深入了解
ONTAP Select

NetApp
May 07, 2026

目录

深入了解

存储

ONTAP Select 存储：一般概念和特征

在探索特定存储组件之前，先了解适用于 ONTAP Select 环境的一般存储概念。

存储配置阶段

ONTAP Select 主机存储的主要配置阶段包括：

- 部署前先决条件
  - 确保每个虚拟机监控程序主机都已配置并准备好进行 ONTAP Select 部署。
  - 配置涉及物理驱动器、RAID 控制器和组、LUN 以及相关的网络准备。
  - 此配置在 ONTAP Select 之外执行。
- 使用虚拟机监控程序管理员实用程序进行配置
  - 您可以使用虚拟机管理程序管理实用程序（例如，VMware 环境中的 vSphere）配置存储的某些方面。
  - 此配置在 ONTAP Select 之外执行。
- 使用 ONTAP Select Deploy 管理实用程序进行配置
  - 您可以使用 Deploy 管理实用程序配置核心逻辑存储构造。
  - 这可以通过 CLI 命令显式执行，也可以作为部署的一部分由实用程序自动执行。
- 部署后配置
  - ONTAP Select 部署完成后，您可以使用 ONTAP CLI 或 System Manager 配置集群。
  - 此配置在 ONTAP Select Deploy 之外执行。

受管存储与非受管存储

由 ONTAP Select 访问和直接控制的存储是托管存储。同一虚拟机监控程序主机上的任何其他存储都是非托管存储。

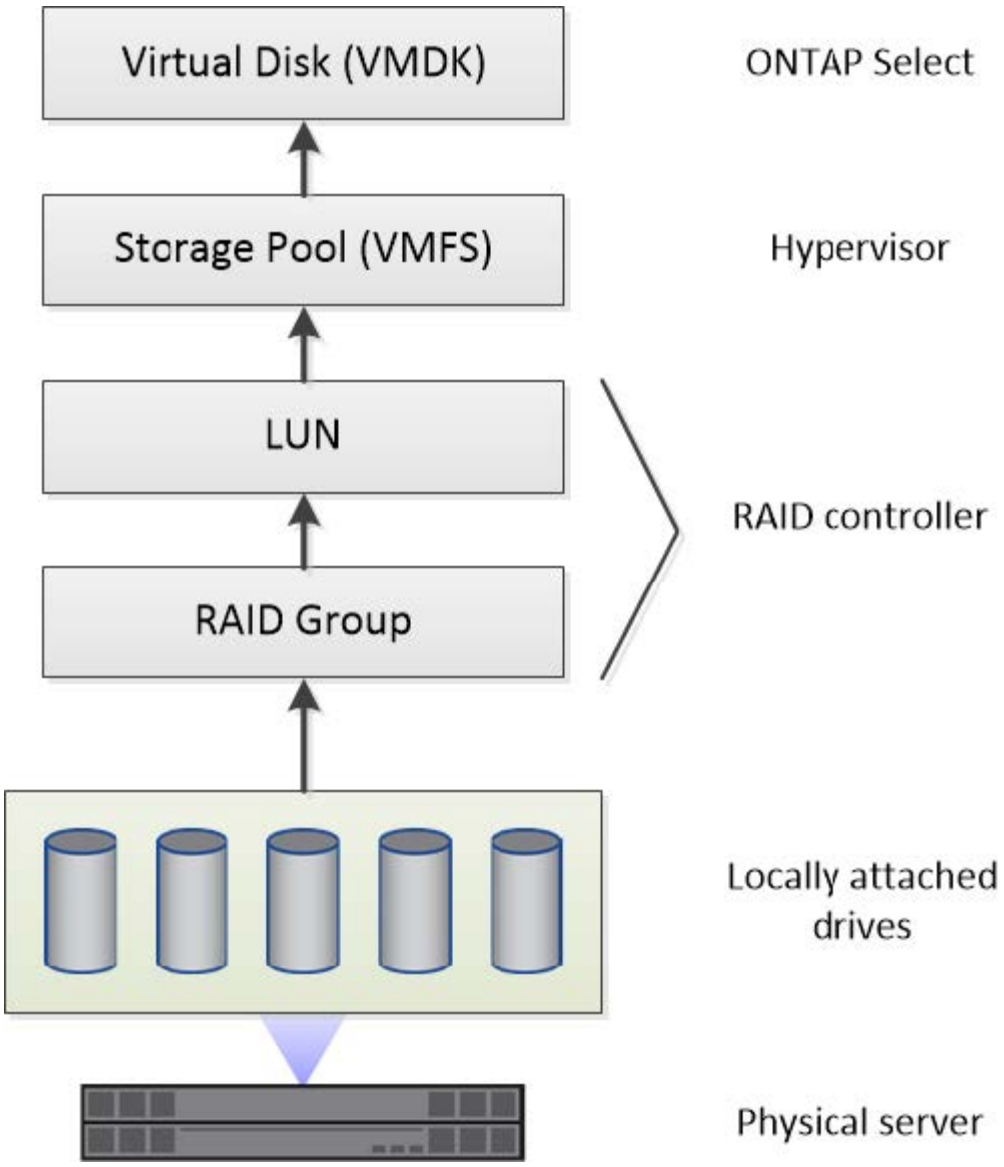
同构物理存储

组成 ONTAP Select 托管存储的所有物理驱动器必须是同质的。也就是说，关于以下特性，所有硬件必须相同：

- 类型（SAS、NL-SAS、SATA、SSD）
- 速度 (RPM)

本地存储环境示意图

每个虚拟机监控程序主机都包含可供 ONTAP Select 使用的本地磁盘和其他逻辑存储组件。这些存储组件从物理磁盘开始以分层结构排列。



本地存储组件的特性

有几个概念适用于 ONTAP Select 环境中使用的本地存储组件。在准备 ONTAP Select 部署之前，您应该熟悉这些概念。这些概念按类别排列：RAID 组和 LUN、存储池和虚拟磁盘。

将物理驱动器分组到 **RAID** 组和 **LUN** 中

一个或多个物理磁盘可以本地连接到主机服务器并可用于 ONTAP Select。物理磁盘被分配给 RAID 组，然后将其作为一个或多个 LUN 呈现给虚拟机监控程序主机操作系统。每个 LUN 都作为物理硬盘驱动器呈现给虚拟机管理程序主机操作系统。

配置 ONTAP Select 主机时，应注意以下事项：

- 所有托管存储都必须通过单个 RAID 控制器进行访问
- 根据供应商的不同，每个 RAID 控制器支持每个 RAID 组的最大驱动器数量

一个或多个 **RAID** 组

每个 ONTAP Select 主机必须具有单个 RAID 控制器。您应该为 ONTAP Select 创建单个 RAID 组。但是，在某些情况下，您可能会考虑创建多个 RAID 组。请参阅 "最佳实践摘要"。

存储池注意事项

在准备部署 ONTAP Select 时，您应该了解与存储池相关的几个问题。

在 VMware 环境中，存储池与 VMware 数据存储区是同义词。

存储池和 **LUN**

每个 LUN 都被视为虚拟机监控程序主机上的本地磁盘，可以作为一个存储池的一部分。每个存储池都使用虚拟机监控程序主机操作系统可以使用的文件系统进行格式化。

您必须确保存储池作为 ONTAP Select 部署的一部分正确创建。您可以使用虚拟机管理程序管理工具创建存储池。例如，使用 VMware，您可以使用 vSphere 客户端创建存储池。然后将存储池传递给 ONTAP Select Deploy 管理实用程序。

管理 **ESXi** 上的虚拟磁盘

在准备部署 ONTAP Select 时，您应该了解与虚拟磁盘相关的几个问题。

虚拟磁盘和文件系统

为 ONTAP Select 虚拟机分配了多个虚拟磁盘驱动器。每个虚拟磁盘实际上是存储池中包含的文件，由虚拟机管理程序维护。ONTAP Select 使用几种类型的磁盘，主要是系统磁盘和数据磁盘。

您还应注意以下有关虚拟磁盘的几点：

- 必须先提供存储池，然后才能创建虚拟磁盘。
- 在创建虚拟机之前无法创建虚拟磁盘。
- 必须依靠 ONTAP Select Deploy 管理实用程序来创建所有虚拟磁盘（即，管理员不得在 Deploy 实用程序之外创建虚拟磁盘）。

配置虚拟磁盘

虚拟磁盘由 ONTAP Select 管理。当您使用 Deploy 管理实用程序创建集群时，它们会自动创建。

ESXi 上的外部存储环境示意图

ONTAP Select vNAS 解决方案使 ONTAP Select 能够使用驻留在虚拟机监控程序主机外部存储上的数据存储区。数据存储区可以使用 VMware vSAN 通过网络访问，也可以直接在外部存储阵列上访问。

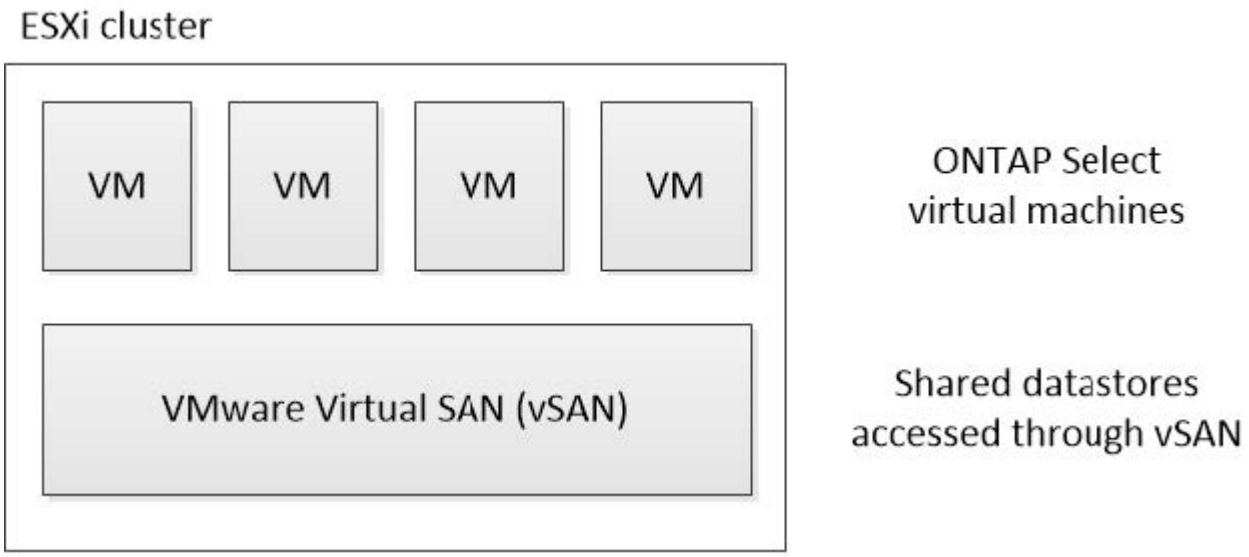
ONTAP Select 可以配置为使用以下类型的 VMware ESXi 网络数据存储，这些数据存储位于虚拟机管理程序主机外部：

