，SATA，SCSI，非易失性存储器快速（Non-Volatile Memory Express，NVMe）和通用串行总线（universal serial bus，USB）。如果计划对存储池使用故障转移群集（failover clustering），则不能使用 SATA，SCSI 或 USB 磁盘。
- 存储池（Storage pool）。存储池是可用于创建虚拟磁盘的一个或多个物理磁盘的集合。您可以将一个或多个可用的未格式化的物理磁盘添加到存储池，但您只能将物理磁盘连接到一个存储池。
- 虚拟磁盘或存储空间（Virtual disk or storage space）。从用户和应用程序的角度来看，这类似于物理磁盘。但是，虚拟磁盘更灵活，因为它们包括固定配置和精简配置（也称为即时（just-in-time，JIT）分配）。它们还通过内置功能（如镜像和奇偶校验）更有弹性地应对物理磁盘故障。这些类似于独立磁盘冗余阵列（RAID）技术，但存储空间存储数据的方式与 RAID 不同。
- 磁盘驱动器（Disk drive）。这是可以从 Windows 操作系统访问的卷，例如，使用驱动器盘符。



注意：在规划存储空间部署时，需要验证存储机柜（storage enclosure）是否已针对 Windows Server 2016 中的存储空间进行了验证。对于存储空间，要使存储空间按插槽识别磁盘并使用阵列的故障和标识/定位指示灯，阵列必须支持 SCSI Enclosure 服务（SES）版本 3。



附加阅读：有关详细信息，请参阅：“Windows Server 目录”，网址为：
<http://aka.ms/Rdpy8>

您可以使用 FAT32 文件系统，新技术文件系统（New Technology File System，NTFS）文件系统或复原文件系统（Resilient File System，ReFS）格式化存储空间虚拟磁盘。如果计划将存储空间用作群集共享卷（Clustered Shared Volume，CSV），重复数据删除或文件服务器资源管理器（File Server Resource Manager，FSRM）的一部分，则需要使用 NTFS 格式化虚拟磁盘。

存储空间的组件和功能 (Components and features of Storage Spaces)

配置存储空间时的一个重要步骤是规划虚拟磁盘。要配置存储空间以满足您的要求，在实施虚拟磁盘之前，必须考虑下表中描述的存储空间功能。

在您的环境中最佳地使用存储空间的功能

功能	选项
存储布局	简单 双向或三向镜像 校验
磁盘扇区大小	512 或 512e 4 KB
驱动器分配	数据存储 手动 热备用
配置方案	精简配置空间 固定配置空间
条带参数	列的个数 交错

功能	描述
存储布局 (Storage layout)	<p>存储布局是定义从存储池中分配磁盘数量的特性之一。有效选项包括：</p> <ul style="list-style-type: none"> 简单 (Simple)。一个简单的空间具有数据条带化，但没有冗余。在数据条带化中，逻辑顺序数据跨多个磁盘进行分段，以使不同的物理存储驱动器访问这些连续的段。条带化可以提高性能，因为可以同时访问多个数据段。要启用数据条带化，必须至少部署两个磁盘。简单的存储布局不提供任何冗余，因此如果存储池中的一个磁盘出现故障，您将丢失所有数据，除非您有备份。 双向和三向镜像 (Two-way and three-way mirrors)。镜像有助于防止丢失一个或多个磁盘。镜像空间维护它们所托管的数据的两个或三个副本。具体来说，双向镜像保持两个数据副本，三向镜像保持三个数据副本。每次写入都会发生复制，以确保所有数据副本始终为最新。镜像空间还可以跨多个物理驱动器分割数据。要实现镜像，必须部署至少两个物理磁盘。镜像提供了防止一个或多个磁盘丢失的保护，因此在存储重要数据时使用镜像。使用镜像的缺点是数据在多个磁盘上复制，因此磁盘使用效率低下。 奇偶校验 (Parity)。奇偶校验空间类似于简单空间，因为数据写入多个磁盘上。但是，当您使用奇偶校验存储布局时，奇偶校验信息也写在磁盘上。如果丢失磁盘，可以使用奇偶校验信息计算数据。奇偶校验信息使存储空间即使在驱动器发生故障时也能继续执行读写请求。奇偶校验信息始终在可用磁盘之间轮换，以启用 I/O 优化。存储空间至少需要三个物理驱动器用于奇偶校验空间。奇偶校验空间通过日志记录增加了弹性。奇偶校验存储布局提供冗余，但在磁盘空间利用方面比在镜像方面更有效。 <p> 注意：给定存储空间的列数也会可能影响磁盘数。</p>
磁盘扇区大小 (Disk sector size)	<p>存储池的扇区大小在创建时设置。其默认大小设置如下：</p> <ul style="list-style-type: none"> 如果正在使用的驱动器列表只包含 512 和 512e 驱动器，则池扇区大小 (pool sector size) 将设置为 512e。一个 512 磁盘使用 512 字节扇区。一个 512e 驱动器是一个具有 4,096 字节扇区的硬盘，其模拟 512 字节扇区。 如果列表包含至少一个 4 千字节 (KB) 驱动器，则池扇区大小将设置为 4 KB。
群集磁盘要求 (Cluster disk requirement)	<p>如果出现计算机故障，故障转移群集可防止工作中断。对于支持故障转移群集的池，池中的所有驱动器都必须支持 SAS。</p>

功能	描述
驱动器分配 (Drive allocation)	<p>驱动器分配定义驱动器如何分配给池。选项包括：</p> <ul style="list-style-type: none">• 数据存储 (Data-store)。这是将任何驱动器添加到池时的默认分配。存储空间可以自动选择数据存储驱动器上的可用容量，用于存储空间创建和 JIT 分配。• 手动 (Manual)。手动驱动器不用作存储空间的一部分，除非在创建该存储空间时特别选择它。此驱动器分配属性允许管理员指定特定类型的驱动器，仅供特定存储空间使用。• 热备用 (Hot spare)。这些是在创建存储空间时未使用但已添加到的池中的预留驱动器。如果托管存储空间列的驱动器发生故障，则会调用这些预留驱动器之一来替换发生故障的驱动器。
配置方案 (Provisioning schemes)	<p>您可以使用以下两种方案之一置备虚拟磁盘：</p> <ul style="list-style-type: none">• 精简配置空间 (Thin provisioning space)。精简配置能够在刚好足够的 JIT 基础上轻松分配存储。池中的存储容量被组织为在数据集需要存储之前未分配的配置板 (provisioning slabs)。代替传统的固定存储分配方法，其中存储容量的大部分被分配但是可能保持未使用，精简配置通过使用称为 <i>修剪 (trim)</i> 的过程来回收不再需要的存储来优化任何可用存储的使用。• 固定的配置空间 (Fixed provisioning space)。在存储空间中，固定配置空间也使用灵活的配置板。区别在于它在创建空间时预先分配存储容量。 <p>您可以在同一存储池中创建精简和固定配置虚拟磁盘。在同一个存储池中同时配置两种类型是很方便的，尤其是当它们与相同的工作负载相关时。例如，您可以选择对包含用户文件的共享文件夹使用精简配置空间，对需要高磁盘 I/O 的数据库使用固定配置空间。</p>
条带参数 (Stripe parameters)	<p>您可以通过在多个物理磁盘上分割数据来提高虚拟磁盘的性能。创建虚拟磁盘时，可以使用两个参数 <i>NumberOfColumns</i> 和 <i>Interleave</i> 配置条带。</p> <ul style="list-style-type: none">• 条带 (<i>stripe</i>) 表示写入存储空间的一次数据，数据写入多个条带或通过。• 列 (<i>Columns</i>) 与底层物理磁盘相关联，在其中写入存储空间的一个数据条。• 交错 (<i>Interleave</i>) 表示每个条带写入单列的数据量。 <p><i>NumberOfColumns</i> 和 <i>Interleave</i> 参数确定条带的宽度 (例如，$\text{stripe width} = \text{NumberOfColumns} * \text{Interleave}$)。条带宽度决定了多个数据和奇偶校验存储空间写入多个磁盘以提高应用程序的性能。通过使用带有 <i>NumberOfColumns</i> 和 <i>Interleave</i> 参数的 Windows PowerShell cmdlet <i>New-VirtualDisk</i>，可以在创建新虚拟磁盘时控制列数和条带交错。</p>

在创建池时，存储空间可以使用任何 DAS 设备。您可以使用在计算机内部连接的 SATA 和 SAS 驱动器 (甚至更老的集成驱动器电子设备 (integrated drive electronics) [IDE] 和 SCSI 驱动器)。规划存储空间存储子系统时，必须考虑以下因素：

- 容错。您是否希望在物理磁盘出现故障时数据依然可用？如果是这样，您必须使用多个物理磁盘并通过使用镜像或奇偶校验配置虚拟磁盘。

- 性能。您可以通过使用虚拟磁盘的奇偶校验布局 (parity layout) 来提高读取和写入操作的性能。在确定性能时, 还需要考虑每个单独物理磁盘的速度。或者, 您可以使用不同类型的磁盘为存储提供分层系统。例如, 您可以对需要快速和频繁访问的数据使用 SSD, 对于不经常访问的数据, 可以使用 SATA 驱动器。
- 可靠性。奇偶校验布局中的虚拟磁盘提供一些可靠性。您可以通过使用热备用物理磁盘来提高可靠性程度, 防止物理磁盘发生故障。
- 可扩展性。使用存储空间的一个主要优点是将来能够通过添加物理磁盘来扩展存储。创建存储空间后, 您可以在任何时间将物理磁盘添加到存储池, 以扩展其存储容量或提供容错。

演示: 配置存储空间 (Configuring Storage Spaces)

在本演示中, 您将了解如何:

- 创建存储池。
- 创建虚拟磁盘和卷。

演示步骤 (Demonstration Steps)

创建一个存储池 (Create a storage pool)

1. 在 LON-SVR1 上, 在 Server Manager 中, 访问 File and Storage Services 和 Storage Pools。
2. 在 STORAGE POOLS 窗格中, 创建名为 StoragePool1 的 New Storage Pool, 然后添加一些可用磁盘。

创建一个虚拟磁盘和卷 (Create a virtual disk and a volume)

1. 在 VIRTUAL DISKS 窗格中, 使用以下设置创建 New Virtual Disk :
 - 存储池: StoragePool1
 - 磁盘名称: Simple vDisk
 - 存储布局: Simple
 - 配置类型: Thin
 - 容量: 2 GB
2. 在 View results 页面上, 等待任务完成, 然后确保选中了 Create a volume when this wizard closes 复选框。
3. 在 New Volume Wizard 中, 使用以下设置创建卷:
 - 虚拟磁盘: Simple vDisk
 - 文件系统: ReFS
 - 卷标: Simple Volume
4. 等待任务完成, 然后单击 Close。

Windows Server 2016 文件和存储服务提供的新功能 (Chan Changes to file and storage services in Windows Server 2016)

文件和存储服务包括帮助您部署和管理一个或多个文件服务器的技术。

Windows Server 2016 中的新功能 (New features in Windows Server 2016)

Windows Server 2016 中新增或改进了以下文件和存储服务功能：

- 存储空间直通 (Storage Spaces Direct)。此功能使您能够通过使用仅具有本地存储的存储节点来构建高可用性存储系统。您将在本单元稍后了解有关此功能的更多信息。
- 存储副本 (Storage Replica)。Windows Server 2016 中的此新功能允许在位于相同位置或不同站点的服务器或群集之间进行复制，以用于灾难恢复。存储副本包括同步和异步复制，用于站点之间更短或更长的距离。这使您能够以更低成本实现存储复制。
- 存储 QoS (Storage Quality of Service , QoS)。通过此功能，您可以在横向扩展文件服务器 (Scale-Out File Server) 上创建集中式 QoS 策略，并将其分配给 Hyper-V 虚拟机的虚拟磁盘。QoS 确保存储的性能随着存储负载的变化而适应策略。
- 重复数据删除 (Data Deduplication)。此功能在 Windows Server 2012 中引入，并在 Windows Server 2016 中的以下方面进行了改进 (有关重复数据删除的更多信息将在本单元的后面部分介绍)：
 - 支持最大 64 TB 的卷大小。该功能在 Windows Server 2016 中已重新设计，现在是多线程的，并且能够利用每个卷的多个 CPU 在高达 64 TB 的大小的卷上提高优化吞吐量。
 - 支持高达 1 TB 的文件大小。通过使用新的流图结构 (stream map structures) 和其他改进来提高优化吞吐量和访问性能，Windows Server 2016 中的重复数据删除在高达 1 TB 的文件上运行良好。
 - 为虚拟化备份应用程序简化重复数据删除配置。在 Windows Server 2016 中，为卷启用重复数据删除时，可简化虚拟化备份应用程序的重复数据删除配置。
 - 支持 Nano 服务器。在 Windows Server 2016 中的新部署选项，Nano 服务器完全支持重复数据删除。
- 支持群集滚动升级 (Support for cluster rolling upgrades)。您可以将现有 Windows Server 2012 R2 群集中的每个节点升级到 Windows Server 2016，而不会导致停机时间，这样可以一次升级所有节点。
- 服务器消息块 (Server Message Block , SMB) 增强改进。在 Windows Server 2016 中，域控制器上的 Active Directory 域服务 (Active Directory Domain Service) 默认 SYSVOL 和 NETLOGON 共享的客户端连接现在需要 SMB 签名和相互身份验证 (例如 Kerberos 身份验证)。这种变化降低了中间人攻击 (man-in-the-middle attacks) 的可能性。如果 SMB 签名和相互身份验证不可用，则 Windows Server 2016 计算机将不会处理基于域的组策略和脚本。

• Windows Server 2016 提供以下新的文件和存储服务功能：

- 存储空间直通
- 存储副本
- 存储 QoS
- 重复数据删除 (改进)：
 - 支持最大 64 TB 的卷大小
 - 支持高达 1 TB 的文件大小
 - 虚拟化备份应用程序的简化重复数据删除配置
 - 支持 Nano 服务器
 - 支持群集滚动升级
- SMB 强化改进



注意：默认情况下，这些设置的注册表值不存在；但是，强制规则 (hardening rules) 仍然适用，直到组策略或其他注册表值覆盖它们。

Windows Server 2012 和 Windows Server 2012 R2 提供的新功能 (New features in Windows Server 2012 and Windows Server 2012 R2)