- vSAN（虚拟 SAN）
- VMFS

- NFS

vSAN 数据存储库

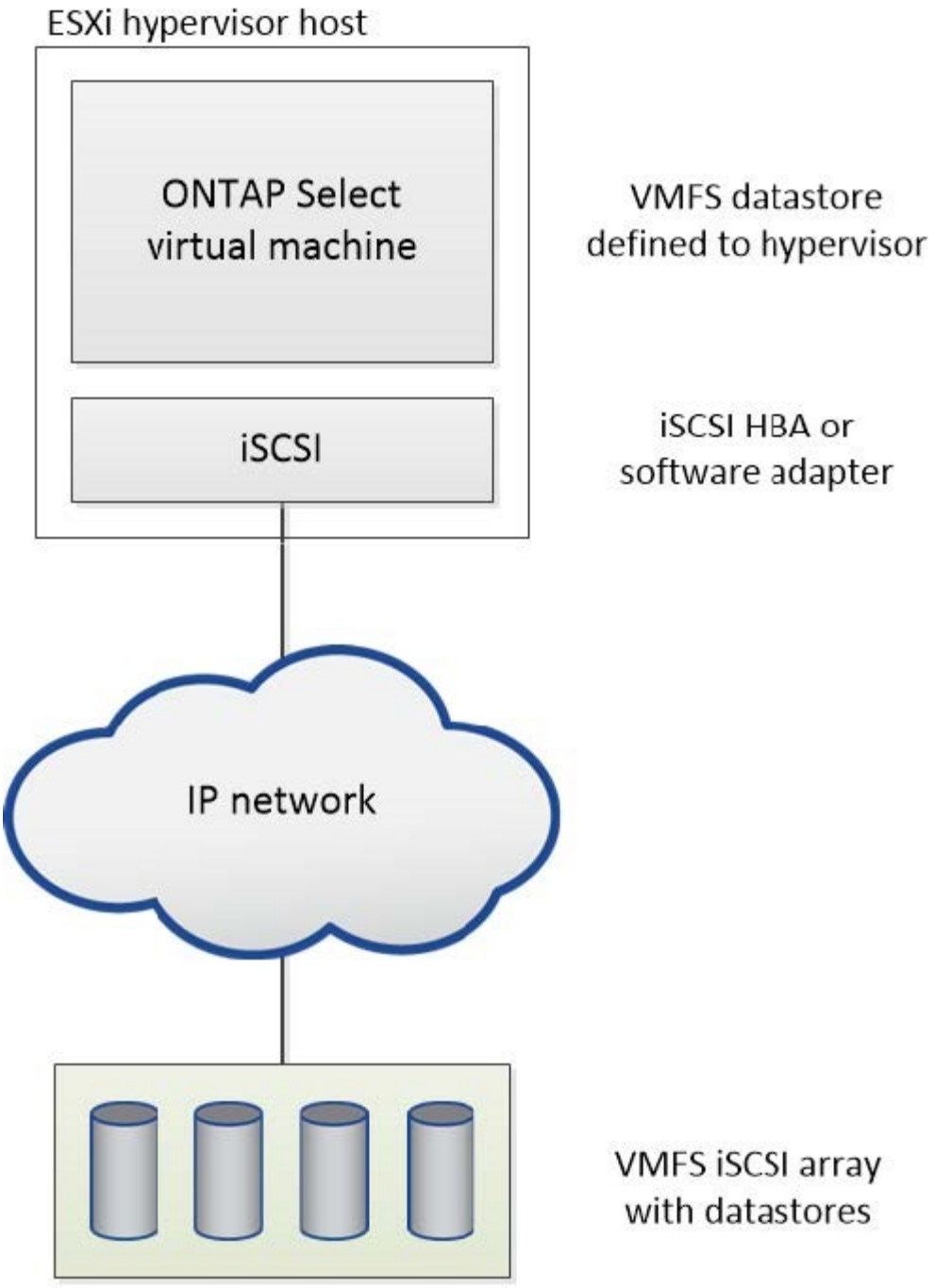
每个 ESXi 主机都可以有一个或多个本地 VMFS 数据存储。通常，这些数据存储只能由本地主机访问。但是，VMware vSAN 允许 ESXi 集群中的每个主机共享集群中的所有数据存储，就像它们是本地数据存储一样。下图说明了 vSAN 如何创建在 ESXi 集群中的主机之间共享的数据存储池。



外部存储阵列上的 **VMFS** 数据存储库

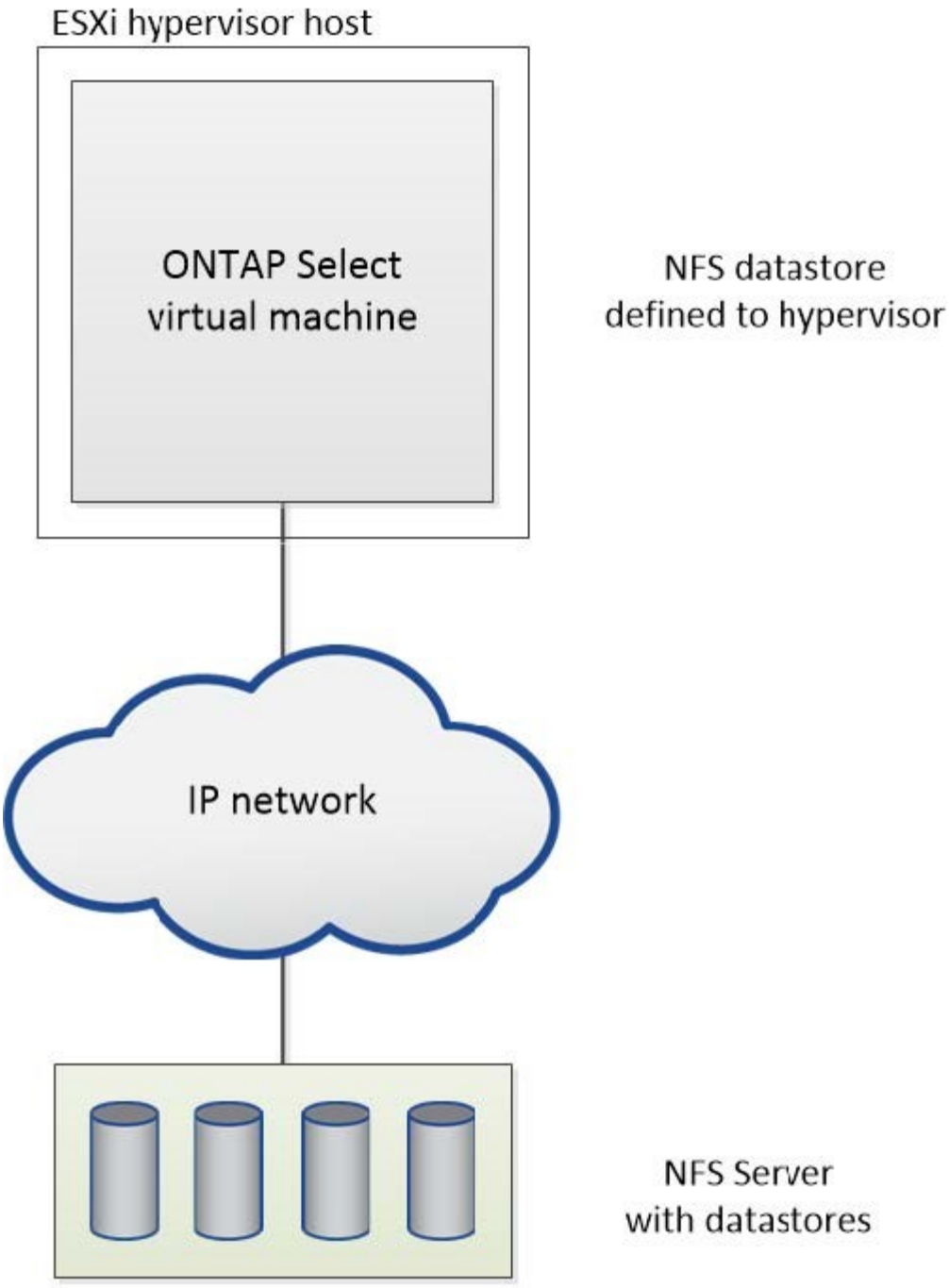
您可以创建驻留在外部存储阵列上的 VMFS 数据存储。使用几种不同的网络协议之一访问存储。下图说明了使用 iSCSI 协议访问的外部存储阵列上的 VMFS 数据存储。

ONTAP Select 支持 VMware Storage/SAN Compatibility 文档中描述的所有外部存储阵列，包括 iSCSI、Fiber Channel 和 Fiber Channel over Ethernet。



外部存储阵列上的 **NFS** 数据存储库

您可以创建驻留在外部存储阵列上的 NFS 数据存储。使用 NFS 网络协议访问存储。下图说明了通过 NFS 服务器设备访问的外部存储上的 NFS 数据存储。



适用于 **ONTAP Select** 本地连接存储的硬件 **RAID** 服务

当硬件 RAID 控制器可用时，ONTAP Select 可以将 RAID 服务移动到硬件控制器，以提高写入性能并防止物理驱动器故障。因此，ONTAP Select 群集中所有节点的 RAID 保护由本地连接的 RAID 控制器提供，而不是通过 ONTAP 软件 RAID 提供。

ONTAP Select 数据聚合配置为使用 RAID 0，因为物理 RAID 控制器正在向底层驱动器提供 RAID 条带化。不支持其他 RAID 级别。

用于本地连接存储的 **RAID** 控制器配置

所有为 ONTAP Select 提供后备存储的本地连接磁盘都必须位于 RAID 控制器后面。大多数商品服务器都配有多个不同价位的 RAID 控制器选项，每个选项都有不同的功能级别。目的是支持尽可能多的这些选项，前提是它们满足对控制器的某些最低要求。

> 您无法从使用硬件 RAID 配置的 ONTAP Select VM 中分离虚拟磁盘。分离磁盘仅支持使用软件 RAID 配置的 ONTAP Select VM。有关详细信息，请参见 "更换 ONTAP Select 软件 RAID 配置中的故障驱动器"。

管理 ONTAP Select 磁盘的 RAID 控制器必须满足以下要求：

- 硬件 RAID 控制器必须具有电池备份单元 (BBU) 或闪存后备写入缓存 (FBWC)，并支持 12Gbps 的吞吐量。
- RAID 控制器必须支持能够承受至少一个或两个磁盘故障的模式（RAID 5 和 RAID 6）。
- 必须将驱动器缓存设置为已禁用。
- 必须将写入策略配置为回写模式，并在 BBU 或闪存故障时回退到直写模式。
- 必须将 I/O 读取策略设置为已缓存。

所有为 ONTAP Select 提供后备存储的本地连接磁盘必须放置在运行 RAID 5 或 RAID 6 的 RAID 组中。对于 SAS 驱动器和 SSD，使用多达 24 个驱动器的 RAID 组可使 ONTAP 从将传入读取请求分散到更多磁盘中获益。这样做可以显著提高性能。对于 SAS/SSD 配置，针对单 LUN 配置与多 LUN 配置进行了性能测试。没有发现明显的差异，因此，为了简单起见，NetApp 建议创建支持您的配置需求所需的最少数量的 LUN。

NL-SAS 和 SATA 驱动器需要一套不同的最佳实践。出于性能原因，磁盘的最小数量仍为 8 个，但 RAID 组大小不应大于 12 个驱动器。NetApp 还建议每个 RAID 组使用一个备用磁盘；但是，可以对所有 RAID 组使用全局备用磁盘。例如，您可以为每三个 RAID 组使用两个备件，每个 RAID 组由 8 到 12 个驱动器组成。

> 较旧 ESXi 版本的最大范围和数据存储区大小为 64TB，这可能会影响支持这些大容量驱动器提供的总原始容量所需的 LUN 数量。

RAID 模式

许多 RAID 控制器最多支持三种操作模式，每种模式都代表了写入请求所采用的数据路径的显著差异。这三种模式如下：

- 直写。所有传入的 I/O 请求都被写入 RAID 控制器缓存，然后在确认请求返回到主机之前立即刷新到磁盘。
- 写入绕过。所有传入的 I/O 请求都直接写入磁盘，绕过 RAID 控制器缓存。
- 写回。所有传入的 I/O 请求都直接写入控制器缓存并立即确认回主机。使用控制器将数据块异步刷新到磁盘。

写回模式提供最短的数据路径，I/O 确认在块进入缓存后立即发生。此模式为混合读/写工作负载提供最低延迟和最高吞吐量。但是，如果没有 BBU 或非易失性闪存技术，如果系统在此模式下运行时发生电源故障，用户将面临丢失数据的风险。

ONTAP Select 需要电池备份或闪存单元；因此，我们可以确信，在发生此类故障时，缓存的块会刷新到磁盘。为此，需要将 RAID 控制器配置为写回模式。

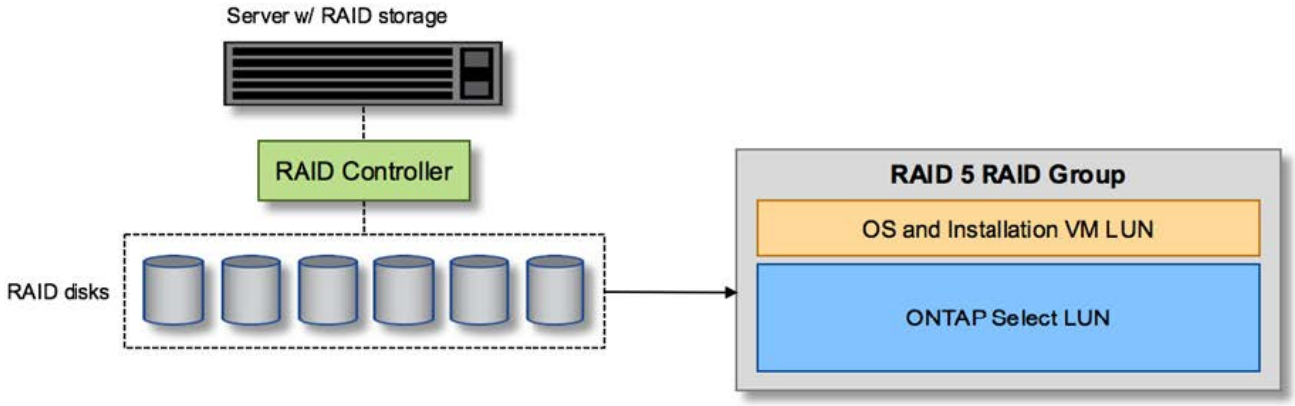
ONTAP Select 和 **OS** 之间共享的本地磁盘

最常见的服务器配置是所有本地连接的磁盘轴位于单个 RAID 控制器后面。应至少配置两个 LUN：一个用于虚拟机管理程序，一个用于 ONTAP Select VM。

例如，假设 HP DL380 g8 具有六个内部驱动器和一个 Smart Array P420i RAID 控制器。所有内部驱动器均由此 RAID 控制器管理，系统上不存在其他存储。

下图显示了此配置样式。在此示例中，系统上不存在其他存储；因此，系统管理程序必须与 ONTAP Select 节点共享存储。

仅使用 **RAID** 管理的磁盘轴的服务器 **LUN** 配置



从与 ONTAP Select 相同的 RAID 组配置 OS LUN 允许虚拟机监控程序操作系统（以及从该存储中配置的任何客户端 VM）受益于 RAID 保护。此配置可防止单驱动器故障导致整个系统停机。

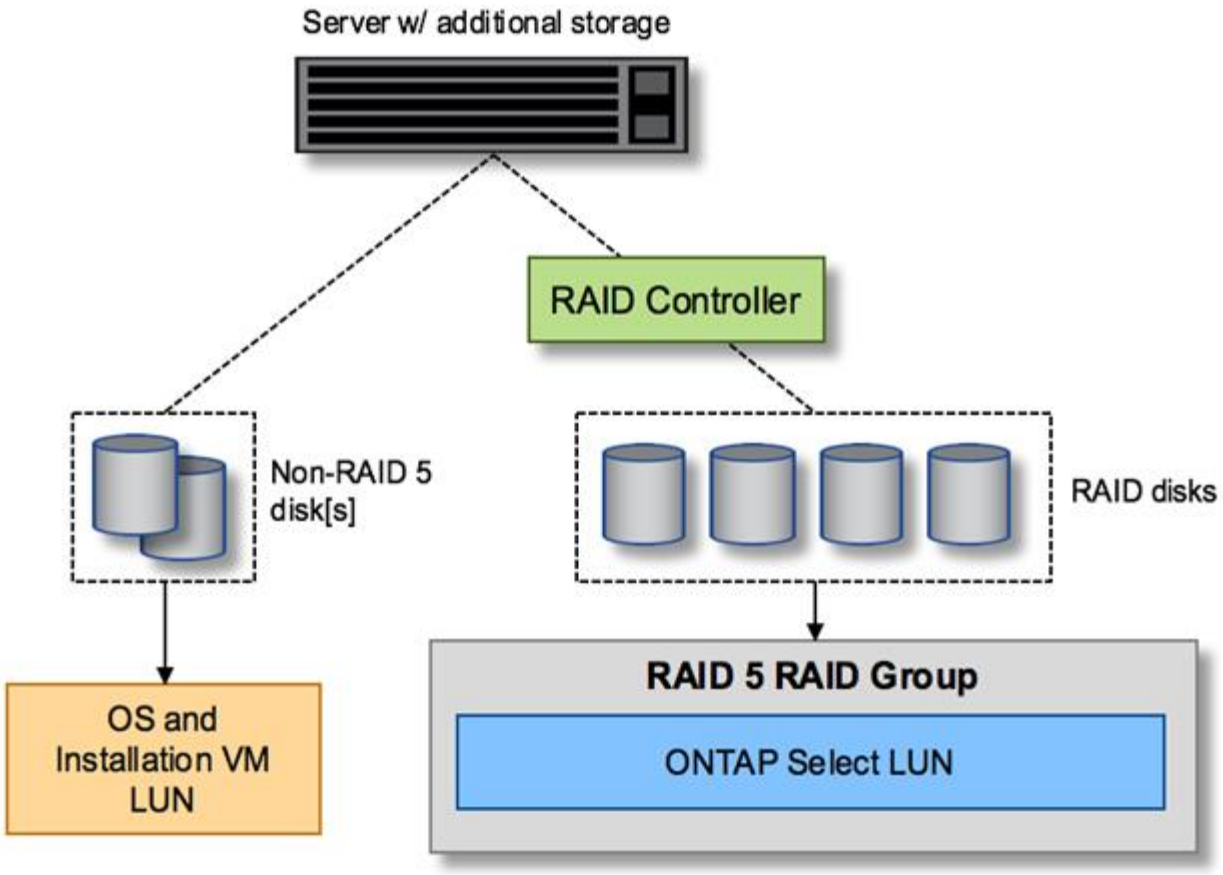
本地磁盘在 **ONTAP Select** 和 **OS** 之间拆分

服务器供应商提供的其他可能配置包括使用多个 RAID 或磁盘控制器配置系统。在此配置中，一组磁盘由一个磁盘控制器管理，该控制器可能提供也可能不提供 RAID 服务。第二组磁盘由能够提供 RAID 5/6 服务的硬件 RAID 控制器管理。

通过这种配置风格，可以提供 RAID 5/6 服务的 RAID 控制器后面的主轴集应由 ONTAP Select VM 独家使用。根据管理的总存储容量，应将磁盘主轴配置为一个或多个 RAID 组和一个或多个 LUN。然后，这些 LUN 将用于创建一个或多个数据存储，所有数据存储都受到 RAID 控制器的保护。

第一组磁盘是为虚拟机监控程序操作系统和不使用 ONTAP 存储的任何客户端虚拟机保留的，如下图所示。

混合 **RAID**/非 **RAID** 系统上的服务器 **LUN** 配置



多个 **LUN**

在两种情况下，必须更改单 RAID 组/单 LUN 配置。使用 NL-SAS 或 SATA 驱动器时，RAID 组大小不得超过 12 个驱动器。此外，单个 LUN 可以大于底层虚拟机管理程序存储限制，无论是单个文件系统范围最大大小还是总存储池最大大小。然后，必须将底层物理存储拆分为多个 LUN，才能成功创建文件系统。

VMware vSphere 虚拟机文件系统限制

某些版本的 ESXi 上数据存储库的最大大小为 64TB。

如果服务器连接的存储容量超过 64TB，则可能需要调配多个 LUN，每个 LUN 小于 64TB。创建多个 RAID 组以缩短 SATA/NL-SAS 驱动器的 RAID 重建时间也会导致调配多个 LUN。

当需要多个 LUN 时，一个主要考虑点是确保这些 LUN 具有相似且一致的性能。如果要在单个 ONTAP 聚合中使用所有 LUN，则这一点尤其重要。或者，如果一个或多个 LUN 的子集具有明显不同的性能配置文件，则强烈建议将这些 LUN 隔离在单独的 ONTAP 聚合中。

可以使用多个文件系统范围来创建单个数据存储区，最多可达数据存储区的最大大小。要限制需要 ONTAP Select 许可证的容量，请确保在群集安装期间指定容量上限。此功能允许 ONTAP Select 仅使用（因此仅需要许可证）数据存储区中的一部分空间。

或者，可以从在单个 LUN 上创建单个数据存储区开始。当需要更大的 ONTAP Select 容量许可证的额外空间时，可以将该空间作为扩展区添加到同一个数据存储区，直到数据存储区的最大大小。达到最大大小后，可以创建新数据存储区并将其添加到 ONTAP Select。支持这两种类型的容量扩展操作，可以通过使用 ONTAP Deploy storage-add 功能来实现。每个 ONTAP Select 节点可配置为支持高达 400TB 的存储。从多个数据存储区调配容

量需要两个步骤的过程。

初始集群创建可用于创建占用初始数据存储中部分或全部空间的 ONTAP Select 集群。第二步是使用其他数据存储执行一个或多个容量添加操作，直到达到所需的总容量。此功能详见章节 "增加存储容量"。

> VMFS 开销为非零（请参阅 VMware KB 1001618），并且尝试使用数据存储区报告为可用的整个空间导致在群集创建操作期间出现虚假错误。

每个数据存储区中有 2% 的缓冲区未使用。此空间不需要容量许可证，因为它未被 ONTAP Select 使用。ONTAP Deploy 自动计算缓冲区的确切千兆字节数，只要未指定容量上限即可。如果指定了容量上限，则首先强制执行该大小。如果容量上限大小在缓冲区大小范围内，则集群创建失败，并显示一条错误消息，指定可用作容量上限的正确最大大小参数：

```
“InvalidPoolCapacitySize: Invalid capacity specified for storage pool
“ontap-select-storage-pool”, Specified value: 34334204 GB. Available
(after leaving 2% overhead space): 30948”
```

VMFS 6 既支持新安装，也支持作为现有 ONTAP Deploy 或 ONTAP Select VM 的存储 vMotion 操作的目标。

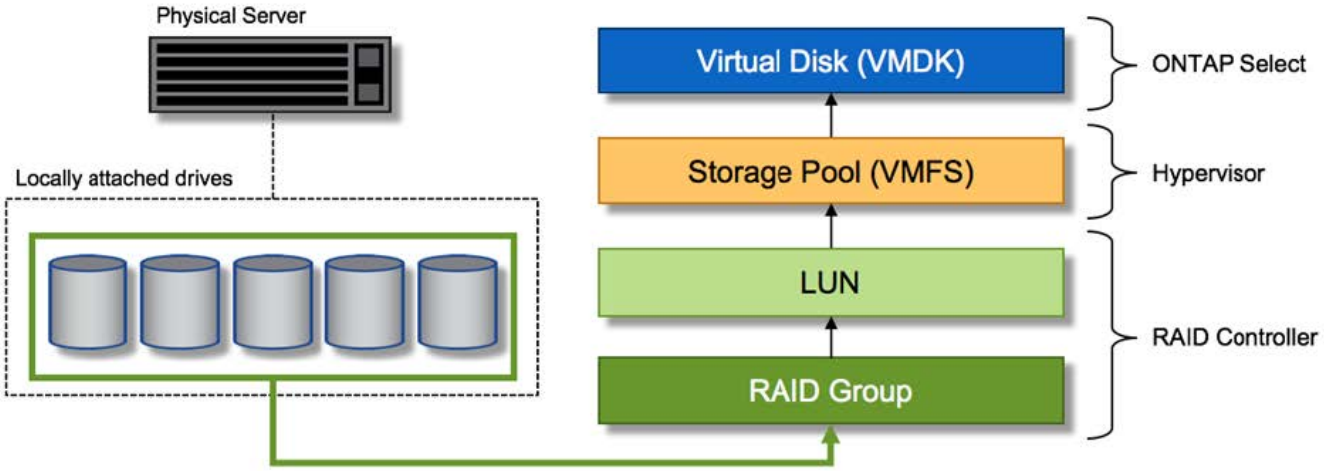
VMware 不支持从 VMFS 5 到 VMFS 6 的现场升级。因此，存储 vMotion 是允许任何虚拟机从 VMFS 5 数据存储区过渡到 VMFS 6 数据存储区的唯一机制。但是，对 ONTAP Select 和 ONTAP Deploy 的存储 vMotion 支持已扩展，以涵盖除了从 VMFS 5 过渡到 VMFS 6 的特定目的之外的其他场景。

ONTAP Select 虚拟磁盘

在其核心，ONTAP Select 向 ONTAP 提供一组从一个或多个存储池配置的虚拟磁盘。ONTAP 具有一组虚拟磁盘，它将其视为物理磁盘，并且存储堆栈的剩余部分由虚拟机管理程序抽象化。下图更详细地显示了这种关系，突出显示了物理 RAID 控制器、虚拟机监控程序和 ONTAP Select VM 之间的关系。

- RAID 组和 LUN 配置从服务器的 RAID 控制器软件中进行。使用 VSAN 或外部阵列时，不需要此配置。
- 存储池配置从虚拟机监控程序内部进行。
- 虚拟磁盘由单个虚拟机创建和拥有；在本示例中，由 ONTAP Select 创建和拥有。

虚拟磁盘到物理磁盘映射



虚拟磁盘配置

为了提供更流畅的用户体验，ONTAP Select 管理工具 ONTAP Deploy 会自动从关联的存储池配置虚拟磁盘，并将它们连接到 ONTAP Select VM。在初始设置期间以及执行存储添加操作期间会自动执行此操作。如果 ONTAP Select 节点是 HA 对的一部分，则会自动将虚拟磁盘分配给本地和镜像存储池。

ONTAP Select 会将底层连接的存储拆分为大小相等的虚拟磁盘，且每个虚拟磁盘不超过 16 TB。如果 ONTAP Select 节点是 HA 对的一部分，则会在每个集群节点上至少创建两个虚拟磁盘，并将其分配给本地和镜像丛，以便在镜像聚合中使用。

例如，ONTAP Select 可以分配一个 31TB 的数据存储区或 LUN（部署虚拟机并配置系统和根磁盘后剩余的空间）。然后创建四个约 7.75TB 的虚拟磁盘，并将其分配给相应的 ONTAP 本地和镜像丛。