Windows Server 2012 R2 和 Windows Server 2012 在其前身提供了几个新的和改进的文件和存储服务功能，包括：

- 多兆字节卷 (Multiterabyte volumes)。此功能部署多兆字节 NTFS 文件系统卷, 支持合并方案并最大化存储使用。主引导记录 (MBR) 格式化磁盘上的 NTFS 卷大小可以高达 2 TB。全局唯一标识符 (GUID) 分区表 (GPT) 格式化磁盘上的卷最多可达 18 艾字节 (18 EB)。
- 重复数据删除 (Data deduplication)。此功能通过在卷上存储相同数据的单个副本来节省磁盘空间。
- iSCSI 目标服务器 (iSCSI Target Server)。iSCSI 目标服务器通过使用 iSCSI 标准为网络上的其他服务器和应用程序提供块存储。Windows Server 2012 R2 还通过使用存储管理计划规范 (Storage Management Initiative Specification) 来支持 VHDX 和端到端管理。
- 存储空间和存储池 (Storage spaces and storage pools)。此功能使您能够通过将行业标准磁盘分组到存储池中来虚拟化存储, 然后从存储池中的可用容量创建存储空间。Windows Server 2012 R2 中的存储空间使您能够创建一个分层存储解决方案, 透明地在容量和性能之间实现适当的平衡, 从而满足企业工作负载的需求。
- 服务器管理器中的文件和存储服务的统一远程管理 (Unified remote management of File and Storage Services in Server Manager.)。您可以使用服务器管理器远程管理多个文件服务器, 包括其角色服务和存储。
- 文件和存储服务 (File and Storage Services) 的 Windows PowerShell cmdlet。您可以使用 Windows PowerShell cmdlet 执行文件和存储服务的大多数管理任务。
- ReFS。Windows Server 2012 中引入的新的复原文件系统 (ReFS) 为基于文件的数据存储提供了增强的完整性, 可用性, 可扩展性和错误保护。
- 服务器消息块 (Server Message Block , SMB) 3.0。SMB 协议是一种网络文件共享协议, 允许应用程序读取和写入文件并从网络上的服务器程序请求服务。
- 卸载数据传输 (Offloaded Data Transfer , ODX)。ODX 功能使支持 ODX 的存储阵列绕过主机计算机, 并在兼容的存储设备之间或之间直接传输数据。
- Chkdsk。新版本的 Chkdsk 在后台自动运行并监视系统卷的运行状况; 使组织能够部署多兆字节 NTFS 文件系统卷, 而不必担心危及其可用性。Chkdsk 工具引入了一种新方法。它优先考虑卷可用性, 并允许在卷保持联机状态时检测损坏, 并且其数据在维护期间仍可供用户使用。

存储空间使用场景 (Storage Spaces usage scenarios)

在考虑在给定情况下是否使用存储空间时, 应该权衡以下好处和限制。存储空间功能旨在使存储管理员能够:

- 实施并轻松管理可扩展, 可靠和便宜的存储。
- 将单个驱动器聚合到存储池中, 这些存储池作为单个实体进行管理。
- 使用廉价存储, 带或不带外部存储。
- 在同一池中使用不同类型的存储 (例如 SATA , SAS , USB , SCSI)。
- 根据需要增加存储池。
- 需要时从先前创建的存储池调配存储。
- 将特定驱动器指定为热备用磁盘。
- 自动修复包含热备用磁盘的池。
- 按池委派管理。

• 存储空间功能:

- 实现和轻松管理可扩展, 可靠和廉价的存储
- 使用具有或不具有外部存储的廉价存储
- 将多个驱动器合并到存储池中, 管理员可以将其作为单个实体进行管理
- 在同一池中实现不同类型的存储
- 根据需要增加存储池
- 根据需要从现有存储池中调配存储
- 将特定驱动器指定为热备用磁盘



- 使用现有的备份和恢复工具，以及卷影复制服务（VSCS）创建快照。
- 管理可以是本地或远程，通过使用 Microsoft 管理控制台（MMC）或 Windows PowerShell。
- 利用具有故障转移群集的存储空间。



注意：虽然上面的列表提到 USB 作为支持的存储介质，在池中使用 USB 可能在 Windows 8 客户端上或在开发概念证明时更实用。该技术的性能还取决于您选择集中在一起放在池中的存储的性能。

然而，存储空间中存在固有的限制。例如，在 Windows Server 2016 中，以下是在规划时应考虑的一些限制：

- 引导卷或系统卷不支持存储空间卷。
- 将驱动器引入存储池时，驱动器的内容将丢失。
- 您应该仅添加未格式化或未分区的驱动器。
- 在简单存储池中必须至少有一个驱动器。
- 容错配置具有特定要求：
- 镜像池至少需要两个驱动器。
- 三向镜像需要至少五个驱动器。
- 奇偶校验至少需要三个驱动器。
- 池中的所有驱动器必须使用相同的扇区大小。
- 抽象物理磁盘的存储层与存储空间不兼容，包括：
- 虚拟机（VM）中的 VHD 和直通（pass-through）磁盘。
- 部署在单独的 RAID 层中的存储子系统。
- 不支持光纤通道和 iSCSI。
- 故障转移群集仅限于 SAS 作为存储介质。



注意：Microsoft 支持仅在环境中的物理机而不是虚拟机上部署存储空间时提供故障诊断帮助。此外，您实施的 JBOD（just a bunch of disks）硬件解决方案必须经过 Microsoft 认证。

在规划环境中特定工作负载的可靠性时，存储空间提供不同的弹性类型。因此，一些工作负载更适合特定的弹性场景。下表描述了这些建议的工作负载类型。

弹性类型	维护的数据副本个数	工作负载建议
镜像	2 (两路镜像) 3 (三路镜像)	建议用于所有工作负载
奇偶校验	2 (单奇偶校验) 3 (双奇偶校验)	具有大单位读/写的顺序工作负载，例如归档
简单	1	不需要弹性的工作负载，或提供备用弹性机制

存储空间直通部署场景 (Storage Spaces Direct deployment scenarios)

存储空间直通直接消除了对共享 SAS 架构的需求, 简化了部署和配置。相反, 它使用现有网络作为存储结构, 利用 SMB 3.0 和 SMB Direct 实现高速, 低延迟 CPU 高效存储。要向外扩展, 您只需添加更多服务器以增加存储容量和 I/O 性能。

存储空间直通可以部署为支持 Hyper-V 虚拟机 (VM) 文件的主存储或 Hyper-V 副本 (Hyper-V Replica) 虚拟机文件的辅助存储。在 Windows Server 2016 中, 这两个选项都为 Hyper-V 提供存储, 特别关注针对服务提供商和企业的 Hyper-V IaaS (基础架构即服务)。

在分解 (disaggregated) 部署方案中, Hyper-V 服务器 (计算组件) 位于与存储空间直通服务器 (存储组件) 不同的群集中。虚拟机配置为将其文件存储在横向扩展文件服务器 (Scale-Out File Server, SOFS) 上。SOFS 被设计为用作服务器应用程序数据的文件共享, 并且使用 SMB 3.0 协议通过网络访问。这允许独立地扩展 Hyper-V 群集 (计算) 和 SOFS 群集 (存储)。

在超融合部署方案中, Hyper-V (计算) 和存储空间直通 (存储) 组件位于同一群集上。此选项不需要部署 SOFS, 因为虚拟机文件存储在 CSV 上。这允许将 Hyper-V 计算群集和存储一起扩展, 并且不需要配置文件服务器访问和权限。一旦配置存储空间直通并且 CSV 卷可用后, 配置 Hyper-V 是相同的过程, 并使用与故障转移群集上任何其他 Hyper-V 部署相同工具。

您还可以部署存储空间直通支持 SQL Server 2012 或更高版本, 它可以存储系统和用户数据库文件。SQL Server 配置为将这些文件存储在 SMB 3.0 文件共享上, 用于 SQL Server 的独立和群集实例。数据库服务器使用 SMB 3.0 协议通过网络访问 SOFS。此方案需要 Windows Server 2012 或更高版本的文件服务器和数据库服务器。



注意: 存储空间目前不支持 Exchange Server 工作负载。

与 Azure 虚拟机场景的互操作性 (Interoperability with Azure virtual machines scenarios)

您可以使用 Azure 虚拟机内的存储空间来组合多个虚拟硬盘驱动器, 从而创建比单个 Azure 虚拟硬盘驱动器更多的存储容量或性能。在 Azure 虚拟机中使用存储空间有三种支持的方案, 但是您应遵循一些限制和最佳实践, 如下所述。

- 作为虚拟机的高性能和/或容量存储。
- 作为 System Center Data Protection Manager 的备份目标。
- 作为 Azure 站点恢复的存储。

多租户场景 (Multi-tenant scenarios)

您可以通过访问控制列表 (ACL) 提供对存储池的管理委派。您可以基于每个存储池进行委派, 从而支持需要租户隔离的托管方案。由于存储空间使用 Windows 安全模型, 因此它可以与 Active Directory 域服务完全集成。

可以使存储空间仅对文件群集中的节点的子集可见。这可以在某些情况下用于利用较大共享群集的成本和管理优势, 并且为了性能或访问目的将这些群集分段。此外, 您可以在存储堆栈的各个级别应用 ACL (例如, 文件共享, CSV 和存储空间)。在多租户场景中, 这意味着可以集中共享和管理完整的存储基础设施, 并且可以设计专用和受控访问存储基础架构的各个部分。您可以为特定客户配置专用于它们的 LUN, 存储池, 存储空间, 群集共享卷和文件共享, 并且 ACL 可以确保只有租户可以访问它们。

此外, 通过使用 SMB 加密, 您可以确保对基于文件的存储的所有访问都被加密, 以防止篡改和窃听攻击。与更通用的解决方案 (如 IPsec) 相比, 使用 SMB 加密的最大好处是, 除了更改服务器上的 SMB 设置之外, 没有部署要求或成本。所使用的加密算法是 AES-CCM, 其还提供数据完整性验证。

讨论：比较存储空间和其他存储解决方案 (Comparing Storage Spaces to other storage solutions)

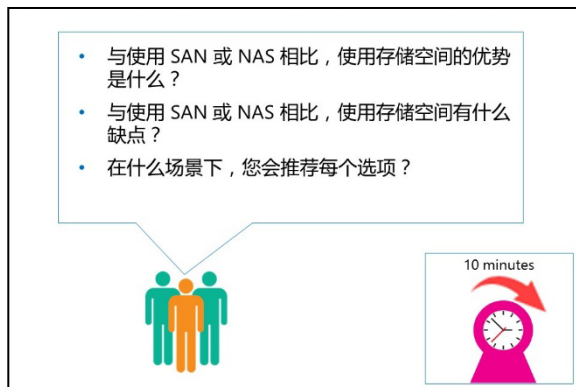
Windows Server 2016 中的存储空间提供了使用更传统存储解决方案 (如 SAN 和网络连接存储 (NAS)) 的备选方案。

考虑以下问题以准备课堂讨论：

问题：与使用 SAN 或 NAS 相比，使用存储空间有什么优势？

问题：与使用 SAN 或 NAS 相比，使用存储空间有什么缺点？

问题：在什么情况下，您会推荐每个选项？



第 2 课

管理存储空间 (Managing Storage Spaces)

实现存储空间后，您必须知道如何管理和维护它们。本课程探讨如何使用存储空间来缓解磁盘故障，扩展存储池，以及使用日志和性能计数器来确保存储的最佳行为。

课程目标 (Lesson Objectives)

完成本课后，您将能够：

- 描述如何管理存储空间。
- 解释如何使用存储空间来缓解存储故障。
- 解释如何扩展存储池。
- 描述如何使用事件日志和性能计数器监视存储空间。

管理存储空间 (Managing Storage Spaces)

存储空间与故障转移群集集成以实现高可用性，并与群集共享卷 (cluster shared volumes , CSV) 集成，用于 SOFS 部署。您可以使用以下命令管理存储空间：

- 服务器管理器
- Windows PowerShell
- 故障转移群集管理器 (Failover Cluster Manager)
- System Center Virtual Machine Manager
- Windows 管理规范 (Windows Management Instrumentation , WMI)

• 您可以使用以下方法管理存储空间：

• 服务器管理器

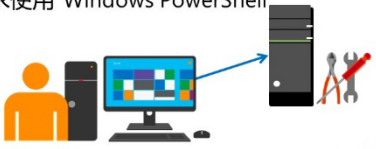
• Windows PowerShell

• 故障转移群集管理器

• 系统中心虚拟机管理器

• Windows 管理规范

• 高级管理要求使用 Windows PowerShell



使用服务器管理器进行管理 (Manage using Server Manager)

服务器管理器使您能够执行虚拟磁盘和存储池的基本管理。在服务器管理器中，您可以创建存储池；添加和删除池中的物理磁盘；并创建，管理和删除虚拟磁盘。例如，在服务器管理器中，您可以查看连接到虚拟磁盘的物理磁盘。如果任何这些磁盘不健康，您将在磁盘名称旁看到不健康的磁盘图标。

使用 Windows PowerShell 进行管理 (Manage using Windows PowerShell)

Windows PowerShell 为虚拟磁盘和存储池提供高级管理选项。下表列出了管理 cmdlet 的一些示例。

Windows PowerShell cmdlet	描述
Get-StoragePool	列出存储池
Get-VirtualDisk	列出虚拟磁盘
Repair-VirtualDisk	修复虚拟磁盘
Get-PhysicalDisk Where {\$_.HealthStatus -ne "Healthy"}	列出不健康的物理磁盘

Reset-PhysicalDisk	从存储池里移除物理磁盘
Get-VirtualDisk Get-PhysicalDisk	列出用于虚拟磁盘的物理磁盘
Optimize-Volume	优化卷，在支持的卷和系统 SKU 上执行此类任务，如碎片整理，修剪（slab），碎片合并（slab consolidation）和存储层处理。



附加阅读：有关详细信息，请参阅：“Windows PowerShell 中的 Storage Cmdlet”，网址：

<http://aka.ms/po9qve>

要在 Windows PowerShell 中使用存储空间 cmdlet，必须下载 StorageSpaces 模块才能在 Windows Server 2016 中使用。有关详细信息，请参阅：“Windows PowerShell 中的存储空间 Cmdlet”，网址：<http://aka.ms/M1fccc>

监控存储层性能（Monitoring storage tier performance）

在规划存储分层时，应评估存储环境的工作负载特征，以便您可以根据使用方式最具成本效益地存储数据。在 Windows Server 2016 中，服务器通过将更频繁访问的数据透明地移动到更快的固态驱动器（SSD 层），并将不太活跃的数据移动到更便宜但容量更大的硬盘驱动器 HDD 层）。

在许多环境中，最常见的工作负载特征包括大数据集，其中大多数数据通常是冷的。冷数据是您不经常访问的文件，并且具有更长的使用寿命。相比之下，最常见的工作负载特征还包括典型地热数据的较小部分。热数据，通常称为工作集，是您当前正在出来的文件；这部分数据集是高度活跃的，随着时间的推移而变化。