> 向 ONTAP Select VM 添加容量可能会导致不同大小的 VMDK。有关详细信息，请参阅部分 "增加存储容量"。与 FAS 系统不同，不同大小的 VMDK 可以存在于同一个聚合中。ONTAP Select 在这些 VMDK 上使用 RAID 0 条带，从而能够充分利用每个 VMDK 中的所有空间，无论其大小如何。

虚拟化 **NVRAM**

NetApp FAS 系统通常配备物理 NVRAM PCI 卡，这是一种包含非易失性闪存的高性能卡。此卡通过授予 ONTAP 立即向客户端确认传入写入的能力，大大提高了写入性能。它还可以在称为转储的过程中安排将修改的数据块移动回较慢的存储介质。

商品系统通常不配备此类设备。因此，此 NVRAM 卡的功能已虚拟化，并放置在 ONTAP Select 系统引导磁盘上的分区中。正是出于这个原因，实例的系统虚拟磁盘的放置非常重要。这也是为什么该产品需要具有用于本地连接存储配置的弹性缓存的物理 RAID 控制器。

NVRAM 放置在自己的 VMDK 上。在自己的 VMDK 中拆分 NVRAM 允许 ONTAP Select VM 使用 vNVMe 驱动程序与其 NVRAM VMDK 进行通信。它还要求 ONTAP Select VM 使用与 ESXi 8.0 及更高版本兼容的硬件版本 13。

数据路径说明：**NVRAM** 和 **RAID** 控制器

虚拟化 NVRAM 系统分区和 RAID 控制器之间的交互可以通过遍历写入请求进入系统时所采用的数据路径来最好地突出显示。

对 ONTAP Select VM 的传入写入请求以 VM 的 NVRAM 分区为目标。在虚拟化层，此分区存在于 ONTAP Select 系统磁盘中，该磁盘是连接到 ONTAP Select VM 的 VMDK。在物理层，这些请求会缓存在本地 RAID 控制器中，就像所有针对底层主轴的块更改一样。从这里，写入会被确认回主机。

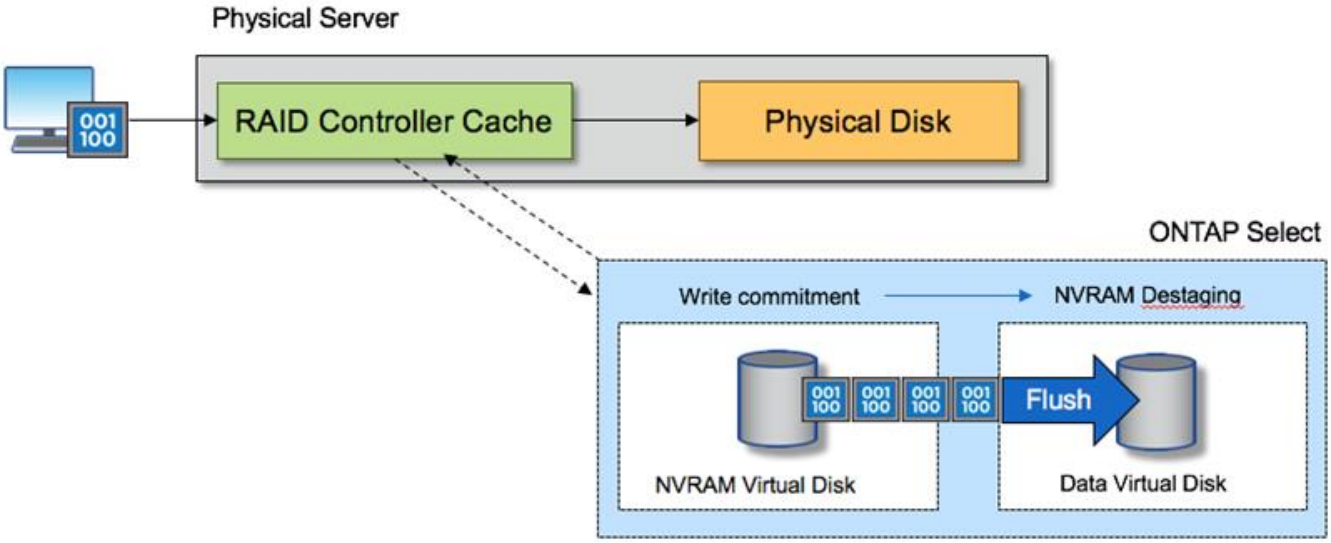
此时，物理上，块驻留在 RAID 控制器缓存中，等待刷新到磁盘。逻辑上，块驻留在 NVRAM 中，等待转储到相应的用户数据磁盘。

由于更改的块会自动存储在 RAID 控制器的本地缓存中，因此对 NVRAM 分区的传入写入会自动缓存并定期刷新到物理存储介质。这不应与将 NVRAM 内容定期刷新回 ONTAP 数据磁盘混淆。这两个事件无关，发生在不同的时间和频率。

下图显示了传入写入所采取的 I/O 路径。它突出了物理层（由 RAID 控制器缓存和磁盘表示）和虚拟层（由虚拟机的 NVRAM 和数据虚拟磁盘表示）之间的区别。

尽管在 NVRAM VMDK 上更改的块被缓存在本地 RAID 控制器缓存中，但缓存无法识别 VM 构造或其虚拟磁盘。它存储系统上的所有更改块，NVRAM 只是其中的一部分。这包括绑定到 hypervisor 的写入请求，如果它是从相同的支持主轴配置的。

对 **ONTAP Select VM** 的传入写入



NVRAM 分区在其自己的 VMDK 上分离。该 VMDK 使用 ESXi 8.0 或更高版本中提供的 vNVME 驱动程序连接。此更改对于具有软件 RAID 的 ONTAP Select 安装最为重要，因为它们无法从 RAID 控制器缓存中受益。

ONTAP Select 软件 RAID 配置服务，适用于本地连接存储

软件 RAID 是在 ONTAP 软件堆栈中实现的 RAID 抽象层。它提供与传统 ONTAP 平台（如 FAS）中 RAID 层相同的功能。RAID 层执行驱动器奇偶校验计算，并提供针对 ONTAP Select 节点内单个驱动器故障的保护。

与硬件 RAID 配置无关，ONTAP Select 还提供软件 RAID 选项。在某些环境中，硬件 RAID 控制器可能不可用或不受欢迎，例如当 ONTAP Select 部署在小型商品硬件上时。软件 RAID 扩展了可用的部署选项，以包括此类环境。要在您的环境中启用软件 RAID，请记住以下几点：

- 它可与 Premium 或 Premium XL 许可证一起使用。
- 它仅支持用于 ONTAP 根磁盘和数据磁盘的 SSD 或 NVMe（需要 Premium XL 许可证）驱动器。
- ONTAP Select VM 启动分区需要一个单独的系统磁盘。
  - 选择单独的磁盘（SSD 或 NVMe 驱动器），为系统磁盘（NVRAM、Boot/CF 卡、Coredump 和多节点设置中的 Mediator）创建数据存储区。

- 术语服务磁盘和系统磁盘可互换使用。
  - 服务磁盘是 ONTAP Select VM 中使用的虚拟磁盘 (VMDK)，用于为集群、引导等各种项目提供服务。
  - 服务磁盘物理上位于单个物理磁盘上（从主机视角来看，统称为服务/系统物理磁盘）。该物理磁盘必须包含一个 DAS 数据存储。ONTAP Deploy 在集群部署期间为 ONTAP Select 虚拟机创建这些服务磁盘。
- 无法跨多个数据存储区或多个物理驱动器进一步分离 ONTAP Select 系统磁盘。
- 硬件 RAID 未被弃用。

适用于本地连接存储的软件 **RAID** 配置

当使用软件 RAID 时，没有硬件 RAID 控制器是理想的，但是，如果系统确实具有现有的 RAID 控制器，则必须遵守以下要求：

- 您必须禁用硬件 RAID 控制器，以便磁盘可以直接呈现给系统（JBOD）。您通常可以在 RAID 控制器 BIOS 中进行此更改。
- 或者硬件 RAID 控制器应处于 SAS HBA 模式。例如，某些 BIOS 配置除了 RAID 之外还允许 "AHCI" 模式，您可以选择启用 JBOD 模式。这启用了直通，以便可以在主机上按原样看到物理驱动器。

根据控制器支持的最大驱动器数量，可能需要额外的控制器。使用 SAS HBA 模式时，请确保以最低 6Gbps 的速度支持 I/O 控制器（SAS HBA）。但是，NetApp 建议使用 12Gbps 的速度。

不支持其他硬件 RAID 控制器模式或配置。例如，一些控制器允许 RAID 0 支持，这可以人为地使磁盘通过，但其影响可能是不可取的。支持的物理磁盘大小（仅限 SSD）介于 200GB 和 16TB 之间。

管理员需要跟踪 ONTAP Select VM 正在使用哪些驱动器，并防止在主机上无意中使用这些驱动器。

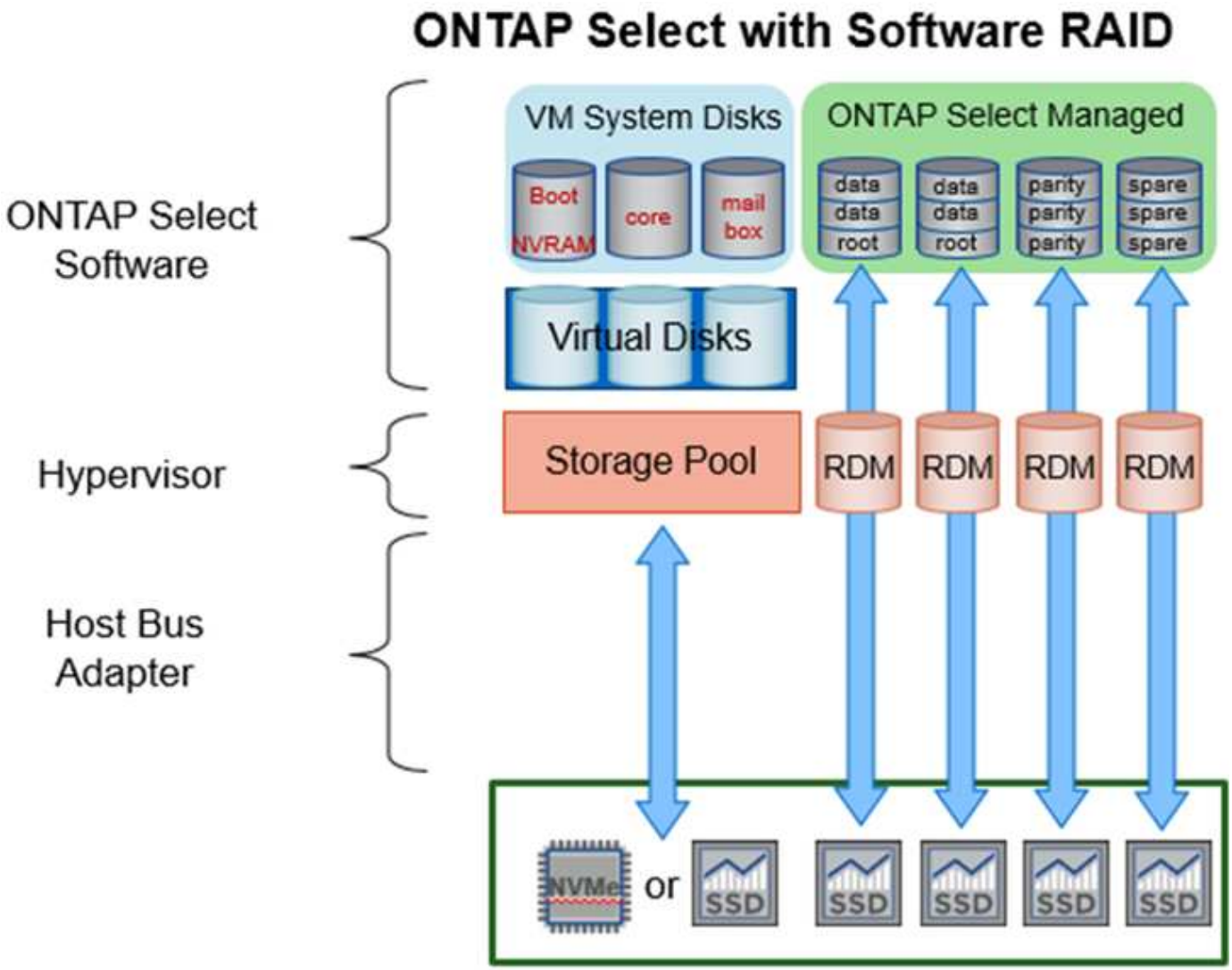
ONTAP Select 虚拟磁盘和物理磁盘

对于带有硬件 RAID 控制器的配置，物理磁盘冗余由 RAID 控制器提供。ONTAP Select 显示了一个或多个 VMDK，ONTAP 管理员可以从中配置数据聚合。这些 VMDK 以 RAID 0 格式进行条带化，因为使用 ONTAP 软件 RAID 是冗余的、低效的，并且由于在硬件级别提供的弹性而无效。此外，用于系统磁盘的 VMDK 与用于存储用户数据的 VMDK 位于同一个数据存储区中。

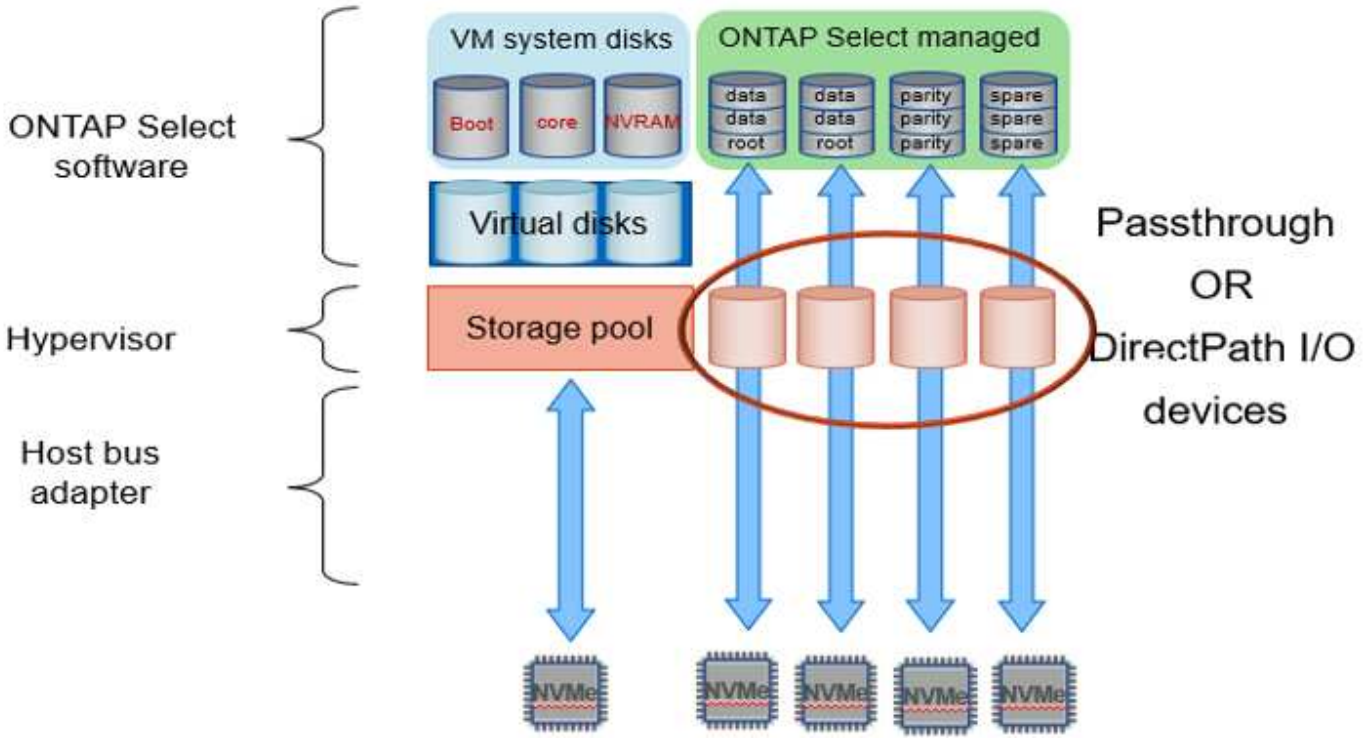
使用软件 RAID 时， ONTAP Deploy 将为 ONTAP Select 提供一组 VMDK 和物理磁盘原始设备映射 [RDM] （用于 SSD ）以及直通或 DirectPath IO 设备（用于 NVMe ）。

下图更详细地显示了这种关系，突出了用于 ONTAP Select VM 内部的虚拟化磁盘与用于存储用户数据的物理磁盘之间的差异。

ONTAP Select 软件 **RAID**：使用虚拟化磁盘和 **RDM**



系统磁盘 (VMDK) 驻留在同一个数据存储区和同一个物理磁盘上。虚拟 NVRAM 磁盘需要快速耐用的介质。因此，仅支持 NVMe 和 SSD 类型的数据存储。



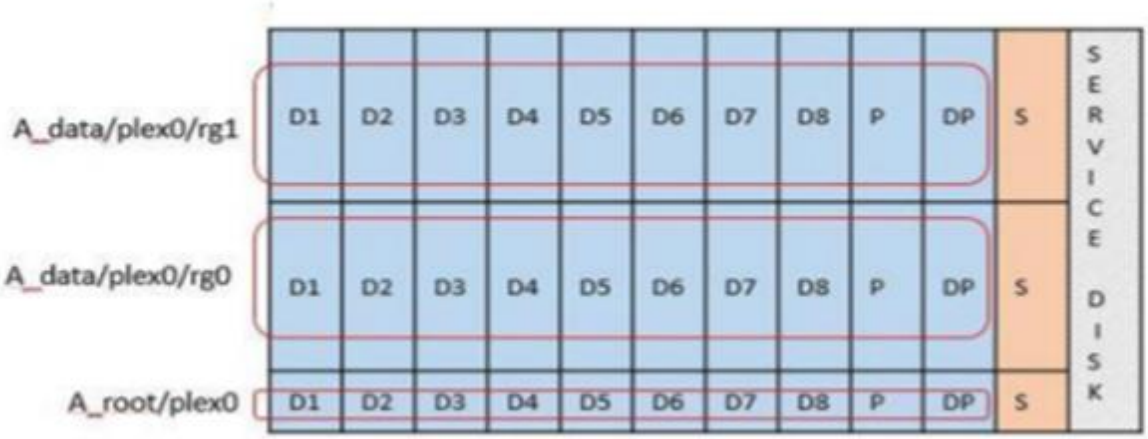
系统磁盘（VMDK）位于同一数据存储区和同一物理磁盘上。虚拟 NVRAM 磁盘需要快速耐用的介质。因此，仅支持 NVMe 和 SSD 类型的数据存储。将 NVMe 驱动器用于数据时，出于性能原因，系统磁盘也应为 NVMe 设备。对于全 NVMe 配置的系统磁盘，INTEL Optane 卡是一个很好的选择。

在当前版本中，无法跨多个数据存储或多个物理驱动器进一步分离 ONTAP Select 系统磁盘。

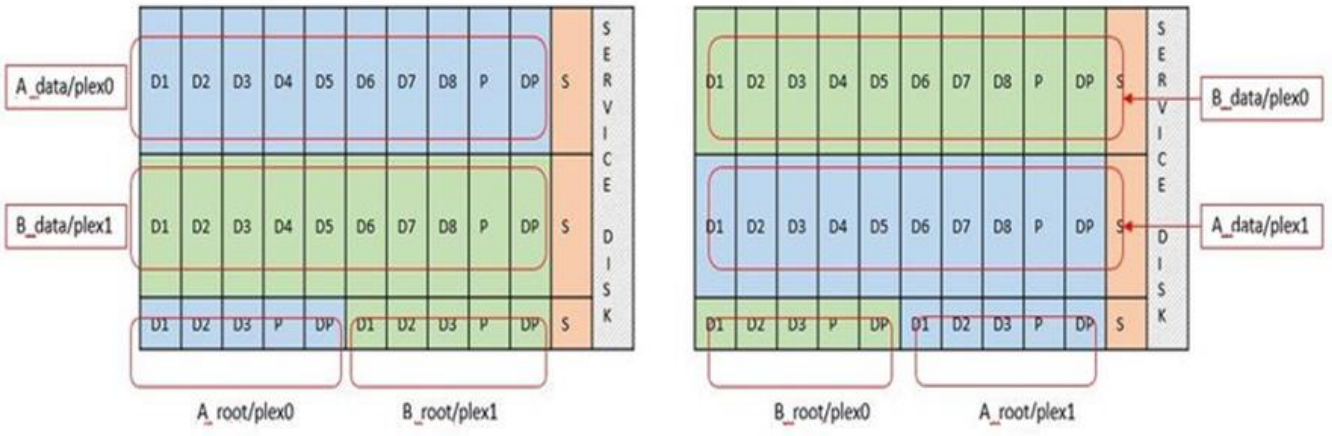
每个数据磁盘分为三个部分：一个小根分区（条带）和两个大小相等的分区，以创建在 ONTAP Select VM 中看到的两个数据磁盘。分区使用 Root Data Data (RD2) 架构，如下图所示，适用于单个节点群集和高可用性 (HA) 对中的节点。

`P` 表示奇偶校验驱动器， `DP` 表示双奇偶校验驱动器， `S` 表示备用驱动器。

单节点集群的 **RDD** 磁盘分区



多节点集群（**HA** 对）的 **RDD** 磁盘分区



ONTAP 软件 RAID 支持以下 RAID 类型：RAID 4、RAID-DP 和 RAID-TEC。这些是 FAS 和 AFF 平台使用的相同 RAID 结构。对于根配置，ONTAP Select 仅支持 RAID 4 和 RAID-DP。当对数据聚合使用 RAID-TEC 时，整体保护为 RAID-DP。ONTAP Select HA 使用无共享架构，将每个节点的配置复制到其他节点。这意味着每个节点必须存储其根分区和对等节点根分区的副本。数据磁盘具有单个根分区。这意味着数据磁盘的最小数量取决于 ONTAP Select 节点是否是 HA 对的一部分。

对于单节点集群，所有数据分区都用于存储本地（活动）数据。对于属于 HA 对的节点，一个数据分区用于存储该节点的本地（活动）数据，第二个数据分区用于镜像来自 HA 对等的活动数据。

直通（**DirectPath IO**）设备与原始设备映射（**RDM**）

ESXi 和 KVM 虚拟机管理程序不支持 NVMe 磁盘作为原始设备映射 (RDM)。要允许 ONTAP Select 直接控制 NVMe 磁盘，必须将这些驱动器配置为 ESXi 或 KVM 内的直通设备。将 NVMe 设备配置为直通设备时，需要服务器 BIOS 的支持，并且可能需要重新启动主机。此外，每个主机可以分配的直通设备数量也有限制，这可能因平台而异。但是，ONTAP Deploy 将此限制为每个 ONTAP Select 节点 14 个 NVMe 设备。这意味着 NVMe 配置以牺牲总容量为代价，提供非常高的 IOPS 密度 (IOPS/TB)。或者，如果您需要具有更大存储容量的高性能配置，建议配置为大型 ONTAP Select VM 大小、用于系统磁盘的 INTEL Optane 卡以及用于数据存储的标称数量的 SSD 驱动器。

要充分利用 NVMe 性能，请考虑较大的 ONTAP Select VM 大小。

直通设备和 RDM 之间还有一个额外的区别。RDM 可以映射到正在运行的 VM。直通设备需要重新启动 VM。这意味着任何 NVMe 驱动器更换或容量扩展（驱动器添加）过程都需要重新启动 ONTAP Select VM。驱动器更换和容量扩展（驱动器添加）操作由 ONTAP Deploy 中的工作流驱动。ONTAP Deploy 管理单节点集群的 ONTAP Select 重新启动和 HA 对的故障转移/故障回复。但是，请注意使用 SSD 数据驱动器（不需要 ONTAP Select 重新启动/故障转移）和使用 NVMe 数据驱动器（需要 ONTAP Select 重新启动/故障转移）之间的区别。

物理和虚拟磁盘配置

为了提供更流畅的用户体验，ONTAP Deploy 会自动从指定的数据存储区（物理系统磁盘）配置系统（虚拟）磁盘，并将它们连接到 ONTAP Select VM。此操作在初始设置期间会自动执行，以便 ONTAP Select VM 可以启动。对 RDM 进行分区，并自动生成根聚合。如果 ONTAP Select 节点是 HA 对的一部分，则会自动将数据分区分配给本地存储池和镜像存储池。此分配在群集创建操作和存储添加操作期间会自动执行。

由于 ONTAP Select 虚拟机上的数据磁盘与底层物理磁盘相关联，因此创建具有更多物理磁盘的配置会影响性能。

根聚合的 RAID 组类型取决于可用磁盘的数量。ONTAP Deploy 选择适当的 RAID 组类型。如果有足够的磁盘分配给节点，则使用 RAID-DP，否则创建 RAID-4 根聚合。