注意：存储层优化过程移动数据，而不是文件；数据被映射并在子文件级被映射和移动。

例如，如果虚拟硬盘上只有 30% 的数据是热的，那么只有 30% 的数据被移动到 SSD 层。

此外，在规划存储分层时，应评估文件放置在特定层时是否最有效的情况。例如，您需要将一个重要文件放在快速层中，或者需要将一个备份文件放在缓慢的层中。对于这些情况，您的存储解决方案可能会选择将文件分配给特定层，也称为将文件固定到层。

在创建存储空间之前，请先做好计划，在观察到您的工作负载后，给自己留出空间来调整存储空间。在观察每秒的输入/输出操作（input/output operations per second，IOPS）和延迟后，您将能够更准确地预测每个工作负载的存储要求。以下是一些建议：

- 不要立即为您的存储空间分配所有可用的 SSD 容量。将存储池中的一些 SSD 容量保留下来，以便在工作负载需要时可以增加 SSD 层的大小。
- 不要将文件固定到存储层，直到您看到存储层优化可以优化存储性能。当租户或工作负载需要特定级别的性能时，您可以将文件固定到存储层，以确保在该层上执行所有 I/O 活动。
- 如果您通过 VDI 提供池桌面，请考虑将父 VHDX 文件固定到 SSD 层。如果已部署虚拟桌面基础架构（Virtual Desktop Infrastructure，VDI）为用户提供池桌面，则应考虑将用于克隆用户桌面的主映像固定到 SSD 层。
- 在观察或监视工作负载时，应使用存储层优化报告（Storage Tier Optimization Report）。此报告用于检查存储层的性能，并确定可能优化其性能的更改。作为性能分析的一部分，报告提供了回答问题的数据，例如“我的工作集有多大？”和“通过添加 SSD 容量我能获得多少？”



附加阅读：有关详细信息，请参考：“监视存储层性能”，网址是：<http://aka.ms/Sz4zfj>


海量视频题库 myipub.com QQ:5565462

使用存储空间管理磁盘故障 (Managing disk failure with Storage Spaces)

在部署之前，您应该计划存储空间来处理磁盘和 JBOD 机箱故障，对服务的影响最小，数据丢失的风险最小。使用任何存储解决方案，您应该预期会发生硬件故障；这在大规模存储解决方案中尤其如此。为了帮助避免由于硬件故障导致的问题，您的存储计划应考虑在您的环境中可能发生的故障的类型和数量。您还应规划解决方案应如何处理每个故障，而不会中断服务。

- 增强磁盘容错能力：
 - 设计一个完整的容错存储解决方案
 - 部署高可用性存储池
 - 验证硬件和固件组件
 - 立即更换故障磁盘
 - 保留一些未分配的空间
 - 准备好多个磁盘故障
 - 在机箱级别提供容错功能

- 设计一个完整的容错存储解决方案。例如，如果希望存储解决方案能够容忍任何级别的单个故障，则需要执行此最低设置：
 - 双向镜像或单奇偶校验存储空间。
 - 文件服务器群集。
 - 每个文件服务器节点和每个 JBOD 之间的冗余 SAS 连接。
 - 冗余网络适配器和网络交换机。
 - 足够的 JBOD 机箱，以容许整个 JBOD 出现故障或断开连接。
- 部署高度可用的存储池。在存储空间中使用镜像或奇偶校验虚拟磁盘为存储资源提供了一些容错和高可用性。但是，因为所有物理磁盘连接到单个系统，所以该系统本身成为单个故障点。如果物理磁盘连接的系统出现故障，则对存储资源的访问将不复存在。Windows Server 2016 中的存储空间支持在使用镜像空间，奇偶校验空间和简单空间时创建群集存储池。要群集存储空间，您的环境必须满足以下要求：
 - 存储池中的所有存储空间必须使用固定配置。
 - 双向镜像空间必须使用三个或更多物理磁盘。
 - 三向镜像空间必须使用五个或更多物理磁盘。
 - 必须使用 SAS 连接群集池中的所有物理磁盘。
 - 所有物理磁盘必须支持持久保留，并通过故障转移群集验证测试

 **注意：** SAS JBOD 必须物理连接到将使用存储池的所有群集节点。具有存储空间的群集存储池不支持未连接到所有群集节点的直接连接存储。

- 除非部署了高可用性存储池，否则如果系统失败，则在另一台服务器上导入存储池。在 Windows Server 2016 中，存储空间将有关存储池的配置直接写入磁盘。因此，如果单点故障系统发生故障，并且服务器硬件需要更换或完全重新安装，则可以在另一台服务器上安装存储池。
- 存储空间的大多数问题是由于不兼容的硬件或固件问题。要减少问题，请遵循以下最佳实践：
 - 仅使用经认证的 SAS 连接的 JBOD。这些机箱型号已使用存储空间进行了测试，可以轻松地识别机箱和物理磁盘的插槽。
 - 不要混合和匹配 JBOD 中的磁盘模型。请使用一个型号的固态硬盘 SSD 和一个型号的 HDD 用于 JBOD 中的所有磁盘（假设您使用存储层），并确保磁盘与 JBOD 模型完全兼容。
 - 在所有磁盘上安装最新的固件和驱动程序版本。安装在 Windows Server Catalog 中列出的为设备批准的固件版本，或者由硬件供应商推荐。在 JBOD 中，同一型号的所有磁盘具有相同的固件版本是很重要的。

- 按照供应商的建议进行磁盘放置。在硬件供应商推荐的插槽中安装磁盘。由于冷却和其他原因，JBODs 对 SSD 和 HDD 的放置通常有不同的要求。
- 除非启用热备用磁盘，否则自动删除缺少的磁盘。处理从存储池丢失的物理磁盘（- RetireMissingPhysicalDisks = Auto）的默认策略只是将磁盘标记为缺少（丢失通信），并且不会对虚拟磁盘执行修复操作。如果磁盘临时脱机，此策略避免潜在的 I/O 密集型虚拟磁盘修复，但存储池的运行状况将保持降级，如果在管理员采取操作之前另一个磁盘发生故障，则会降低恢复能力。除非使用热备盘，否则我们建议您将 RetireMissingPhysicalDisks 策略更改为 Always，以便在磁盘与系统失去通信时自动启动虚拟磁盘修复操作，尽快恢复池的健康状况和相关存储空间。
- 始终在从存储池中删除驱动器之前更换物理磁盘。在更换机箱中的物理磁盘之前更改存储池配置可能会导致 I/O 故障或启动虚拟磁盘修复，从而导致“STOP 0x50”错误和潜在的数据丢失。
- 作为常规规则，在池中保留未分配的磁盘空间以进行虚拟磁盘修复，而不是使用热备盘。在 Windows Server 2016 中，您可以选择使用池中现有磁盘上的可用容量进行磁盘修复操作，而不是使热备用磁盘联机。这使存储空间能够通过将数据复制到池中的多个磁盘来自动修复具有故障磁盘的存储空间，与使用热备份相比，显著减少从故障磁盘进行恢复所需的时间，并且可以使用所有磁盘，而不是设置热备份。
 - 要更正虚拟磁盘或存储池中的故障磁盘，必须删除导致该问题的磁盘。碎片整理，扫描磁盘或使用 chkdsk 等操作无法修复存储池。
 - 要更换故障磁盘，必须向池中添加新磁盘。在日常维护期间进行磁盘维护时，新磁盘会自动重新同步。或者，您可以手动触发磁盘维护。
- 配置列计数时，请确保有足够的物理磁盘来支持自动虚拟磁盘修复。通常，您应该使用 3-4 列配置虚拟磁盘，以实现良好的吞吐量平衡和低延迟。增加列计数可增加条带化虚拟磁盘的物理磁盘数，从而提高该虚拟磁盘的吞吐量和 IOPS。但是，增加列计数也会增加延迟。因此，您应该使用包含 3-4 列（使用镜像时）的多个虚拟磁盘或使用奇偶校验空间时的 7 列来优化整体群集性能。由于并行使用多个虚拟磁盘，整个群集的性能仍然很高，弥补了减少了列数。
- 准备好多个磁盘故障。如果您同时购买机箱中的所有磁盘，则磁盘的年龄相同，并且一个磁盘的故障可能会很快由其他磁盘故障导致。即使存储空间在初始磁盘修复后恢复到运行状况，也应该尽快更换故障磁盘，以避免出现其他磁盘故障的风险，这可能会危及存储健康和可用性以及面临数据丢失的风险。如果您希望能够在下次计划的维护之前安全地延迟磁盘修复，请将存储空间配置为容忍两个磁盘故障。
- 在机箱级别提供容错。如果需要提供更高级别的容错级别（在机箱级别），请部署多个支持机箱感知的兼容 JBOD。在机箱感知存储解决方案中，存储空间将每个数据副本写入特定的 JBOD 机箱。因此，如果一个机箱发生故障或脱机，数据在另一个或多个备用机箱中保持可用。要使用存储空间进行机箱感知，您的环境必须满足以下要求：
 - JBOD 存储机箱必须支持 SCSI 机箱服务（SCSI Enclosure Services，SES）。
 - 存储空间必须配置为镜像。
 - 要允许具有双向镜像的故障机箱，您需要三个兼容的存储机箱。
 - 要允许具有三向镜像的两个故障机箱，您需要五个兼容的存储机箱。

存储池扩展 (Storage pool expansion)

使用存储空间的一个主要优点是能够通过添加额外的存储来扩展存储池。但是，有时候，您必须先调查存储在池中的磁盘上的使用方式，然后才能扩展存储空间。这是因为，各种虚拟磁盘的块分布在存储池中的物理磁盘上，该配置基于您在创建池时选择的存储布局选项。根据具体情况，即使池中有可用空间，您也可能无法扩展存储。

范例 (Example)

请考虑以下示例：

在第一个图中，存储池由五个磁盘组成，其中 disk1 大于其他磁盘。 vdisk1 在所有五个磁盘上消耗空间，而 vdisk2 仅消耗磁盘 1 至 3 上的空间

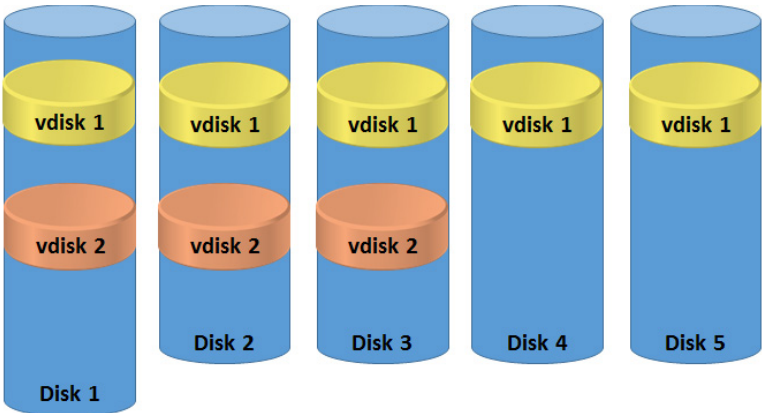
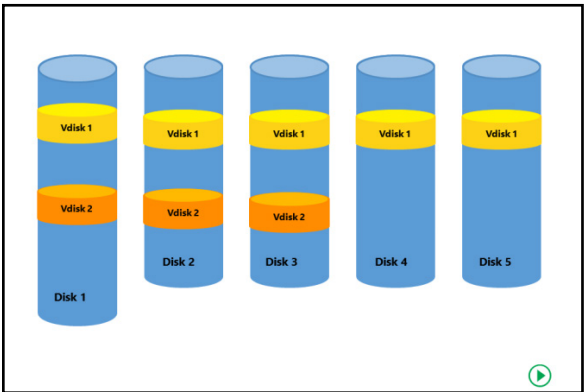


图 4.1 : A STORAGE POOL CONSISTING OF FIVE DISKS

在第二个图中，第六个磁盘已添加到存储池。

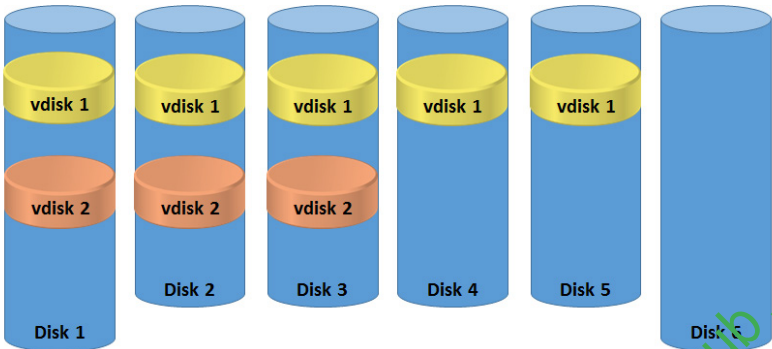


图 4.2 : A STORAGE POOL CONSISTING OF SIX DISKS

- 如果尝试扩展 vdisk1，尽管池中的磁盘 6 中有更多空间可用，但该磁盘的最大可用空间已被使用。这是因为 vdisk1 所需的布局 - 由于创建时选择的选项（例如镜像和奇偶校验） - 需要五个磁盘。因此，要扩展 vdisk1，您需要添加四个额外的磁盘。
- 但是，如果尝试扩展 vdisk2，则可以这样做，因为该磁盘当前分布在三个设备上，并且这三个设备之间有可用空间来扩展它。



注意：在存储空间中，已阻止的存储被排列为列。因此，在预扩展状态下，vdisk1 使用五列，vdisk2 使用三列。

- Vdisk2 可能只是使用双向镜像的虚拟磁盘。这意味着 disk1 上的数据在 disk2 和 disk3 上重复。如果要使用双向镜像扩展虚拟磁盘，则必须具有适当数量的列以适应虚拟磁盘的需要

确定列使用情况 (Determining Column Usage)

在将存储添加到存储池之前，必须通过确定列的使用情况来确定跨设备的块的当前分布。为此，您可以使用 Windows PowerShell cmdlet `Get-VirtualDisk`。



注意：有关详细信息，请参阅：“存储空间常见问题 (FAQ)”，网址是：
<http://aka.ms/knx5zg>

扩展存储池 (Expanding a storage pool)

在必要时确定列使用情况之后，可以使用以下选项之一扩展存储池：

- 服务器管理器。打开 Server Manager，选择 File and Storage Services，然后单击 Storage Pools。您可以通过右键单击池添加物理磁盘，然后单击 Add Physical Disk。
- Windows PowerShell。您可以使用 Windows PowerShell cmdlet `Add-PhysicalDisk` 将物理磁盘添加到存储池。例如：

```
Add-PhysicalDisk -VirtualDiskFriendlyName UserData -PhysicalDisks (Get-PhysicalDisk -
FriendlyName PhysicalDisk3, PhysicalDisk4)
```

演示: 使用 Windows PowerShell 管理存储空间 (Managing Storage Spaces by using Windows PowerShell)

在本演示中，您将了解如何使用 Windows PowerShell：

- 查看存储池的属性。
- 将物理磁盘添加到存储池。

演示步骤 (Demonstration Steps)

查看存储池属性 (View the properties of a storage pool)

1. 在 LON-SRV1 上，打开 Windows PowerShell。
2. 在服务器管理器中查看当前存储配置。
3. 运行下列命令：

- a. 要返回具有当前运行状况和运行状态的存储池列表，请运行以下命令：

```
Get-StoragePool
```

- b. 要返回有关 StoragePool1 的更多信息，请运行以下命令：

```
Get-StoragePool StoragePool1 | fl
```

- c. 要返回有关虚拟磁盘的详细信息，包括配置类型，奇偶校验布局和运行状况，请运行以下命令：

```
Get-VirtualDisk | fl
```


- d. 要返回可以池化的物理磁盘的列表，请运行以下命令：

```
Get-PhysicalDisk | Where {$_.canpool -eq "true"}
```

添加物理磁盘到存储池（Add physical disks to a storage pool）

1. 运行下列命令：

- a. 要在 StoragePool1 中创建新的虚拟磁盘，请运行以下命令：

```
New-VirtualDisk -StoragePoolFriendlyName StoragePool1 -FriendlyName Data -Size 2GB
```

您可以在服务器管理器中看到这个新的虚拟磁盘。

- b. 要添加可以合并到变量的物理磁盘列表，请运行以下命令：

```
$canpool = Get-PhysicalDisk -CanPool $true
```

- c. 要将变量中的物理磁盘添加到 StoragePool1，请运行以下命令：

```
Add-PhysicalDisk -PhysicalDisks $canpool -StoragePoolFriendlyName StoragePool1
```

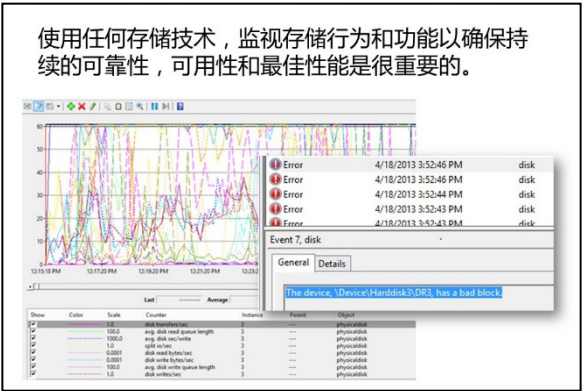
2. 在服务器管理器中查看其他物理磁盘。

事件日志和性能计数器（Event logs and performance counters）

使用任何存储技术，重要的是监控存储行为和功
能，以确保持续的可靠性，可用性和最佳性能。

使用事件日志（Using the Event Log）

当存储体系结构发现问题时，存储空间会生成错
误，然后将这些错误记录到事件日志中。您可以
使用事件日志工具或通过使用服务器管理器或
Windows PowerShell cmdlet 访问记录的错误来访问
这些事件。下表标识与有问题的存储关联的常见
事件 ID。



事件 ID	信息	原因
100	物理驱动器 %1 无法读取存储池 %2 的配置或返回的损坏数据。因此，内存中配置可能不是配置的最新副本。返回代码：%3。	由于以下原因，物理驱动器可能无法读取配置或返回存储池的损坏数据。 物理驱动器可能会出现设备 I/O 错误的请求失败。 物理驱动器可能包含损坏的存储池配置数据。 物理驱动器可能包含不足的内存资源。

Event ID	信息	原因
102	存储池 %1 的大多数物理驱动器的配置更新失败，导致池进入故障状态。返回代码：%2。	<ul style="list-style-type: none"> 将存储池配置写入物理驱动器时可能会出现写入失败，原因如下： 物理驱动器可能会出现设备 I/O 错误的请求失败。 没有足够数量的物理驱动器联机，并使用其最新配置进行更新。 物理驱动器可能包含不足的内存资源。
103	存储池 %1 的容量使用量已超过池上设置的阈值限制。返回代码：%2。	存储池的容量消耗已超过池上设置的阈值限制。
104	存储池 %1 的容量消耗现在低于在池上设置的阈值限制。返回代码：%2。	存储池的容量消耗将返回到低于池上设置的阈值限制的级别。
200	Windows 无法读取物理驱动器 %1 的驱动器头。如果您知道驱动器仍然可用，则使用命令行或 GUI 重置驱动器运行状况可能会清除此故障情况，并允许您将驱动器重新分配给其存储池。返回代码：%2。	Windows 无法读取物理驱动器的驱动器头。
201	物理驱动器 %1 具有无效的元数据。使用命令行或 GUI 重置运行状况状态可能会将物理驱动器带到原始池。返回代码：%2。	物理驱动器上的元数据已损坏。
202	物理驱动器 %1 具有无效的元数据。使用命令行或 GUI 重置运行状况状态可能会解决问题。返回代码：%2。	物理驱动器上的元数据已损坏。
203	物理驱动器 %1 上发生 I/O 故障。返回代码：%2。	物理驱动器上发生 I/O 故障。
300	物理驱动器 %1 无法读取存储空间 %2 的配置或返回的损坏数据。因此，内存中配置可能不是配置的最新副本。返回代码：%3。	<ul style="list-style-type: none"> 物理驱动器可能无法读取配置或返回损坏的数据，原因如下： 物理驱动器可能会失败具有设备 I/O 错误的请求。 物理驱动器可能包含损坏的存储空间配置数据。 物理驱动器可能包含不足的内存资源。
301	所有池驱动器无法读取存储空间 %1 的配置或返回的损坏数据。因此，存储空间将不会附加。返回代码：%2。	<ul style="list-style-type: none"> 您可能会遇到所有物理驱动器无法读取其配置或返回存储空间的损坏数据的原因如下： 物理驱动器可能会出现设备 I/O 错误的请求失败。 物理驱动器可能包含损坏的存储池配置数据。

		<ul style="list-style-type: none"> 物理驱动器可能包含不足的内存资源。
302	大多数存储空间 %1 的池驱动器托管空间元数据在空间元数据更新时失败，这导致存储池进入失败状态。返回代码：%2。	<ul style="list-style-type: none"> 托管存储空间的空间元数据的大多数池驱动器可能会失败元数据更新，原因如下： 物理驱动器可能会出现设备 I/O 错误的请求失败。 物理驱动器数量不足，具有联机存储空间元数据。 物理驱动器可能包含不足的内存资源。
303	托管存储空间数据的驱动器失败或丢失。因此，没有可用的数据副本。返回代码：%2。	如果存储池中的驱动器出现故障或已删除，则可能会发生此事件。
304	托管存储空间 %1 的数据的一个或多个驱动器失败或丢失。因此，至少有一个数据副本不可用。但是，至少有一个数据副本仍然可用。返回代码：%2。	托管存储空间数据的一个或多个驱动器失败或丢失。因此，至少有一个数据副本不可用。但是，至少有一个数据副本仍然可用。
306	尝试映射或分配存储空间 %1 的更多存储失败。这是因为更新存储空间元数据时涉及写入失败。返回代码：%2。	尝试映射或分配更多存储空间已失败。需要更多的物理驱动器。
307	尝试取消映射或修剪存储空间 %1 失败。返回代码：%2。	尝试取消映射或修剪所列出的存储空间失败。
308	驱动程序发起了对存储空间 %1 的修复尝试。返回代码：%2。	驱动程序启动了对存储空间的修复尝试。这是正常情况。不需要采取进一步行动。

性能监视 (Performance monitoring)

大多数关于存储架构配置的决策都会影响存储架构的性能。使用存储空间实现存储架构也是如此。由于多种因素 (包括成本, 可靠性, 可用性, 功耗和易用性) 之间的平衡, 性能会更好或更差。

存储架构中有多个组件可以处理存储请求, 包括:

- 文件高速缓存管理 (File cache management)。
- 文件系统架构。
- 卷管理。
- 物理存储硬件。
- 存储空间配置选项。

您可以使用 Windows PowerShell 和性能监视器 (Performance Monitor) 来监视存储池的性能。如果要使用 Windows PowerShell, 则必须为 Windows PowerShell 安装存储空间性能分析 (Storage Spaces Performance Analysis) 模块。



注意: 要下载 “用于 Windows PowerShell 的存储空间性能分析模块”, 请转到:

<http://aka.ms/b1d52u>

要使用 Windows PowerShell 生成和收集性能数据, 请在 Windows PowerShell 提示符下运行以下 cmdlet:

```
Measure-StorageSpacesPhysicalDiskPerformance -StorageSpaceFriendlyName StorageSpace1 -  
MaxNumberOfSamples 60 -SecondsBetweenSamples 2 -ReplaceExistingResultsFile -  
ResultsFilePath StorageSpace1.blg -SpacetoPDMMappingPath PDMAP.csv
```

此 cmdlet:

- 监视与名为 StorageSpace1 的存储空间相关联的所有物理磁盘的性能。
- 以两秒钟的间隔捕获 60 秒的性能数据。
- 替换结果文件 (如果已存在)。
- 将性能日志存储在名为 StorageSpace1.blg 的文件中。
- 将物理磁盘映射信息存储在名为 PDMAP.csv 的文件中。

您可以使用性能监视器查看在上面的 cmdlet 中指定的两个文件中收集的数据, 名为 StorageSpace1.blg 和 PDMAP.csv。

实验 A: 实现存储空间 (Implementing Storage Spaces)

场景 (Scenario)

A. Datum 公司已经购买了一些硬盘驱动器和 SSD, 你已经被委托创建一个存储解决方案, 可以充分利用这些新的设备。在 A. Datum 中对数据访问和冗余有混合要求, 因此必须确保为不需要快速磁盘读写访问的关键数据提供冗余解决方案。您还必须为需要快速读取和写入访问的数据创建解决方案。

您决定使用存储空间和存储分层 (storage tiering) 来满足要求。

目标 (Objectives)

完成本实验后, 您将能够:

- 创建存储空间。
- 启用和配置存储分层。

实验设置 (Lab Setup)

预计时间: 40 分钟

虚拟机: 28740B-LON-DC1, 28740B-LON-SVR1

用户名: Adatum\Administrator

密码: Pa55w.rd

对于本实验, 您需要使用可用的虚拟机环境。在开始实验之前, 您必须完成以下步骤:

1. 在主机计算机上, 启动 Hyper-V Manager。
2. 在 Hyper-V Manager 中, 单击 28740B-LON-DC1, 然后在 Actions 窗格中单击 Start。
3. 在 Actions 窗格中, 单击 Connect。等待直到虚拟机启动。
4. 使用以下凭据登录:
 - 用户名: Administrator
 - 密码: Pa55w.rd
 - 域: Adatum
5. 对 28740B-LON-SVR1 重复步骤 2 到 4。

练习 1: 创建一个存储空间 (Creating a Storage Space)

场景 (Scenario)

您的服务器没有基于硬件的 RAID 卡, 但是您已被要求配置冗余存储。要支持此功能, 必须创建一个存储池。

创建存储池后, 必须创建冗余虚拟磁盘。由于数据很关键, 因此对冗余存储的请求指定必须使用三向镜像卷。在卷使用后不久, 磁盘发生故障, 您必须通过向存储池添加另一个磁盘来替换它。

本练习的主要任务如下:

1. 从连接到服务器的六个磁盘创建存储池。
2. 创建三向镜像虚拟磁盘 (至少需要五个物理磁盘)。
3. 将文件复制到卷, 并验证它在文件资源管理器中可见。
4. 删除物理驱动器以模拟驱动器故障。

5. 验证文件是否仍然可用。
6. 将新磁盘添加到存储池，然后删除损坏的磁盘。

► **任务 1: 从连接到服务器的六个磁盘创建存储池**

1. 在 LON-SVR1 上，打开 Server Manager。
2. 在左窗格中，单击 File and Storage Services，然后在 Servers 窗格中，单击 Storage Pools。
3. 使用以下设置创建存储池：
 - 名称：StoragePool1
 - 物理磁盘：前 6 个磁盘。

► **任务 2: 创建三向镜像虚拟磁盘（至少需要五个物理磁盘）**

1. 在 LON-SVR1 上，在服务器管理器，在 VIRTUAL DISKS 窗格中，使用以下设置创建虚拟磁盘：
 - 存储池：StoragePool1
 - 名称：Mirrored Disk
 - 存储布局：Mirror
 - 弹性设置：Three-way mirror
 - 配置类型：Thin
 - 虚拟磁盘大小：10 GB



注意：如果三向弹性设置不可用，请继续执行实验室中的下一步。

2. 在 New Volume Wizard 中，使用以下设置创建卷：
 - 虚拟磁盘：Mirrored Disk
 - 驱动器号：H
 - 文件系统：ReFS
 - 卷标签：Mirrored Volume

► **任务 3: 将文件复制到卷，并验证它在文件资源管理器中可见**

1. 在 LON-SVR 上，打开 Command Prompt。
2. 键入以下命令，然后按 Enter 键：

```
Copy C:\windows\system32\write.exe H:\
```

3. 从任务栏 (taskbar) 打开 File Explorer，然后访问 Mirrored Volume (H:)。您应该在文件列表中看到 write.exe。

► **任务 4: 删除物理驱动器来模拟驱动器故障**

- 在主机计算机上，在 Hyper-V Manager，在 Virtual Machines 窗格中，更改 28740B-LON-SVR1 设置为：
 - 删除以 28740B-LON-SVR1-Disk1 开头的硬盘驱动器。

► **任务 5: 验证文件仍然可以使用**

1. 切换到 LON-SVR1。
2. 打开 File Explorer，然后转到 H:\。

3. 验证 write .exe 是否仍然可用。
4. 在服务器管理器中的 STORAGE POOLS 窗格中, 在菜单栏上, 单击 Refresh "Storage Pools"。



注意： 请注意 Mirrored Disk 旁边显示的警告。

5. 打开 Mirrored Disk Properties 对话框, 然后访问 Health 窗格。



注意： 请注意, 健康状态表示警告。操作状态应指示以下一个或多个: 不完全, 未知或降级 (Degraded)。

6. 关闭 Mirrored Disk Properties 对话框。

► 任务 6: 将新磁盘添加到存储池, 并删除损坏的磁盘

1. 在 LON-SVR1 上, 在服务器管理器中, 在 STORAGE POOLS 窗格中, 在菜单栏上, 单击 Refresh "Storage Pools"。
2. 在 STORAGE POOLS 窗格中, 右键单击 StoragePool1, 单击 Add Physical Disk, 然后在列表中添加第一个磁盘。
3. 要删除断开连接的磁盘, 请打开 Windows PowerShell, 然后运行以下命令:
 - a. Get-PhysicalDisk