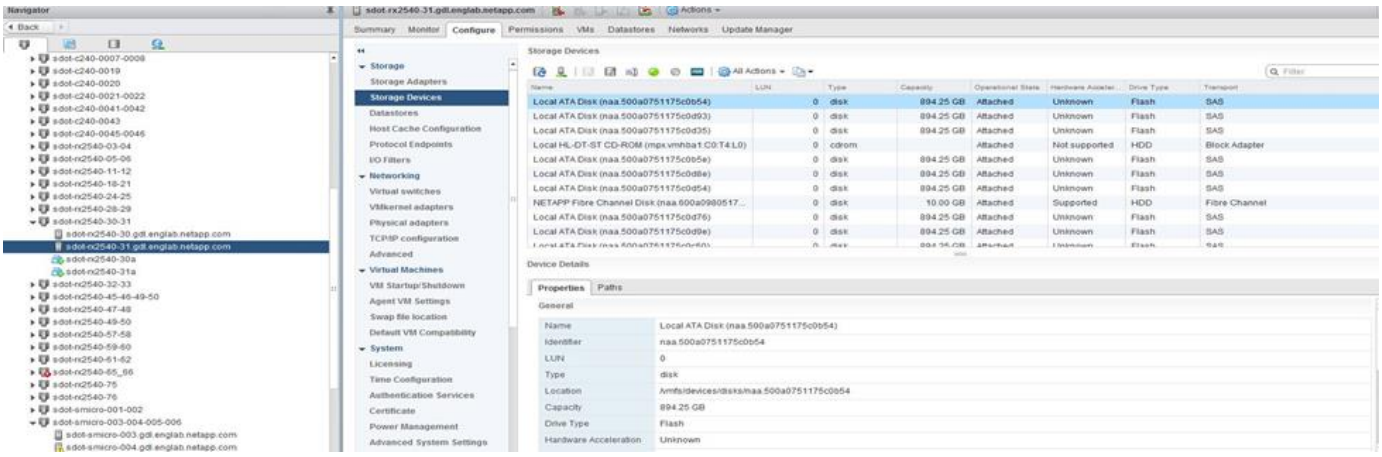
使用软件 RAID 向 ONTAP Select VM 添加容量时，管理员必须考虑物理驱动器大小和所需驱动器的数量。有关详细信息，请参见 "增加存储容量"。

与 FAS 和 AFF 系统类似，您只能将容量相同或更大的驱动器添加到现有的 RAID 组。容量更大的驱动器会调整为合适的大小。如果您要创建新的 RAID 组，则新的 RAID 组大小应与现有的 RAID 组大小相匹配，以确保整体聚合性能不会恶化。

将 **ONTAP Select** 磁盘与相应的 **ESXi** 或 **KVM** 磁盘匹配

ONTAP Select 磁盘通常标记为 NET x.y。您可以使用以下 ONTAP 命令获取磁盘 UUID：

```
<system name>::> disk show NET-1.1
Disk: NET-1.1
Model: Micron_5100_MTFD
Serial Number: 1723175C0B5E
UID:
*500A0751:175C0B5E*:00000000:00000000:00000000:00000000:00000000:00000000:
00000000:00000000
BPS: 512
Physical Size: 894.3GB
Position: shared
Checksum Compatibility: advanced_zoned
Aggregate: -
Plex: -This UID can be matched with the device UID displayed in the
‘storage devices’ tab for the ESX host
```



在 ESXi 或 KVM shell 中，您可以输入以下命令来闪烁给定物理磁盘（由其 naa.unique-id 标识）的 LED。

ESXi

```
esxcli storage core device set -d <naa_id> -l=locator -L=<seconds>
```

KVM

```
cat /sys/block/<block_device_id>/device/wwid
```

使用软件 **RAID** 时出现多个驱动器故障

系统可能会遇到多个驱动器同时处于故障状态的情况。系统的行为取决于聚合 RAID 保护和故障驱动器的数量。

RAID4 聚合可以在一个磁盘故障中存活，RAID-DP 聚合可以在两个磁盘故障中存活，RAID-TEC 聚合可以在三个磁盘故障中存活。

如果故障磁盘数小于 RAID 类型支持的最大故障数，且备用磁盘可用，则自动启动重建过程。如果备用磁盘不可用，则聚合以降级状态提供数据，直至添加备用磁盘。

如果故障磁盘数大于 RAID 类型支持的最大故障数，则本地丛标记为故障，聚合状态降级。从驻留在 HA 合作伙伴上的第二个丛提供数据。这意味着节点 1 的任何 I/O 请求都通过集群互连端口 e0e（iSCSI）发送到物理上位于节点 2 上的磁盘。如果第二个丛也失败，则聚合将被标记为失败，并且数据不可用。

必须删除并重新创建失败的丛，才能恢复正确的数据镜像。请注意，导致数据聚合降级的多磁盘故障也会导致根聚合降级。ONTAP Select 使用根数据数据 (RDD) 分区架构将每个物理驱动器拆分为一个根分区和两个数据分区。因此，丢失一个或多个磁盘可能会影响多个聚合，包括本地根或远程根聚合的副本，以及本地数据聚合和远程数据聚合的副本。

以下输出示例中删除并重新创建了失败的 plex：

```
C3111E67::> storage aggregate plex delete -aggregate aggr1 -plex plex1
Warning: Deleting plex "plex1" of mirrored aggregate "aggr1" in a non-
shared HA configuration will disable its synchronous mirror protection and
disable
         negotiated takeover of node "sti-rx2540-335a" when aggregate
"aggr1" is online.
Do you want to continue? {y|n}: y
[Job 78] Job succeeded: DONE

C3111E67::> storage aggregate mirror -aggregate aggr1
Info: Disks would be added to aggregate "aggr1" on node "sti-rx2540-335a"
in the following manner:
      Second Plex
        RAID Group rg0, 5 disks (advanced_zoned checksum, raid_dp)
                                                              Usable
Physical
          Position   Disk                      Type           Size
```

```
Size
          ---------- ------------------------ ---------- --------
--------
          shared     NET-3.2                  SSD               -
-
          shared     NET-3.3                  SSD               -
-
          shared     NET-3.4                  SSD         208.4GB
208.4GB
          shared     NET-3.5                  SSD         208.4GB
208.4GB
          shared     NET-3.12                 SSD         208.4GB
208.4GB

      Aggregate capacity available for volume use would be 526.1GB.
      625.2GB would be used from capacity license.
Do you want to continue? {y|n}: y

C3111E67::> storage aggregate show-status -aggregate aggr1
Owner Node: sti-rx2540-335a
 Aggregate: aggr1 (online, raid_dp, mirrored) (advanced_zoned checksums)
  Plex: /aggr1/plex0 (online, normal, active, pool0)
   RAID Group /aggr1/plex0/rg0 (normal, advanced_zoned checksums)
                                                                   Usable
Physical
     Position Disk                      Pool Type     RPM     Size
Size Status
     -------- ------------------------- ---- ----- ------ --------
-------- ----------
     shared   NET-1.1                     0   SSD        -  205.1GB
447.1GB (normal)
     shared   NET-1.2                     0   SSD        -  205.1GB
447.1GB (normal)
     shared   NET-1.3                     0   SSD        -  205.1GB
447.1GB (normal)
     shared   NET-1.10                    0   SSD        -  205.1GB
447.1GB (normal)
     shared   NET-1.11                    0   SSD        -  205.1GB
447.1GB (normal)
  Plex: /aggr1/plex3 (online, normal, active, pool1)
   RAID Group /aggr1/plex3/rg0 (normal, advanced_zoned checksums)
                                                                   Usable
Physical
     Position Disk                      Pool Type     RPM     Size
Size Status
     -------- ------------------------- ---- ----- ------ --------
```

```
-------- ----------
     shared   NET-3.2                    1   SSD        -  205.1GB
447.1GB (normal)
     shared   NET-3.3                    1   SSD        -  205.1GB
447.1GB (normal)
     shared   NET-3.4                    1   SSD        -  205.1GB
447.1GB (normal)
     shared   NET-3.5                    1   SSD        -  205.1GB
447.1GB (normal)
     shared   NET-3.12                   1   SSD        -  205.1GB
447.1GB (normal)
10 entries were displayed..
```

要测试或模拟一个或多个驱动器故障，请使用 `storage disk fail -disk NET-x.y -immediate` 命令。如果系统中有备件，聚合体将开始重建。您可以使用命令 `storage aggregate show` 检查重建的状态。您可以使用 ONTAP Deploy 删除模拟故障驱动器。请注意，ONTAP 已将驱动器标记为 `Broken`。驱动器实际上并未损坏，可以使用 ONTAP Deploy 重新添加。要擦除"损坏"标签，请在 ONTAP Select CLI 中输入以下命令：

```
set advanced
disk unfail -disk NET-x.y -spare true
disk show -broken
```

最后一个命令的输出应为空。

虚拟化 **NVRAM**

NetApp FAS 系统传统上配备物理 NVRAM PCI 卡。此卡是包含非易失性闪存的高性能卡，可显著提高写入性能。它通过授予 ONTAP 立即确认传入写入回客户端的能力来实现这一点。它还可以在称为 destaging 的过程中安排将修改的数据块移动回较慢的存储介质。

商品系统通常不配备此类设备。因此，NVRAM 卡的功能已被虚拟化并放置在 ONTAP Select 系统引导磁盘上的分区中。正是出于这个原因，实例的系统虚拟磁盘的放置非常重要。

ONTAP Select vSAN 和外部阵列配置

虚拟 NAS (vNAS) 部署支持虚拟 SAN (vSAN) 上的 ONTAP Select 集群、某些 HCI 产品和外部阵列类型的数据存储。这些配置的底层基础设施提供了数据存储弹性。

最低要求是，您使用的虚拟机管理程序（受支持的 Linux 主机上的 VMware ESXi 或 KVM）支持底层配置。如果虚拟机管理程序是 ESXi，则应在相应的 VMware HCL 上列出。

vNAS 架构

vNAS 命名法用于所有不使用 DAS 的设置。对于多节点 ONTAP Select 集群，这包括同一 HA 对中的两个 ONTAP Select 节点共享单个数据存储（包括 vSAN 数据存储）的架构。节点也可以安装在来自相同共享外部阵列的单独数据存储上。这允许阵列端存储效率降低整个 ONTAP Select HA 对的整体占用空间。ONTAP Select

vNAS 解决方案的架构与带有本地 RAID 控制器的 DAS 上的 ONTAP Select 非常相似。也就是说，每个 ONTAP Select 节点继续拥有其 HA 合作伙伴数据的副本。ONTAP 存储效率策略具有节点范围。因此，阵列端存储效率是首选的，因为它们可以潜在地应用于来自两个 ONTAP Select 节点的数据集。

HA 对中的每个 ONTAP Select 节点也可能使用单独的外部阵列。这是使用带有外部存储的 ONTAP Select MetroCluster SDS 时的常见选择。

当为每个 ONTAP Select 节点使用单独的外部阵列时，这两个阵列提供与 ONTAP Select VM 相似的性能特性非常重要。

vNAS 架构与带有硬件 RAID 控制器的本地 DAS

vNAS 架构在逻辑上与具有 DAS 和 RAID 控制器的服务器架构最为相似。在这两种情况下，ONTAP Select 会消耗数据存储空间。该数据存储空间被划分为 VMDK，这些 VMDK 构成了传统的 ONTAP 数据聚合。ONTAP Deploy 确保 VMDK 在集群创建和存储添加操作期间大小正确并分配给正确的丛（在 HA 对的情况下）。

vNAS 和带有 RAID 控制器的 DAS 之间有两个主要区别。最直接的区别在于 vNAS 不需要 RAID 控制器。vNAS 假设底层外部阵列提供具有 RAID 控制器设置的 DAS 所能提供的数据持久性和弹性。第二个也是更微妙的区别与 NVRAM 性能有关。

vNAS NVRAM

ONTAP Select NVRAM 是 VMDK。这意味着 ONTAP Select 在块可寻址设备 (VMDK) 之上模拟字节可寻址空间（传统 NVRAM）。但是，NVRAM 的性能对 ONTAP Select 节点的整体性能至关重要。

对于带有硬件 RAID 控制器的 DAS 设置，硬件 RAID 控制器缓存充当 NVRAM 缓存，因为对 NVRAM VMDK 的所有写入都首先托管在 RAID 控制器缓存中。

对于 VNAS 架构，ONTAP Deploy 使用称为单实例数据日志记录 (SIDL) 的引导参数自动配置 ONTAP Select 节点。当此引导参数存在时，ONTAP Select 绕过 NVRAM 并将数据有效负载直接写入数据聚合。NVRAM 仅用于记录 WRITE 操作更改的块的地址。此功能的好处是它避免了双重写入：一次写入 NVRAM，第二次写入是在 NVRAM 被清空时。此功能仅对 vNAS 启用，因为对 RAID 控制器缓存的本地写入具有可忽略的额外延迟。

SIDL 功能不兼容所有 ONTAP Select 存储效率功能。可以使用以下命令在聚合级别禁用 SIDL 功能：

```
storage aggregate modify -aggregate aggr-name -single-instance-data
-logging off
```

如果关闭 SIDL 功能，写入性能将受到影响。在禁用该聚合中所有卷上的所有存储效率策略后，可以重新启用 SIDL 功能：

```
volume efficiency stop -all true -vserver * -volume * (all volumes in the
affected aggregate)
```

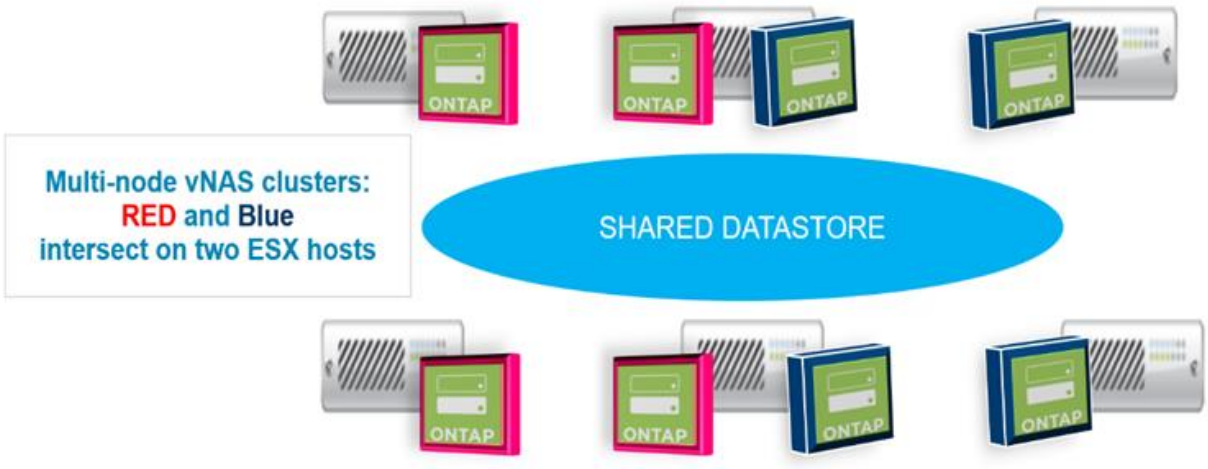
在 ESXi 上使用 vNAS 时并置 ONTAP Select 节点

ONTAP Select 包括对共享存储上的多节点 ONTAP Select 集群的支持。ONTAP Deploy 启用在同一 ESXi 主机上配置多个 ONTAP Select 节点，只要这些节点不属于同一集群。

此配置仅对 VNAS 环境（共享数据存储）有效。使用 DAS 存储时，每个主机不支持多个 ONTAP Select 实例，因为这些实例竞争相同的硬件 RAID 控制器。

ONTAP Deploy 可确保多节点 VNAS 集群的初始部署不会将来自同一集群的多个 ONTAP Select 实例放置到同一主机上。下图显示了在两台主机上交叉的两个四节点集群的正确部署示例。

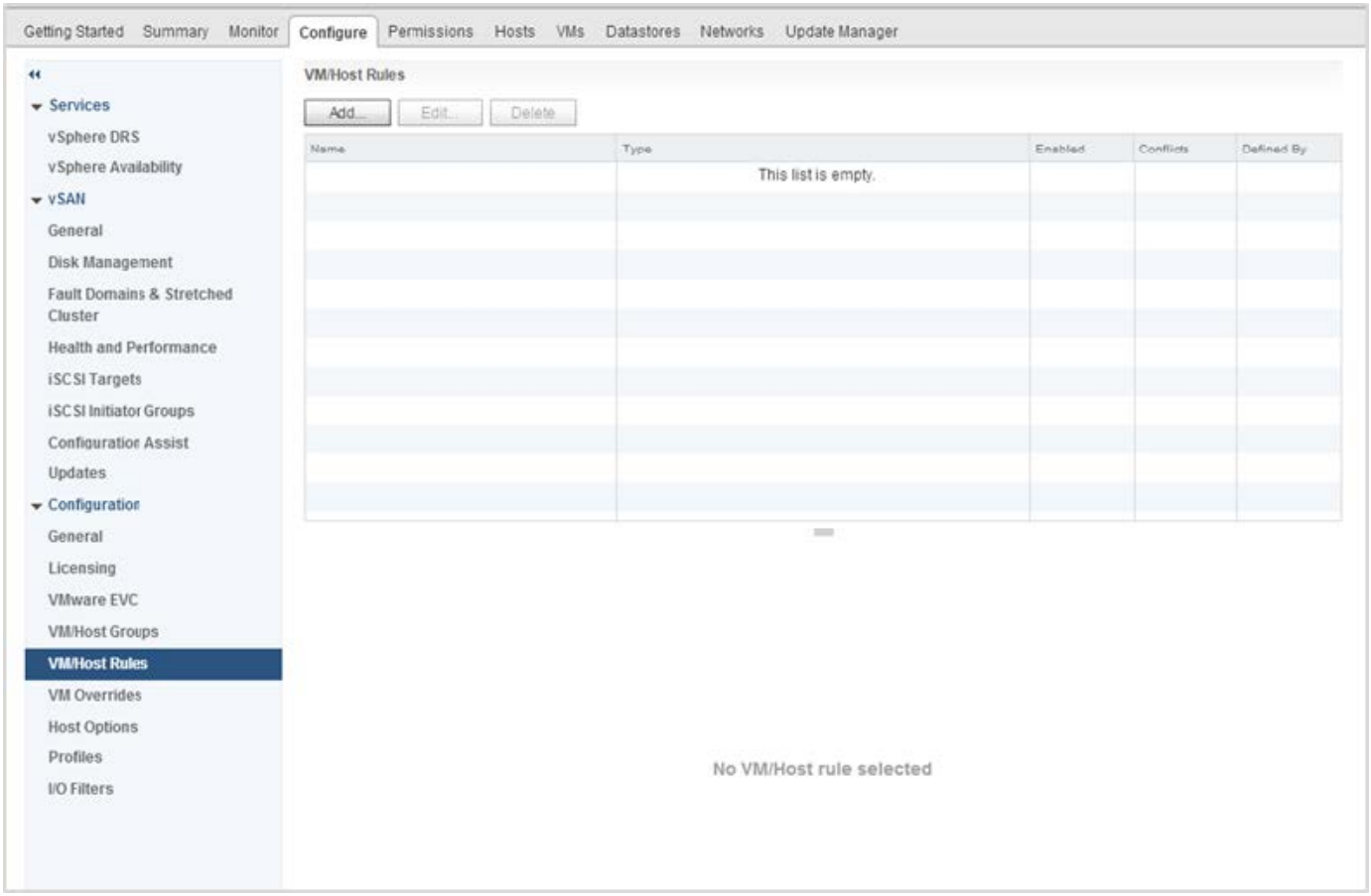
多节点 **VNAS** 集群的初始部署

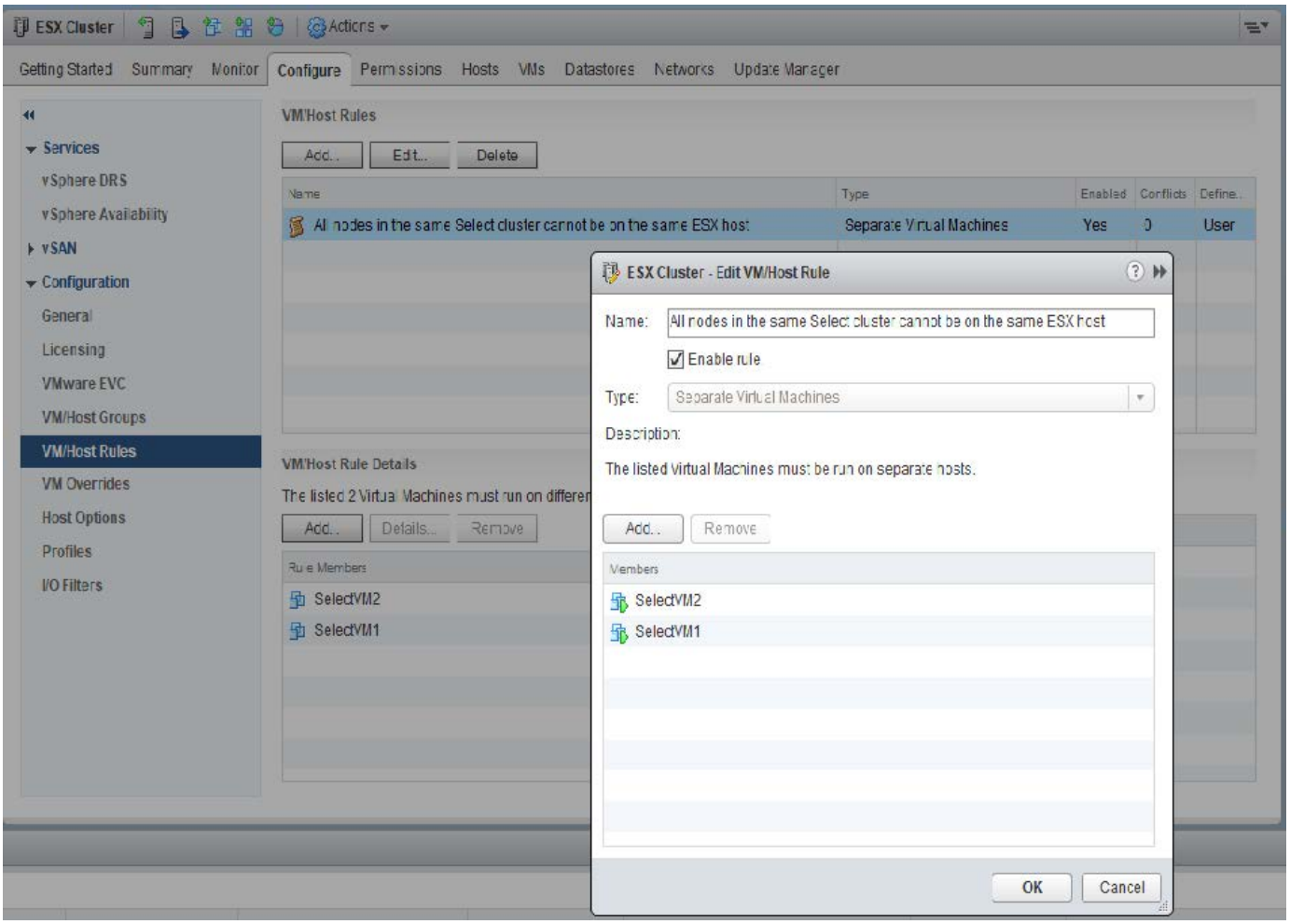


部署后，可以在主机之间迁移 ONTAP Select 节点。这可能会导致来自同一个集群的两个或多个 ONTAP Select 节点共享同一个底层主机的非优化和不受支持的配置。NetApp 建议手动创建 VM 反关联规则，以便 VMware 自动维护同一个集群的节点之间的物理分离，而不仅仅是来自同一个 HA 对的节点。

反关联规则要求在 ESXi 群集上启用 DRS。

请参阅以下示例，了解如何为 ONTAP Select VM 创建反关联规则。如果 ONTAP Select 集群包含多个 HA 对，则集群中的所有节点都必须包含在此规则中。


Getting Started
Summary
Monitor
Configure
Permissions
Hosts
VMs
Datastores
Networks
Update Manager
Services
vSphere DRS
vSphere Availability
vSAN
General
Disk Management
Fault Domains & Stretched Cluster
Health and Performance
iSCSI Targets
iSCSI Initiator Groups
Configuration Assist
Updates
Configuration
General
Licensing
VMware EVC
VM/Host Groups
VM/Host Rules
VM Overrides
Host Options
Profiles
I/O Filters
VM/Host Rules
Add...
Edit...
Delete
Name
Type
Enabled
Conflicts
Defined By
This list is empty.
No VM/Host rule selected



来自同一 ONTAP Select 集群的两个或多个 ONTAP Select 节点可能会因以下原因之一而位于同一 ESXi 主机上：

- 由于 VMware vSphere 许可证限制或未启用 DRS，DRS 不存在。
- DRS 反关联规则被绕过，因为 VMware HA 操作或管理员启动的虚拟机迁移优先。

ONTAP Deploy 不会主动监控 ONTAP Select VM 位置。但是，集群刷新操作会在 ONTAP Deploy 日志中反映此不受支持的配置：

增加 **ONTAP Select** 存储容量

ONTAP Deploy 可用于为 ONTAP Select 集群中的每个节点添加和许可其他存储。

ONTAP Deploy 中的存储添加功能是增加管理下的存储的唯一方法，不支持直接修改 ONTAP Select VM。下图显示了启动存储添加向导的"+"图标。

以下注意事项对于容量扩展操作的成功非常重要。添加容量需要现有许可证覆盖总空间量（现有容量加新容量）。导致节点超出其许可容量的存储添加操作会失败。应首先安装具有足够容量的新许可证。