注意： 请注意 OperationalStatus 显示为失去通信 (Lost Communication) 磁盘的 FriendlyName。在下一个命令中使用此磁盘名称代替 diskname

- b. \$Disk = Get-PhysicalDisk -FriendlyName 'diskname'
 - c. Remove-PhysicalDisk -PhysicalDisks \$disk -StoragePoolFriendlyName StoragePool1
4. 在服务器管理器中, 刷新存储池视图以查看警告消失

结果： 完成此练习后, 您已成功创建存储池并向其中添加了五个磁盘。此外, 您应该已经从存储池创建了一个三向镜像的, 精简配置的虚拟磁盘。您还应该已经将一个文件复制到新卷, 然后验证它是否可访问。接下来, 在删除物理驱动器之后, 您应该已经验证虚拟磁盘仍然可用, 并且您可以访问它。最后, 您应该已将另一个物理磁盘添加到存储池。

练习 2: 启用并配置存储分层 (Enabling and configuring storage tiering)

场景 (Scenario)

管理层希望您实施存储层以利用多个 SSD 的高性能属性, 同时将较便宜的硬盘驱动器用于访问频率较低的数据。

本练习的主要任务如下:

1. 使用 Get-PhysicalDisk cmdlet 查看系统上的所有可用磁盘。
2. 创建新的存储池。
3. 查看介质类型。
4. 指定样本磁盘的介质类型, 并验证介质类型是否已更改。

5. 使用 Windows PowerShell 创建池级存储层 (pool-level storage tier)。
6. 使用 New Virtual Disk Wizard 创建具有存储分层的新虚拟磁盘。
7. 准备下一个实验。

► **任务 1: 使用 Get-PhysicalDisk cmdlet 查看系统上的所有可用磁盘**

- 在 LON-SVR1, 在 Windows PowerShell (Admin) 中, 运行以下命令：

```
Get-PhysicalDisk
```

► **任务 2: 创建一个新的存储池**

1. 在 Windows PowerShell 中, 运行以下命令：

```
$scanpool = Get-PhysicalDisk -CanPool $true
New-StoragePool -FriendlyName "TieredStoragePool" -StorageSubsystemFriendlyName
"Windows Storage*" -PhysicalDisks $scanpool
```

2. 打开 File Explorer, 然后运行 D:\Labfiles\Mod04\mod4.ps1 脚本。这将配置用于练习下一部分的磁盘名称。

► **任务 3: 查看媒体类型**

- 如果要验证媒体类型, 从 LON-SVR1, 在 Windows PowerShell 中, 运行以下命令：

```
Get-StoragePool -FriendlyName TieredStoragePool | Get-PhysicalDisk | Select
FriendlyName, MediaType, Usage, BusType
```

► **任务 4: 指定样本磁盘的介质类型, 并验证介质类型是否已更改**

1. 要配置介质类型, 在 LON-SVR1 上, 在 Windows PowerShell 中, 运行以下命令：

```
Set-PhysicalDisk -FriendlyName PhysicalDisk1 -MediaType SSD
Set-PhysicalDisk -FriendlyName PhysicalDisk2 -MediaType HDD
```

2. 要验证介质类型, 请运行以下命令：

```
Get-PhysicalDisk | Select FriendlyName, Mediatype, Usage, BusType
```

► **任务 5: 使用 Windows PowerShell 创建池级存储层 (Create pool-level storage tiers by using Windows PowerShell)**

- 要在 LON-SVR1 创建池级存储层, 一个用于 SSD 介质类型, 一个用于 HDD 介质类型, 在 Windows PowerShell 中, 运行以下命令：

```
New-StorageTier -StoragePoolFriendlyName TieredStoragePool -FriendlyName HDD_Tier -
MediaType HDD
New-StorageTier -StoragePoolFriendlyName TieredStoragePool -FriendlyName SSD_Tier -
MediaType SSD
```

► **任务 6: 使用新建虚拟磁盘向导创建具有存储分层的新虚拟磁盘 (Create a new virtual disk with storage tiering by using the New Virtual Disk Wizard)**

1. 在 LON-SVR1 上, 在服务器管理器的 Storage Pools 中, 刷新显示。
2. 在 VIRTUAL DISKS 窗格中, 使用以下设置创建虚拟磁盘：
 - 存储池：TieredStoragePool
 - 名称：TieredVirtDisk

- 存储布局：Simple
 - 配置类型：Fixed
 - 虚拟磁盘大小：4 GB (**每块物理磁盘** 2 GB)
3. 在 New Volume Wizard 中，使用以下设置创建卷：
- 虚拟磁盘：TieredVirtDisk
 - 驱动器号：R
 - 文件系统：ReFS
 - 卷标：Tiered Volume



注意： 如果文件系统下拉菜单中没有 ReFS，请选择 NTFS.

4. 在 TieredVirtDisk 的属性中，观察：
- 存储层 (Storage tiers)
 - 容量
 - 分配的空间
 - 使用的池空间
 - 存储布局 (Storage layout)

► **任务 7: 为下一个实验做准备 (Prepare for the next lab)**

- 完成实验后，请保留为下一个实验室运行的虚拟机

结果： 完成此练习后，您应该已成功启用和配置存储分层。

问题：最少必须向存储池添加多少磁盘才能创建三向镜像虚拟磁盘？

问题：连接到 Windows Server 2012 服务器的 USB 连接磁盘，四个 SAS 磁盘和一个 SATA 磁盘。
您希望为用户提供一个可用于文件存储的卷。你会用什么？

第 3 课 实现重复数据删除 (Implementing Data Deduplication)

重复数据删除是 Windows Server 2016 的角色服务。此服务识别和删除数据中的重复，而不影响数据完整性。这样做是为了实现存储更多数据和使用更少物理磁盘空间的最终目标。本课介绍如何在 Windows Server 2016 存储中实现重复数据删除。

课程目标 (Lesson Objectives)

完成本课后，您将能够：

- 描述 Windows Server 2016 中的重复数据删除。
- 识别 Windows Server 2016 中的重复数据删除组件。
- 解释如何部署重复数据删除。
- 描述重复数据删除的常见使用方案。
- 解释如何监控和维护重复数据删除。
- 使用重复数据删除的备份和还原注意事项。

什么是重复数据删除 (What is Data Deduplication) ?

为了应对企业中的数据存储增长，组织正在整合服务器，并使容量扩展和数据优化成为关键目标。重复数据删除提供了实现这些目标的实用方法，包括：

- 容量优化 (Capacity optimization)。重复数据删除可以在较少的物理空间中存储更多数据。与诸如单实例存储 (SIS) 或 NTFS 压缩的功能相比，它实现了更高的存储效率。重复数据删除使用子文件可变大小的分块和压缩，为通用文件服务器提供 2 : 1 的优化比率，并为虚拟化数据提供高达 20 : 1 的优化比率。
- 规模和性能。数据重复数据删除是高度可扩展，资源高效和非侵入性的。虽然在 Windows Server 2012 R2 中每秒可以处理高达 50 MB，在 Windows Server 2012 中每秒可以处理大约 20 MB 的数据，但是通过重复数据删除处理管道中的进步，Windows Server 2016 的性能将大大提高。在此最新版本的 Windows Server 中，重复数据删除可以通过在多个卷上同时使用多个 I/O 队列并行运行多个线程，而不会影响服务器上的其他工作负载。调整 CPU 对所消耗的服务器工作负载和内存资源造成的影响很小；如果服务器很忙，重复数据删除可以完全停止。此外，您可以随时灵活运行重复数据删除作业，设置运行重复数据删除时的计划以及建立文件选择策略。
- 可靠性和数据完整性。将重复数据删除应用于服务器上的卷时，它会维护数据的完整性。重复数据删除使用校验和结果，一致性和身份验证来确保数据完整性。重复数据删除对于所有的元数据和最常被引用的数据都会维护冗余，以确保数据在数据损坏的情况下被修复，或至少可以恢复。
- BranchCache 的带宽效率。通过与 BranchCache 集成，相同的优化技术应用于通过 WAN 传输到分支机构的数据。其结果是更快的文件下载时间和更少的带宽消耗。

- 重复数据删除：
 - 识别和删除重复的数据而不影响数据的完整性和保真度
 - 具有以更少的空间存储更多数据的目标
- 当您在卷上启用重复数据删除时，一个具有低优先级的后台任务开始运行：
 - 将数据分段成小的，可变大小的区块 (32~128KB)
 - 标识重复的块，并保持每个块的单独副本
 - 使用引用替换冗余副本，并使用指针指向共享块
 - 压缩块



- 使用熟悉的工具优化管理。重复数据删除具有内置于服务器管理器和 Windows PowerShell 中的优化功能。默认设置可以立即提供节省, 或者您可以微调设置以查看更多收益。通过使用 Windows PowerShell cmdlet, 您可以启动优化作业或计划以后运行。安装重复数据删除功能并在所选卷上启用重复数据删除也可以通过使用调用 Windows PowerShell 脚本的 Unattend.xml 文件来实现, 并且可以与 Sysprep 配合使用, 以便在系统首次引导时部署重复数据删除。

重复数据删除过程涉及在数据中查找和删除重复, 而不影响其保真度 (fidelity) 或完整性。目标是通过将文件分割成小的可变大小的块 (chunk, 32-128 KB), 识别重复的块, 并保持每个块的单个副本, 以更少的空间存储更多的数据。

在重复数据删除之后, 文件不再存储为独立的数据流, 并且它们被指向存储在公共块存储中的数据块的存储根替换。因为这些文件共享块, 所以这些块只存储一次, 这减少了存储所有文件所需的磁盘空间。在文件访问期间, 正确的块被透明地组合以提供数据, 应用程序或用户无需具有对文件的磁盘变换的任何知识。这使您可以将重复数据删除应用到文件, 而无需担心应用程序的行为发生任何更改或影响正在访问这些文件的用户。数据重复数据删除在存储场景中效果最好, 大量数据不会频繁修改。

重复数据删除角色服务的增强 (Enhancements to the Data Deduplication role service)

Windows Server 2016 包括对 Windows Server 2012 R2 和 Windows Server 2012 中重复数据删除功能的重要改进, 包括:

- 支持高达 64 TB 的卷大小。Windows Server 2012 R2 中的重复数据删除功能在大小超过 10 TB 的卷 (或更高的数据更改率较高的工作负载) 上表现不佳, 但该功能在 Windows Server 2016 中被重新设计。重复数据删除处理管道现在是多线程的, 能够每卷使用多个 CPU, 以提高高达 64 TB 的卷大小的优化吞吐率。这是 VSS 的限制, 重复数据删除依赖于 VSS。
- 支持高达 1 TB 的文件大小。在 Windows Server 2012 R2 中, 非常大的文件不适合用于重复数据删除。但是, 使用新的流图结构和其他改进来提高优化吞吐量和访问性能, Windows Server 2016 中的重复数据删除在高达 1 TB 的文件上运行良好。
- 简化虚拟化备份应用程序的重复数据删除配置。虽然 Windows Server 2012 R2 支持虚拟化备份应用程序的重复数据删除, 但它需要手动调整重复数据删除设置。但是, 在 Windows Server 2016 中, 为卷启用重复数据删除时, 通过预定义的使用类型选项可大大简化虚拟化备份应用程序的重复数据删除配置。
- 支持 Nano 服务器。Nano 服务器是 Windows Server 2016 中的一个新部署选项, 与 Windows Server 使用 Sever Core 部署选项相比, 它具有更小的系统资源占用, 启动明显更快, 并且需要更少的更新和重新启动。此外, Nano 服务器完全支持数据重复数据删除。
- 支持群集滚动升级。运行重复数据删除的故障转移群集中的 Windows 服务器可以包括运行 Windows Server 2012 R2 的节点和运行 Windows Server 2016 的节点的混合。此主要增强功能在群集滚动升级期间提供对所有已重复数据删除卷的完整数据访问。例如, 您可以逐渐将现有 Windows Server 2012 R2 群集中的每个重复数据删除节点升级到 Windows Server 2016, 而不会导致停机时间, 一次升级所有节点。



注意: 虽然两个 Windows Server 版本的重复数据删除都可以访问优化的数据, 但优化作业仅在 Windows Server 2012 R2 重复数据删除节点上运行, 并且在 Windows Server 2016 重复数据删除节点上阻止运行, 直到群集滚动升级完成。

在 Windows Server 2016 中的重复数据删除功能, 允许您有效地存储, 传输和备份较少的位 (bits)。

重复数据删除对卷的要求 (Volume requirements for Data Deduplication)

安装角色服务后, 可以在每个卷的基础上启用重复数据删除。重复数据删除包括以下要求:

- 卷不能是系统或引导卷。由于操作系统使用的大多数文件是不断打开的，因此系统卷上的重复数据删除会对性能产生负面影响，因为在使用文件之前，需要重新展开已删除重复数据的数据。
- 卷可能通过使用主引导记录（MBR）或 GUID 分区表（GPT）格式进行分区，并且必须使用 NTFS 或 ReFS 文件系统进行格式化。
- 卷必须连接到 Windows Server，并且不能显示为不可移动驱动器。这意味着您不能使用 USB 或软盘驱动器进行重复数据删除，也不能使用远程映射驱动器。
- 卷可以位于共享存储上，例如光纤通道，iSCSI SAN 或 SAS 阵列。
- 对于重复数据删除，不会处理具有扩展属性的文件，加密文件，小于 32 KB 的文件以及重新解析点文件（reparse point files）。
- 重复数据删除不适用于 Windows 客户端操作系统。

重复数据删除组件（Data Deduplication components）

重复数据删除角色服务由几个组件组成。这些组件包括：

- 过滤器驱动程序（Filter driver）。此组件监视本地或远程 I/O，并通过与各种作业交互来处理文件系统上的数据块。每个卷都有一个过滤器驱动程序。
- 重复数据删除服务（Deduplication service）。此组件管理以下作业类型：
 - 优化（Optimization）。由多个作业组成，它们根据卷的重复数据删除策略执行文件的重复数据删除和压缩。在对文件进行初始优化之后，如果该文件随后被修改并且满足用于优化的重复数据删除策略阈值，则文件将被再次优化。
 - 垃圾收集（Garbage Collection）。数据重复数据删除包括垃圾回收作业，以处理卷上已删除或已修改的数据，以便清除不再引用的任何数据块。此作业处理先前已删除或逻辑覆盖的优化内容以创建可用卷可用空间。当优化文件被新数据删除或覆盖时，块存储中的旧数据不会立即删除。虽然垃圾回收计划为每周运行，但您可能会考虑在发生大量删除后才运行垃圾回收。
 - 清理（Scrubbing）。重复数据删除具有内置的数据完整性功能，例如校验和验证和元数据一致性检查。它还有关键元数据和最常用的数据块内置冗余。当访问数据或重复数据删除作业处理数据时，如果这些功能遇到损坏，则会在日志文件中记录损坏。清理作业使用这些功能来分析块存储损坏日志，并在可能的情况下进行修复。可能的修复操作包括使用三个冗余数据源：
 - 重复数据删除保留常用块的备份副本，当它们在称为热点的区域中引用了超过100次时。如果工作副本已损坏，重复数据删除将使用其冗余副本，例如软件损坏，例如位翻转（bit flips）或撕毁写入（torn writes）。
 - 如果使用镜像存储空间，重复数据删除可以使用冗余块的镜像来服务 I/O 并修复损坏。
 - 如果文件是用损坏的块处理的，则损坏的块被删除，并且新的传入块用于修复损坏。

- 重复数据删除功能由几个组件组成：
 - 一个过滤驱动程序，用于监视本地或远程 I/O
 - 重复数据删除角色服务，控制三种可用的作业类型：
 - 优化
 - 垃圾收集
 - 清理



注意：由于内置于重复数据删除中的附加验证，重复数据删除系统通常是第一个报告硬件或文件系统中数据损坏的早期迹象的系统。

- 未优化 (Unoptimization)。此作业将在卷上所有优化文件上撤消重复数据删除。使用此类作业的一些常见方案包括：停用具有启用了重复数据删除功能的卷的服务器，故障诊断重复数据删除数据的问题或将数据迁移到不支持重复数据删除的另一个系统。在开始此作业之前，应该使用 Disable-DedupVolume Windows PowerShell cmdlet 在一个或多个卷上禁用进一步的重复数据删除活动。禁用重复数据删除后，卷将保持重复数据删除状态，并且现有重复数据删除数据仍可访问；但是，服务器停止为该卷运行优化作业，并且不会对新数据进行重复数据删除。之后，您将使用未优化作业来撤消卷上的现有重复数据删除数据。在成功的未优化作业结束时，将从卷中删除所有重复数据删除元数据。



注意：在使用取消优化作业时，应该谨慎，因为所有重复数据删除的数据将返回原始逻辑文件大小。因此，您应验证该卷是否有足够的可用空间用于此活动，或者移动/删除某些数据以允许作业成功完成。

重复数据删除流程 (Data Deduplication process)

在 Windows Server 2016 中，重复数据删除可以透明地删除重复，而不更改访问语义。在卷上启用重复数据删除后，重复数据删除被用于来优化卷上的文件数据，它执行以下操作：

- 优化作业（即后台任务）在服务器上以低优先级运行，以处理卷上的文件。
- 通过使用算法，将卷上的所有文件数据分成从 32 KB 到 128 KB 不等的小型可变大小的块。
- 标识卷上具有一个或多个副本的块。
- 将块插入到公共块存储中。
- 将所有重复的块替换为块存储中的块的单个副本的引用或存根。
- 使用重新分析点 (reparse points) 替换原始文件，其中包含对其数据块的引用。
- 压缩块并将其组织在 System Volume Information 文件夹中的容器文件中。
- 删除文件的主数据流。

重复数据删除过程通过本地服务器上的计划任务工作，但您可以通过使用 Windows PowerShell 以交互方式运行该过程。有关这方面的更多信息将在后面的模块中讨论。

重复数据删除不会产生任何写入性能影响，因为在写入文件时不会对数据进行重复数据删除。Windows Server 2016 使用后处理重复数据删除，确保重复数据删除潜力最大化。这种重复数据删除过程的另一个优点是，您的应用程序服务器和客户端计算机卸载所有处理，这意味着对您环境中的其他资源的压力较小。但是，读取重复数据删除的文件时，性能会受到影响。



注意：重复数据删除的三种主要类型是源，目标（或后处理重复数据删除，post-process deduplication）和内联（或传输重复数据删除，transit deduplication）。

重复数据删除可能会处理所选卷上的所有数据，除了大小小于 32 KB 的文件和排除的文件夹中的文件。在启用此功能之前，必须仔细确定服务器及其连接的卷是否适合进行重复数据删除。您还应该考虑在重复数据删除过程中定期备份重要数据。

在启用重复数据删除卷并优化数据后，该卷包含以下元素：

- 未优化的文件 (Unoptimized files)。包括不符合所选文件时间策略设置，系统状态文件，备用数据流，加密文件，扩展属性文件，小于 32 KB 的文件或其他重新分析文件的文件。
- 已优化的文件 (Optimized files)。包括存储为重新分析点 (reparse points) 的文件，其中包含指向块存储中的相应块的映射的指针，这些块在请求时用于恢复文件。

- 块存储 (Chunk store)。优化文件数据的位置。
- 额外的可用空间。优化的文件和块存储空间比优化之前占用的空间要少得多。

部署重复数据删除 (Deploying Data Deduplication)

规划重复数据删除 (Planning a Data Deduplication deployment)

在环境中安装和配置重复数据删除之前，必须使用以下步骤规划部署：

- 目标部署。重复数据删除旨在应用于主数据卷，而不是逻辑扩展数据卷，而无需添加任何额外的专用硬件。您可以根据所涉及的数据类型以及卷或特定文件类型发生的更改的频率和数量来安排重复数据删除。您应考虑对以下数据类型使用重复数据删除：
 - 常规文件共享。组内容发布和共享，用户主文件夹和文件夹重定向/脱机文件。
 - 软件部署共享。软件二进制，图像和更新。
 - VHD 库。用于配置到管理程序的虚拟硬盘 (VHD) 文件存储。
 - VDI 部署。使用 Hyper-V 的虚拟桌面基础结构 (VDI) 部署。
 - 虚拟化备份。作为 Hyper-V 客户端运行的备份应用程序将备份数据保存到挂接的 VHD。
- 确定哪些卷是重复数据删除的候选。重复数据删除对于优化存储和减少磁盘空间消耗非常有效，在应用于正确的数据时，可节省 50% 至 90% 的系统存储空间。使用以下注意事项来评估哪些卷是重复数据删除的理想选择：
 - 是否存在重复数据？

托管用户文档，软件部署二进制文件或虚拟硬盘文件的文件共享或服务往往有大量的重复，并且从重复数据删除中可以节省更多的存储空间。有关重复数据删除的部署候选和受支持/不支持的方案的信息将在本单元稍后讨论。
 - 数据访问模式是否有足够的时间进行重复数据删除？

例如，频繁更改并且经常由用户或应用程序访问的文件不是重复数据删除的好候选。在这些情况下，重复数据删除可能无法处理文件，因为对数据的不断访问和更改可能会取消重复数据删除所产生的任何优化增益。另一方面，好的候选者可以节省重复数据删除文件的时间。
 - 服务器是否有足够的资源和时间运行重复数据删除？

重复数据删除需要读取，处理和写入大量数据，从而消耗服务器资源。当资源利用率低时，服务器通常具有高活动时间和时间；当资源可用时，重复数据删除作业更有效。但是，如果服务器始终处于最大资源容量，则可能不是重复数据删除的理想选择。
- 使用重复数据删除评估工具评估节约的成本。您可以使用重复数据删除评估工具 DDPEval.exe 来确定在特定卷上启用重复数据删除时可获得的预期节省。DDPEval.exe 支持评估本地驱动器和映射或未映射的远程共享。

在环境中安装和配置重复数据删除之前，您需要使用以下步骤规划部署：

- 目标部署
- 确定哪些卷是重复数据删除的候选
- 使用重复数据删除评估工具评估节约额
- 规划部署，可伸缩性和重复数据删除策略



注意：安装重复数据删除功能时，重复数据删除评估工具 (DDPEval.exe) 会自动安装到

\\Windows\\System32\\ 目录。

有关详细信息, 请参阅“计划部署重复数据删除”, 网址为: <http://aka.ms/sxzd2l>

- 规划部署, 可伸缩性和重复数据删除策略。默认重复数据删除策略设置通常足以适用于大多数环境。但是, 如果您的部署具有以下任一条件, 您可以考虑更改默认设置:
 - 传入数据是静态的或预期是只读的, 并且您希望更早地处理卷上的文件。在此方案中, 将 MinimumFileAgeDays 设置更改为较少的天数, 以便较早地处理文件。
 - 您有不希望重复数据删除的目录。将目录添加到排除列表。
 - 您有不希望重复数据删除的文件类型。将文件类型添加到排除列表。
 - 服务器具有与默认值不同的非高峰时间, 并且您想要更改垃圾收集和清除时间表。使用 Windows PowerShell 更新调度时间。

安装和配置数据重复删除 (Installing and configuring Data Deduplication)

完成规划后, 需要使用以下步骤将重复数据删除部署到您环境中的服务器:

- 在服务器上安装重复数据删除组件。使用以下选项在服务器上安装重复数据删除组件:
 - *Server Manager*。在服务器管理器中, 您可以导航到 Add Roles and Features Wizard > 在 Server Roles 下 > 选择 File and Storage Services > 选择 File Services 复选框 > 选择 Data Deduplication 复选框 > 单击 Install, 以安装重复数据删除。
 - *Windows PowerShell*。您可以使用以下命令安装重复数据删除:

```
Import-Module ServerManager
Add-WindowsFeature -Name FS-Data-Deduplication
Import-Module Deduplication
```

- 启用重复数据删除。使用以下选项在服务器上启用重复数据删除:
 - *Server Manager*。从服务器管理器仪表板:
 - i. 右键单击数据卷, 然后选择 Configure Data Deduplication。
 - ii. 在 Data deduplication 框中, 选择要在卷上承载的工作负载。例如, 为运行的虚拟机配置存储时, 请为通用数据文件或 Virtual Desktop Infrastructure (VDI) server 选择 General purpose file server。
 - iii. 输入从文件删除重复数据之前的文件创建日期开始的最少天数, 输入不应删除重复数据的任何文件类型的扩展名, 然后单击 Add 以浏览到不应删除重复数据的文件的任何文件夹。
 - iv. 单击 Apply 以应用这些设置并返回到服务器管理器仪表板, 或单击 Set Deduplication Schedule 按钮继续设置重复数据删除的计划。
 - *Windows PowerShell*。使用以下命令在卷上启用重复数据删除:

```
Enable-DedupVolume -Volume VolumeLetter -UsageType StorageType
```

 **注意:** 使用卷的驱动器号替换 *VolumeLetter*。使用对应于卷的工作负载的预期类型的值替换 *StorageType*。可接受的值包括:

- *HyperV*。Hyper-V 存储的卷。
- *备份*。为虚拟化备份服务器优化的卷。
- *默认*。通用卷。

或者，您可以使用 Windows PowerShell cmdlet `Set-DedupVolume` 来配置其他选项，例如从文件删除重复数据之前的文件创建日期开始的最少天数，不应进行重复数据删除的任何文件类型的扩展名，或应该从重复数据删除中排除的文件夹。

- 配置重复数据删除作业。使用重复数据删除作业，您可以手动运行，按需运行或使用计划。以下列表是您可以对卷执行的作业类型：

- 优化。包括内置作业，它会自动调度以定期优化卷。优化作业根据策略设置对卷上的数据和压缩文件块进行重复数据删除。您还可以使用以下命令按需触发优化作业：

```
Start-DedupJob -Volume VolumeLetter -Type Optimization
```

- 数据清理。自动计划清理作业以按周分析卷，并在 Windows 事件日志中生成摘要报告。您还可以使用以下命令按需触发清理作业：

```
Start-DedupJob -Volume VolumeLetter -Type Scrubbing
```

- 垃圾收集。自动计划垃圾收集作业以每周处理卷上的数据。因为垃圾回收是一个处理密集型操作，您可以考虑等待，直到删除负载达到阈值，才能按需运行此作业或在几小时后计划作业。您还可以使用以下命令按需触发垃圾收集作业：

```
Start-DedupJob -Volume VolumeLetter -Type GarbageCollection
```

- 未优化。未优化作业可根据需要提供，不会自动计划。但是，您可以使用以下命令按需触发未优化作业：

```
Start-DedupJob -Volume VolumeLetter -Type Unoptimization
```



注意：有关详细信息，请参阅“Set-DedupVolume”，网址为：<http://aka.ms/o30xqw>

- 配置重复数据删除计划。在服务器上启用重复数据删除时，默认情况下启用三个计划：优化计划每小时运行一次，垃圾收集计划和清理计划每周运行一次。您可以使用此 Windows PowerShell cmdlet `Get-DedupSchedule` 查看计划。这些计划作业在服务器上的所有卷上运行。但是，如果要仅在特定卷上运行作业，则必须创建新作业。您可以从服务器管理器中的 Deduplication Settings 页面创建，修改或删除作业计划，或使用 Windows PowerShell cmdlet：`New-DedupSchedule`，`Set-DedupSchedule` 或 `Remove-DedupSchedule`。



注意：重复数据删除作业最多只支持每周作业调度。如果需要为每月作业或任何其他自定义时间段创建计划，请使用 Windows 任务计划程序。但是，您将无法使用 Windows PowerShell cmdlet `Get-DedupSchedule` 查看使用 Windows 任务计划程序创建的这些自定义作业计划。

演示：实现数据重复删除 (Implementing Data Deduplication)

在本演示中，您将了解如何：

- 安装重复数据删除角色服务。
- 启用重复数据删除。
- 检查重复数据删除的状态。

海量视频题库 myitpub.com QQ:5565462

演示步骤 (Demonstration Steps)

安装重复数据删除角色服务 (Install the Data Deduplication role service)

- 在 LON-SVR1 上, 在服务器管理器中, 添加重复数据删除角色服务。

启用数据重复删除 (Enable Data Deduplication)

1. 打开 File Explorer, 并观察可用的卷和可用空间。
2. 返回到 File and Storage Services。
3. 单击 Disks。
4. 单击 1 磁盘, 然后单击 D 卷。
5. 启用重复数据删除, 然后单击 General purpose file server 设置。
6. 配置以下设置 :
 - A. 删除超过 (以天为单位) 的文件 : 1
 - B. 启用吞吐量优化 (Enable throughput optimization)
 - C. 排除 : D:\shares

检查重复数据删除状态 (Check the status of Data Deduplication)

1. 切换到 Windows PowerShell。
2. 执行以下命令以验证重复数据删除状态 :

```
a. Get-DedupStatus
b. Get-DedupStatus | fl
c. Get-DedupVolume
d. Get-DedupVolume | fl
e. Start-DedupJob D: -Type Optimization -Memory 50
```

3. 重复命令 2a 和 2c。



注意 : 因为大多数驱动器 D 上的文件很小, 你可能不会注意到大量的节省空间。

4. 关闭所有打开的窗口。

重复数据删除的使用场景 (Usage scenarios for Data Deduplication)

下表突出显示了各种内容类型的典型重复数据删除节省。您的数据存储节省量将因数据类型, 数据混合以及卷的大小和卷包含的文件而异。在启用重复数据删除之前, 应考虑使用重复数据删除评估工具来评估卷。

- 用户文档。这包括组内容发布或共享, 用户主文件夹 (或 MyDocs) 和用于访问脱机文件的配置文件重定向。将重复数据删除应用到这些共享可能会节省高达系统存储空间的 30% 到 50%。
- 软件部署共享。这包括软件二进制文件, cab 文件, 符号文件, 图像和更新。将重复数据删除



除应用于这些共享可能能够节省高达系统存储空间的 70% 到 80%。

- 虚拟化库。这包括用于提供给虚拟机管理程序 (hypervisors) 的虚拟硬盘文件 (即 .vhd 和 .vhdx 文件) 存储。将重复数据删除应用到这些库可能能够节省高达系统存储空间的 80% 到 95%。
- 常规文件共享。这包括上面标识的所有类型的数据的混合。将重复数据删除应用到这些共享可能会节省高达 50% 到 60% 的系统存储空间。

重复数据删除部署候选 (Data Deduplication deployment candidates)

根据 Windows Server 2016 中观察到的节省和典型资源使用情况, 重复数据删除的部署备份排名如下:

- **重复数据删除的理想候选**
 - 文件夹重定向服务器
 - 虚拟化库 (Virtualization depot) 或配置库 (provisioning library)
 - 软件部署共享
 - SQL Server 和 Exchange Server 备份卷
 - 横向扩展文件服务器 (SoFS) CSV
 - 虚拟化备份 VHD (例如, DPM)
 - VDI VHD (仅限个人 VDI)



注意: 在大多数 VDI 部署中, 引导风暴需要进行专门的规划, 这是给大量用户尝试同时登录到 VDI (通常在早上到达工作时) 的现象。反过来, 这会损坏 VDI 存储系统, 并可能导致 VDI 用户长时间的延迟。但是, 在 Windows Server 2016 中, 当在虚拟机启动期间从磁盘重复数据删除存储中读取数据块时, 它们会缓存在内存中。因此, 后续的读取不需要频繁地访问块存储, 因为高速缓存拦截它们; 引导风暴的影响最小化, 因为内存比磁盘快得多。

- **应根据内容进行评估**
 - 业务线服务器
 - 静态内容提供者
 - Web 服务器
 - 高性能计算 (HPC)
- **不是理想的重复数据删除的候选**
 - Hyper-V 主机
 - WSUS
 - SQL Server 和 Exchange Server 数据库卷

重复数据删除互操作性 (Data Deduplication interoperability)

在 Windows Server 2016 中, 在部署重复数据删除时应考虑以下相关技术和潜在问题:

- BranchCache。通过在 Windows 服务器和客户端上启用 BranchCache 可以优化通过网络访问数据。当启用了 BranchCache 的系统通过 WAN 与启用了重复数据删除的远程文件服务器进行通信时, 所有已重复数据删除的文件都已建立索引和散列, 因此可快速计算来自分支机构的数据请求。这类似于预先索引或预先散列启用 BranchCache 的服务器。



注意：BranchCache 是当用户从分支机构位置访问中心办公室中的内容时可以减少广域网（WAN）利用率并增强网络应用响应性的功能。启用 BranchCache 时，从 Web 服务器或文件服务器检索的内容的副本将缓存在分支办公室内。如果分支中的另一客户端请求相同的内容，则客户端可以直接从本地分支网络下载它，而不需要通过使用 WAN 来检索内容。

- **故障转移群集。** Windows Server 2016 完全支持故障转移群集，这意味着重复数据删除的卷将在群集中的节点之间进行正常故障转移。有效地，去重的卷是自包含和便携式单元（即，卷包含的所有数据和配置信息），但要求群集中访问重复数据删除的卷的每个节点都必须运行重复数据删除功能。当群集形成时，在群集中配置重复数据删除调度信息。因此，如果重复数据删除的卷由另一个节点接管，则调度的作业将由新节点在下一个调度的间隔上应用。
- **FSRM 配额。** 虽然您不应在启用重复数据删除的卷根文件夹上创建硬配额，但是可以使用文件服务器资源管理器（FSRM）在启用重复数据删除的卷根上创建软配额。当 FSRM 遇到重复数据删除的文件时，它将标识文件的配额计算的逻辑大小。因此，当重复数据删除处理文件时，配额使用（包括任何配额阈值）不会更改。所有其他 FSRM 配额功能（包括子根文件夹的卷根软配额和配额）将在使用重复数据删除时按预期工作。



注意：文件服务器资源管理器（FSRM）是一套用于 Windows Server 2016 的工具，允许您识别，控制和管理存储在服务器上的数据的数据的数量和类型。FSRM 使您能够在文件夹和卷上配置硬配额或软配额。硬配额可防止用户在达到配额限制后保存文件；而软配额不强制配额限制，但是当卷上的数据达到阈值时生成通知。在启用重复数据删除的卷根文件夹上启用硬配额时，卷上的实际可用空间和卷上的配额限制空间不相同；这可能会导致重复数据删除优化作业失败。

- **DFS 复制。** 重复数据删除与分布式文件系统（DFS）复制兼容。优化或未优化文件不会触发复制，因为文件不会更改。DFS 复制使用远程差分压缩（RDC），而不是块存储中的块，用于线上节省。实际上，如果为重复数据删除启用了副本，则可以通过使用重复数据删除来优化副本实例上的文件。



注意：单实例存储（Single Instance Storage，SIS）是用于 NTFS 文件重复数据删除的文件系统过滤器驱动程序，在 Windows Server 2012 R2 中已弃用，并在 Windows Server 2016 中完全删除。

监控和维护重复数据删除 (Monitoring and maintaining Data Deduplication)

在您的环境中部署重复数据删除后，重要的是要监控和维护启用了重复数据删除和相应数据存储的系统，以确保最佳性能。虽然 Windows Server 2016 中的重复数据删除包括大量自动化（包括优化作业），但重复数据删除过程需要您验证优化的效率；对系统，存储体系结构和卷进行适当的调整；并解决重复数据删除的任何问题。

重复数据删除监控和报表分析 (Monitoring and reporting of Data Deduplication)

在规划您的环境中的重复数据删除时，您将不可避免地问自己，“我的已配置的重复数据删除卷的大小应该是多少？”，虽然 Windows Server 2016 在高达 64 TB 的卷上支持重复数据删除，但必须评估您的环境可以支持的重复数据删除卷的大小。对于许多人来说，这个问题的答案是它取决于您的硬件规格和您的独特工作负载。更具体地，它主要取决于数据量的变化和磁盘存储子系统的数据访问吞吐率的频率。

在您的环境中监视重复数据删除的效率在部署的每个阶段都很有帮助，尤其是在您的规划阶段。如前面模块中所述，Windows Server 2016 中的重复数据删除执行密集的 I/O 和计算操作。在大多数部署中，重复数据删除在后台或每日计划中运行，每天都有新数据或修改数据（即数据更改），只要重复数据删除能够每天优化所有数据更改，则卷大小将适用于重复数据删除。另一方面，一些组织只需创建一个 64 TB 卷，就可以进行重复数据删除，然后想知道为什么它们会遇到低优化率。在这种情况下，最有可能的情况是，重复数据删除无法跟上来自配置卷上过大的数据集的传入流量。虽然 Windows Server 2016 中的重复数据删除在多个卷上同时使用多个 I/O 队列并行运行多个线程，但重复数据删除环境可能需要额外的计算能力。

在估算启用重复数据删除的卷大小时，应考虑以下因素：

- 重复数据删除优化必须能够跟上每日数据更改。
- 数据被的总量与数据量的大小成比例。
- 重复数据删除优化的速度显著取决于磁盘存储子系统的数据访问吞吐率。

因此，要估算已进行重复数据删除的卷的最大容量，您应该熟悉数据流的大小和卷的优化处理速度。您可以选择使用参考数据，如服务器硬件规格，存储驱动器/阵列速度和各种使用类型的重复数据删除速度，以供您进行估计。然而，评估适当卷大小的最准确方法是基于您的数据的代表性样本（如数据更改和重复数据删除处理速度）直接对重复数据删除系统执行测量。

您应该考虑使用以下选项来监视环境中的重复数据删除，并报告其运行状况：

- Windows PowerShell cmdlet。在服务器上启用重复数据删除功能后，您可以使用以下 Windows PowerShell cmdlet：
 - Get-DedupStatus。最常用的 cmdlet，此 cmdlet 返回具有重复数据删除元数据的卷的重复数据删除状态，包括重复数据删除率，优化文件的数量/大小，重复数据删除作业的上次运行时间以及卷上节省的空间量。
 - Get-DedupVolume。此 cmdlet 返回具有重复数据删除元数据的卷的重复数据删除状态。元数据包括重复数据删除率，优化文件的数量/大小以及重复数据删除设置，例如最小文件年龄，最小文件大小，排除文件/文件夹，压缩排除文件类型和块冗余阈值。
 - Get-DedupMetadata。此 cmdlet 返回具有重复数据删除元数据的卷的重复数据删除数据存储的状态信息，其中包括以下数目：
 - 容器中的数据块。

- 使用以下命令监视重复数据删除：
 - Windows PowerShell cmdlet
 - 事件查看器日志
 - 性能监视器数据
 - 文件资源管理器
- 使用 Windows PowerShell cmdlet 维护重复数据删除
- 准备好解决重复数据删除问题

- 数据存储中的容器。
- 容器中的数据流。
- 流映射存储中的容器。
- 容器中的热点。
- 流地图存储中的热点。
- 卷上的损坏。
- Get-DedupJob。此 cmdlet 返回当前运行或排队的重复数据删除作业的重复数据删除状态和信息。

一个常见的情况是评估重复数据删除是否跟上传入数据的速率。您可以使用 Get-DedupStatus cmdlet 来监视优化文件数量与策略内文件数量的比较。这使您能够查看是否处理所有策略内文件。如果策略内文件数量持续上升速度高于优化文件数量, 则应检查硬件规格以了解适当的利用率或卷使用类型上的数据类型, 以确保重复数据删除效率。但是, 如果来自 LastOptimizationResult cmdlet 的输出值为 0x00000000, 则在上一个优化作业期间, 整个数据集已被成功处理。



注意：有关详细信息, 请参阅: “Windows PowerShell 中的存储 Cmdlet”, 网址为:

<http://aka.ms/po9qve>

- 事件查看器日志。监视事件日志还有助于了解重复数据删除事件和状态。要查看重复数据删除事件, 请在事件查看器中导航到 Applications and Services Logs, 单击 Microsoft, 单击 Windows, 然后单击 Deduplication。例如, 事件 ID 6153 将为您提供重复数据删除作业的耗用时间和吞吐率。
- 性能监视器数据。除了使用计数器来监视服务器性能 (如 CPU 和内存) 之外, 还可以使用典型的磁盘计数器来监视当前正在运行的作业的吞吐率, 例如: Disk Read Bytes/sec, Disk Write Bytes/sec, 和 Average Disk sec/Transfer。根据服务器上的其他活动, 您可以使用这些计数器的数据结果, 通过检查正在读取的数据量和每个时间间隔写入的数量来大致估计保存率。您还可以使用资源监视器来识别特定程序/服务的资源使用情况。要查看磁盘活动, 请在 Windows 资源监视器中过滤进程列表以找到 fsdmhost.exe, 并检查 Disk 选项卡下文件的 I/O。



注意：Fsdmhost.exe 是 Microsoft 文件服务器数据管理主机进程的可执行文件, 由 Windows Server 2016 中的重复数据删除进程使用。

- 文件资源管理器 (File Explorer)。虽然不是在整个卷上验证重复数据删除的理想选择, 但您可以使用文件资源管理器找到检查单个文件的重复数据删除。在查看文件的属性时, 您会注意到 Size 显示文件的逻辑大小, 而 Size on Disk 显示文件的真实物理分配。对于优化的文件, Size on Disk 小于实际文件大小。这是因为重复数据删除将文件的内容移动到公共块存储, 并用 NTFS 重分析点存根和元数据替换原始文件。

维护重复数据删除 (Maintaining Data Deduplication)

通过监视收集的数据, 您可以使用以下 Windows PowerShell cmdlet, 以确保在您的环境中实现重复数据删除的最佳效率。

- Update-DedupStatus。某些存储 cmdlet (如 Get-DedupStatus 和 Get-DedupVolume) 从缓存的元数据检索信息。此 cmdlet 扫描卷以计算新的重复数据删除信息以更新元数据。
- Start-DedupJob。此 cmdlet 用于启动即席重复数据删除作业, 例如优化, 垃圾回收, 清理和未优化。例如, 如果重复数据删除的卷由于额外的修改而导致可用空间不足, 则可以考虑启动特别优化作业。

- **Measure-DedupFileMetadata**。此 cmdlet 用于测量卷上的潜在磁盘空间。更具体地说，如果删除一组文件夹并随后运行垃圾收集作业，则此 cmdlet 返回可在卷上回收的磁盘空间量。文件通常具有在其他文件夹中共享的块。重复数据删除引擎计算哪些块是唯一的，并且将在垃圾收集作业之后删除。
- **Expand-DedupFile**。此 cmdlet 将优化的文件扩展至其原始位置。由于与应用程序或其他需求的兼容性，您可能需要扩展优化的文件。确保卷上有足够的空间来存储扩展文件。

重复数据删除负面影响的故障诊断 (Troubleshooting adverse effects of Data Deduplication)

当 Windows Server 2016 中的重复数据删除对应用程序或对文件的访问产生负面影响时，有几个选项可用，包括：

- 通过更改计划或选择手动运行重复数据删除作业，可以使用不同的重复数据删除频率。
- 使用作业选项，例如：
 - **StopWhenSystemBusy**，如果作业干扰服务器的工作负载，则停止重复数据删除。
 - **Preempt**，导致重复数据删除引擎将特定的重复数据删除作业移动到作业队列的顶部，并取消当前作业。
 - **ThrottleLimit**，它设置可以由特定的重复数据删除作业建立的并发操作的最大数量。
 - **Priority**，设置特定重复数据删除作业的 CPU 和 I/O 优先级。
 - **Memory**，指定重复数据删除作业可以使用的物理计算机内存的最大百分比。



注意：建议允许重复数据删除自动管理内存分配，但在某些情况下可能需要调整最大百分比。对于大多数这些情况，您应该考虑在 15 到 50 范围内的最大百分比，以及在您指定 **StopWhenSystemBusy** 参数时计划运行的作业的较高内存消耗。对于通常计划在工作时间之后运行的垃圾回收和清理重复数据删除作业，您可以考虑使用更高的内存消耗，例如 50。

- 如果需要兼容性或性能，请使用 **Expand-DedupFile** cmdlet 展开或取消删除特定文件。
- 使用具有 **Unoptimization** 作业类型的 **Start-DedupJob** cmdlet 禁用卷上的重复数据删除。

重复数据删除损坏的故障诊断 (Troubleshooting Data Deduplication corruptions)

Windows Server 2016 中的重复数据删除提供了检测，报告甚至修复数据损坏的功能。实际上，通过重复数据删除，数据完整性被认为是非常重要的，因为大量的重复数据删除的文件可能引用单个常用的块，这些块被损坏。虽然重复数据删除中内置了许多功能以帮助防止损坏，但仍有一些情况下重复数据删除可能无法自动从损坏中恢复。



附加阅读：有关详细信息，请参阅：“排除重复数据删除损坏故障”，网址为：
<http://aka.ms/Tdz13m>

重复数据删除报告损坏的一些最常见原因是：

- 复制数据时使用的不兼容的 **Robocopy** 选项。使用 **Robocopy** 与卷根上的 **/MIR** 选项作为目标擦除重复数据删除存储。要避免此问题，请使用 **/XD** 选项将 **System Volume Information** 从 **Robocopy** 命令的作用域中排除。



注意：有关详细信息，请参阅：“当在 Windows Server 2012 中使用 **Robocopy/MIR** 时，**FARM** 和重复数据删除可能会受到不利影响”，网址为：
<http://aka.ms/W0ux7m>

海量视频题库 myipub.com QQ:5565462

- 在已消除重复数据的卷上使用的不兼容的备份/恢复程序。您应该验证您的备份解决方案是否支持 Windows Server 2016 中的重复数据删除, 因为不支持的备份解决方案可能会在恢复后引入损坏。有关这方面的更多信息将在本单元的后面部分。
- 将重复数据删除的卷迁移到低级 Windows Server 版本。文件损坏消息可能会从安装在旧版 Windows Server 上的已重复数据删除的卷中访问的文件中报告, 但已在较新版本的操作系统上进行了优化。在这种情况下, 您应该验证访问重复数据删除数据的服务器的版本是否与优化卷上的数据的服务器版本相同。尽管重复数据删除的卷可以重新安装在不同的服务器上, 但是重复数据删除是向后兼容的, 但不能向前兼容; 您可以升级并迁移到较新版本的 Windows Server, 但是较早版本的 Windows Server 不能读取由较新版本的 Windows Server 重复数据删除的数据, 并且可能会在尝试读取时将数据报告为损坏。
- 对已启用重复数据删除的卷根 (root of a volume) 启用压缩。在根启用了压缩的卷上不支持重复数据删除。因此, 这可能会导致重复数据删除文件的损坏和无法访问。



注意：在 Windows Server 2016 中支持压缩文件夹中文件的重复数据删除，并且应正常运行。

- **硬件问题。** 通过使用重复数据删除清理作业，可以及早检测到许多硬件存储问题。有关详细信息，请参阅下面的常规损坏故障诊断步骤。
- **一般损坏。** 您可以使用以下步骤排除重复数据删除报告损坏的大多数常见原因：
 - a. **检查事件日志以了解损坏的详细信息。** 检查重复数据删除清理事件日志中是否有早期文件损坏的情况，以及由清理作业尝试的损坏修复。重复数据删除检测到的任何损坏都会记录到事件日志中。清理通道列出检测到的任何损坏和由作业尝试修复的文件。重复数据删除擦除事件日志位于事件查看器中（位于 Application and Services > Microsoft > Windows > Deduplication > Scrubbing）。此外，在系统事件日志和存储空间事件日志中搜索硬件事件通常会产生有关硬件问题的其他信息。



注意：重复数据删除事件日志中潜在的大量事件可能难以通过事件查看器进行解析。一个公开可用的脚本，它生成一个易于阅读的 HTML，突出显示检测到的损坏情况，以及清理作业中任何尝试的损坏修复的结果。

有关详细信息，请参阅：“生成重复数据删除清理报告”，网址为：<http://aka.ms/N75aww>

- b. 在只读模式下运行 CHKDSK。虽然此命令可以修复卷上的一些数据损坏，但在没有任何参数的情况下运行命令将启动只读扫描。



附加阅读：有关详细信息，请参阅：“CHKDSK”，网址为：<http://aka.ms/Nep9wf>

- c. 运行 deep Scrubbing 作业以修复检测到的损坏。必须进行损坏调查，应使用深度清理作业，以确保所有损坏都记录在事件日志中的重复数据删除清理通道中。清理事件将提供损坏的细分，包括损坏的块，受影响的文件，损坏的确切容器偏移量以及受影响文件的列表（最多 10 K 文件）。

您可以在 Windows PowerShell 中使用以下命令启动深度清理作业：


```
Start-DedupJob VolumeLetter -Type Scrubbing -Full
```

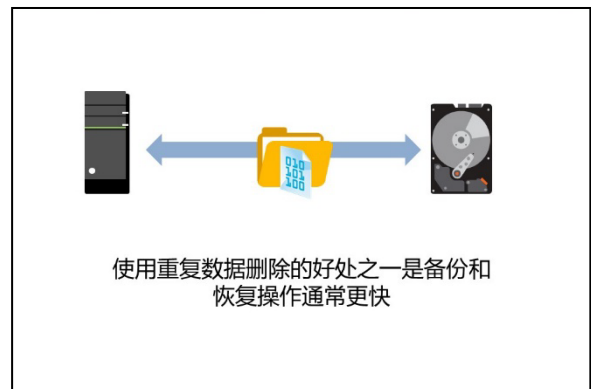


注意：使用卷的驱动器号替换 VolumeLetter。

使用重复数据删除的备份和还原注意事项 (Backup and restore considerations with Data Deduplication)

使用重复数据删除的优点之一是备份和恢复操作更快。这是因为您减少了卷上使用的空间, 这意味着要备份的数据较少。执行优化备份时, 备份也较小。这是因为优化文件, 非优化文件和重复数据删除块存储文件的总大小远小于卷的逻辑大小。

 **注意：**许多基于块的备份系统应该与重复数据删除一起使用, 从而在备份介质上保持优化。不使用重复数据删除的基于文件的备份操作通常会以原始格式复制文件。



Windows Server 2016 中的重复数据删除支持以下备份和还原方案：


- 单个文件备份/恢复
- 完全卷备份/恢复
- 使用 VSS 编写器 (VSS writer) 优化的文件级备份/恢复

另一方面, Windows Server 2016 中的重复数据删除不支持以下备份和还原方案：

- 仅对重分析点 (reparse points) 进行备份或恢复。
- 仅备份或恢复块存储。

此外, 备份应用程序可以执行增量优化的备份, 如下所示：

- 仅备份自上次备份以来创建, 修改或删除的更改的文件。
- 备份更改的块存储容器文件。
- 在子文件级别执行增量备份。

 **注意：**新块被追加到当前块存储容器。当其大小达到大约 1 GB 时, 该容器文件被密封, 并创建一个新的容器文件。

还原操作 (Restore operations)

还原操作也可以从重复数据删除中获益。任何文件级的完全卷恢复操作都可以受益, 因为它们本质上是备份过程的逆过程, 而较少的数据意味着更快的操作。完全卷恢复的方法是：

1. 还原完整的重复数据删除元数据和容器文件集。
2. 还原完整的重复数据删除重分析点集。
3. 还原所有未重复数据删除的文件。

从优化备份进行块级还原会自动进行优化还原, 因为还原过程发生在重复数据删除 (在文件级别工作)。

与第三方供应商的任何产品一样, 您应该验证备份解决方案是否支持 Windows Server 2016 中的重复数据删除, 因为不支持的备份解决方案可能在还原后引入损坏。以下是在 Windows Server 2016 中支持重复数据删除的解决方案的常见方法：

- 一些备份供应商支持未优化的备份, 在备份时重新利用重复数据删除的文件; 即将文件备份为正常的全尺寸文件。

- 一些备份供应商支持全卷备份的优化备份，该备份会按原样备份已重复数据删除的文件；即，作为具有块存储的重分析点存根。
- 一些备份供应商支持这两种。

备份供应商应该能够对其产品支持的内容，使用的方法以及与哪个版本进行评论。



注意：有关详细信息，请参阅：“备份和恢复启用了重复数据删除功能的卷”，网址为：
<http://aka.ms/w8iows>

问题：您可以在启用存储分层的驱动器上启用重复数据删除功能吗？

问题：您可以在 ReFS 格式化的驱动器上启用重复数据删除功能吗？

问题：是否可以在运行虚拟机的卷上启用重复数据删除，并将其应用于这些虚拟机？

实验 B: 实施重复数据删除 (Implementing Data Deduplication)

场景 (Scenario)

测试存储冗余和性能选项后, 您决定最大限度地利用可用的磁盘空间, 尤其是在通用文件服务器上, 这将是有益的。您决定测试重复数据删除解决方案以最大限度地提高用户的存储可用性。

新增: 在测试了存储冗余和性能选项后, 现在决定最大限度地利用可用的磁盘空间也是有益的, 特别是对于不断增长的需求的虚拟机存储。您决定测试重复数据删除解决方案以最大限度地提高虚拟机的存储可用性。

目标 (Objectives)

完成本实验后, 您将能够:

- 安装重复数据删除角色服务。
- 启用重复数据删除。
- 检查重复数据删除的状态。

实验设置 (Lab Setup)

预计时间: 40 分钟

虚拟机: 28740B-LON-DC1, 28740B-LON-SVR1

用户名: Adatum\Administrator

密码: Pa55w.rd

对于本实验, 您必须使用可用的虚拟机环境。这些应已从实验室 A 运行。如果不是, 在开始实验之前, 您必须完成以下步骤, 然后完成实验 A:

1. 在主机计算机上, 启动 Hyper-V Manager。
2. 在 Hyper-V Manager 中, 单击 28740B-LON-DC1, 然后在 Actions 窗格中单击 Start。
3. 在 Actions 窗格中, 单击 Connect。等待直到虚拟机启动。
4. 使用以下凭据登录:
 - 用户名: Administrator
 - 密码: Pa55w.rd
 - 域: Adatum
5. 对 28740B-LON-SVR1 重复步骤 2 到 4。

练习 1: 安装重复数据删除 (Installing Data Deduplication)

场景 (Scenario)

您决定通过使用服务器管理器在密集使用的文件服务器上安装重复数据删除角色服务。

本练习的主要任务如下:

1. 安装重复数据删除角色服务。
2. 检查重复数据删除的状态。
3. 验证虚拟机性能。

► 任务 1：安装重复数据删除角色服务

- 在 LON-SVR1 上，在服务器管理器中，添加重复数据删除角色服务。

► 任务 2：检查重复数据删除的状态

- 切换到 Windows PowerShell。
- 要验证重复数据删除状态，请运行以下命令：

```
Get-DedupVolume
Get-DedupStatus
```

- 这些命令不返回任何结果。这是因为您需要在安装卷后在卷上启用它。

► 任务 3：验证虚拟机性能 (Verify the virtual machine performance)

- 在 LON-SRV1 上，在 Windows PowerShell 中，运行以下命令：

```
Measure-Command -Expression {Get-ChildItem -Path D:\ -Recurse}
```



注意：您将在实验中稍后使用上一个命令返回的值。

结果：完成此练习后，您已成功安装了重复数据删除角色服务，并在其中一个文件服务器上启用了该服务。

练习 2：配置重复数据删除 (Configuring Data Deduplication)

场景 (Scenario)

您确定驱动器 D 被大量使用，并且您怀疑它包含某些文件夹中的重复文件。您决定启用和配置重复数据删除角色，以减少此卷上的占用空间。

本练习的主要任务如下：

- 配置重复数据删除。
- 配置优化以立即运行并查看状态。
- 验证文件是否已优化。
- 再次验证 VM 性能。
- 准备下一个模块。

► 任务 1：配置重复数据删除

- 在服务器管理器中，单击 File and Storage Services
- 单击 Disks
- 单击磁盘 1，然后单击 D 卷。
- 为 General purpose file server 设置启用重复数据删除。
- 配置以下设置：
 - 重复数据删除超过 (天) 的文件：0
 - 启用：throughput optimization.
 - 排除：D:\shares

► 任务 2：配置优化以立即运行并查看状态

- 在 LON-SRV1, 在 Windows PowerShell 中, 运行以下命令：

```
Start-DedupJob D: -Type Optimization -Memory 50
Get-DedupJob -Volume D:
```



注意：从上一个命令验证优化作业的状态。重复上一个命令，直到进度显示为 100%。

► 任务 3：验证文件是否已优化

- 在 LON-SRV1 上, 在文件资源管理器中, 导航到 D:\Labfiles\Mod04 中的文件, 并从几个文件属性中查看以下值：Size on disk 和 Size。
- 在 Windows PowerShell 中, 要验证重复数据删除状态, 请运行以下命令：

```
Get-DedupStatus -Volume D: | fl
Get-DedupVolume -Volume D: | fl
```



注意：观察优化文件的数量。

- 在服务器管理器中, 单击 File and Storage Services, 选择 Disk 1, 然后选择卷 D。
- 刷新显示, 并观察 Deduplication Rate 和 Deduplication Savings。



注意：因为驱动器 D 上的大多数文件都很小, 您可能不会注意到大量的保存空间。

► 任务 4：再次验证虚拟机性能

- 在 PowerShell (Admin) 中, 运行以下命令：

```
Measure-Command -Expression {Get-ChildItem -Path D:\ -Recurse}
```



注意：将上一个命令返回的值与实验室中前一个命令的值进行比较, 以评估系统性能是否已更改。

► 任务 5：准备下一个单元

完成实验后, 将虚拟机还原到其初始状态。

- 在主机计算机上, 启动 Hyper-V Manager。
- 在 Virtual Machines 列表中, 右键单击 28740B-LON-SRV1, 然后单击 Virtual Machines。
- 在 Revert Virtual Machine 对话框中, 单击 Revert。
- 对 28740B-LON-DC1 重复步骤 2 和 3。

结果：完成此练习后, 您已成功配置了 LON-SRV1 上适当数据卷的重复数据删除。

问题：您的经理担心使用重复数据删除会对文件服务器卷的性能造成的影响。这个问题是否有效？

单元复习和作业 (Module Review and Takeaways)

复习问题 (Review Questions)

问题：将五个 2 TB 磁盘附加到 Windows Server 2012 计算机。您想要简化管理磁盘的过程。此外，您要确保如果一个磁盘发生故障，则故障磁盘的数据不会丢失。您可以启用那些功能来实现这个目标？

问题：您的经理要求您考虑在存储架构中使用重复数据删除。在什么情况下，重复数据删除角色服务特别有用？

常见问题和故障诊断技巧 (Common Issues and Troubleshooting Tips)

常见问题	故障诊断技巧
当重复数据删除卷上的可用磁盘空间接近零时，某些文件无法读取。	

海量视频题库 myitpub.com QQ:5565462