如果将额外容量添加到现有的 ONTAP Select 聚合，则新存储池（数据存储区）的性能配置文件应类似于现有存储池（数据存储区）的性能配置文件。请注意，无法将非 SSD 存储添加到安装有类似 AFF 个性（启用闪存）的 ONTAP Select 节点。也不支持混合 DAS 和外部存储。

如果将本地连接的存储添加到系统以提供其他本地 (DAS) 存储池，则必须构建其他 RAID 组和 LUN（或 LUN）。与 FAS 系统一样，如果您要向相同的聚合添加新的空间，则应注意确保新的 RAID 组性能与原始 RAID 组相似。如果要创建新的聚合，则如果充分了解新聚合的性能影响，则新的 RAID 组布局可能会有所不同。

如果数据存储区的总大小不超过支持的最大数据存储区大小，则可以将新空间作为扩展区添加到同一数据存储区。将数据存储区扩展区添加到已安装 ONTAP Select 的数据存储区可以动态完成，并且不会影响 ONTAP Select 节点的操作。

如果 ONTAP Select 节点是 HA 对的一部分，则应考虑一些其他问题。

在 HA 对中，每个节点都包含来自其合作伙伴的数据的镜像副本。向节点 1 添加空间需要向其合作伙伴节点 2 添加相同数量的空间，以便将来自节点 1 的所有数据复制到节点 2。换句话说，作为节点 1 的容量添加操作的一部分添加到节点 2 的空间在节点 2 上不可见或不可访问。空间被添加到节点 2，以便在 HA 事件期间节点 1 数据得到完全保护。

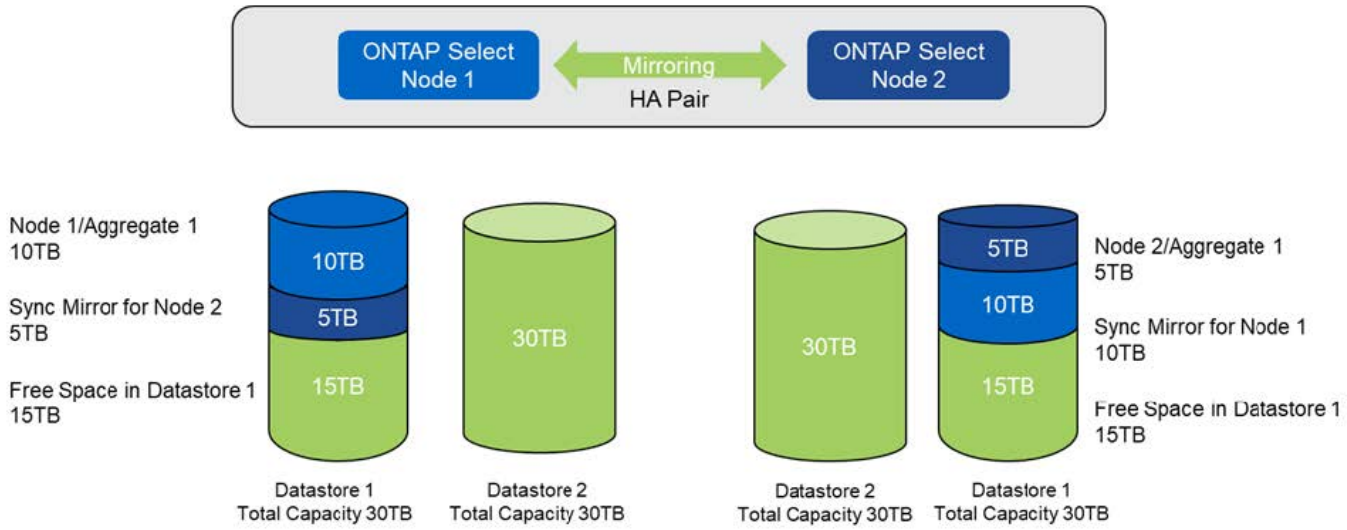
在性能方面还有一个额外的考虑因素。节点 1 上的数据会同步复制到节点 2。因此，节点 1 上新空间（数据存储）的性能必须与节点 2 上新空间（数据存储）的性能相匹配。换句话说，在两个节点上添加空间，但使用不同的驱动器技术或不同的 RAID 组大小，可能会导致性能问题。这是由于 RAID SyncMirror 操作用于在伙伴节点上维护数据副本。

为了增加 HA 对中两个节点上的用户可访问容量，必须执行两个存储添加操作，每个节点一个。每个存储添加操作都需要在两个节点上都有额外的空间。每个节点所需的总空间等于节点 1 所需的空间加上节点 2 所需的空间。

初始设置为两个节点，每个节点具有两个数据存储，每个数据存储中具有 30TB 的空间。ONTAP Deploy 创建一个双节点集群，每个节点从数据存储 1 消耗 10TB 的空间。ONTAP Deploy 为每个节点配置 5TB 的活动空间。

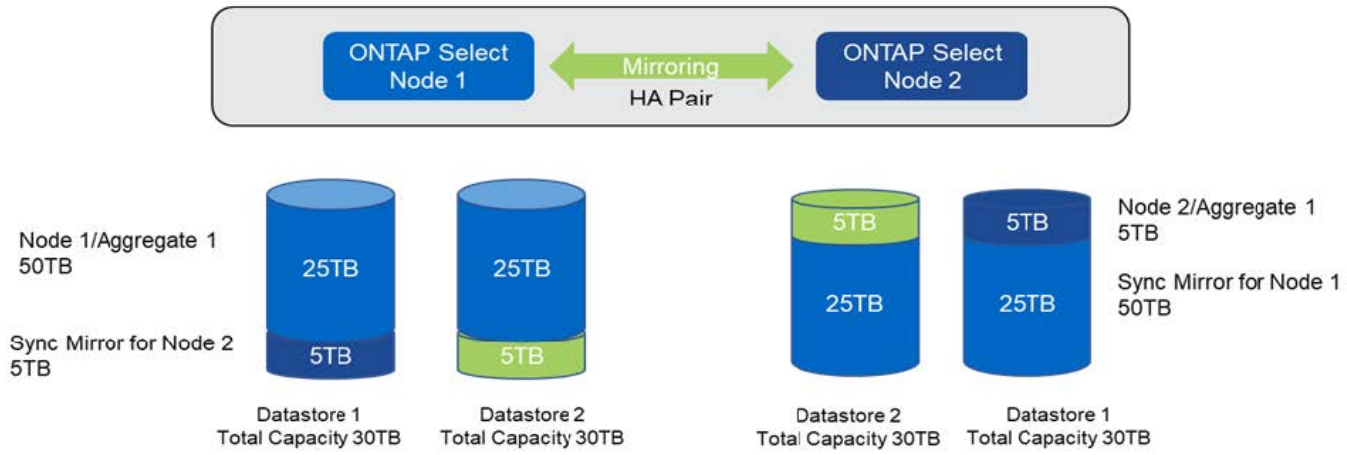
下图显示了节点 1 的单个存储添加操作的结果。ONTAP Select 仍然在每个节点上使用等量的存储（15TB）。但是，节点 1 具有比节点 2（5TB）更多的活动存储（10TB）。两个节点都受到完全保护，因为每个节点都托管另一个节点数据的副本。数据存储区 1 中还有其他可用空间，数据存储区 2 仍然完全可用。

容量分配：单次存储添加操作后的分配和可用空间



节点 1 上的两个额外的存储添加操作消耗了数据存储区 1 的其余部分和数据存储区 2 的一部分（使用容量上限）。第一次存储添加操作消耗了数据存储区 1 中剩余的 15TB 可用空间。下图显示了第二次存储添加操作的结果。此时，节点 1 有 50TB 的活动数据正在管理中，而节点 2 有原始的 5TB。

容量分配：节点 **1** 额外增加两个存储操作后的分配和可用空间



容量添加操作期间使用的最大 VMDK 大小为 16TB。集群创建操作期间使用的最大 VMDK 大小仍为 8TB。ONTAP Deploy 根据您的配置（单节点或多节点集群）和添加的容量创建正确大小的 VMDK。但是，每个 VMDK 的最大大小在集群创建操作期间不应超过 8TB，在存储添加操作期间不应超过 16TB。

通过软件 **RAID** 提高 **ONTAP Select** 的容量

类似地，storage-add 向导可用于使用软件 RAID 增加 ONTAP Select 节点的管理容量。该向导仅显示可用的 DAS SDD 驱动器，并且可以作为 RDM 映射到 ONTAP Select VM。

虽然可以将容量许可证增加单个 TB，但在使用软件 RAID 时，不可能将容量物理增加单个 TB。与将磁盘添加到 FAS 或 AFF 阵列类似，某些因素决定了单个操作中可以添加的最小存储量。

在 HA 对中，将存储添加到节点 1 需要在节点的 HA 对（节点 2）上也有相同数量的驱动器。本地驱动器和远程磁盘都由节点 1 上的一个存储添加操作使用。也就是说，远程驱动器用于确保节点 1 上的新存储在节点 2 上被复制和保护。为了在节点 2 上添加本地可用的存储，必须在两个节点上都提供单独的存储添加操作和单独的相同数量的驱动器。

ONTAP Select 将任何新驱动器分区为与现有驱动器相同的根分区、数据分区和数据分区。分区操作在创建新聚合期间或在现有聚合的扩展期间进行。每个磁盘上的根分区条带大小设置为与现有磁盘上的现有根分区大小相匹配。因此，两个相等的数据分区大小中的每一个都可以计算为磁盘总容量减去根分区大小除以 2。根分区条带大小是可变的，并在初始集群设置过程中按以下方式计算。所需的总根空间（单节点集群为 68GB，HA 对为 136GB）除以初始磁盘数量减去任何备用驱动器和奇偶校验驱动器。在添加到系统的所有驱动器上，根分区条带大小保持不变。

如果要创建新的聚合，所需的最小驱动器数量取决于 RAID 类型以及 ONTAP Select 节点是否是 HA 对的一部分。

如果向现有聚合添加存储，则需要考虑一些额外的注意事项。假设 RAID 组尚未达到最大限制，则可以将驱动器添加到现有 RAID 组。向现有 RAID 组添加主轴的传统 FAS 和 AFF 最佳实践也适用于此，在新主轴上创建热点是一个潜在的问题。此外，只能将数据分区大小相等或更大的驱动器添加到现有 RAID 组。如上所述，数据分区大小与驱动器原始大小不同。如果要添加的数据分区大于现有分区，则新驱动器的大小合适。换言之，每个新驱动器的一部分容量仍未利用。

也可以使用新驱动器作为现有聚合的一部分来创建新的 RAID 组。在这种情况下，RAID 组大小应与现有 RAID 组大小相匹配。

ONTAP Select 存储效率支持

ONTAP Select 提供的存储效率选项与 FAS 和 AFF 阵列上存在的存储效率选项类似。

使用全闪存 VSAN 或通用闪存阵列的 ONTAP Select 虚拟 NAS (vNAS) 部署应遵循使用非 SSD 直接连接存储 (DAS) 的 ONTAP Select 最佳实践。

只要您具有带有 SSD 驱动器和高级许可证的 DAS 存储，就会在新安装上自动启用类似 AFF 的个性。

具有类似 AFF 的个性，在安装过程中会自动启用以下内联 SE 功能：

- 内联零模式检测
- 卷内联重复数据删除
- 卷后台重复数据删除
- 自适应内联压缩
- 内联数据压缩
- 聚合内联重复数据删除
- 聚合后台重复数据删除

要验证 ONTAP Select 是否已启用所有默认存储效率策略，请在新创建的卷上运行以下命令：

```
<system name>::> set diag
Warning: These diagnostic commands are for use by NetApp personnel only.
Do you want to continue? {y|n}: y
twonode95IP15::*> sis config
Vserver:                                SVM1
Volume:                                 _export1_NFS_volume
Schedule                                -
Policy:                                 auto
Compression:                            true
Inline Compression:                     true
Compression Type:                       adaptive
Application IO Si                       8K
Compression Algorithm:                  lzopro
Inline Dedupe:                          true
Data Compaction:                        true
Cross Volume Inline Deduplication:      true
Cross Volume Background Deduplication:  true
```

对于从 9.6 及更高版本升级的 ONTAP Select，您必须使用高级许可证在 DAS SSD 存储上安装 ONTAP Select。此外，在使用 ONTAP Deploy 进行初始集群安装期间，必须选中*启用存储效率*复选框。当未满足先前条件时，在 ONTAP 升级后启用类似 AFF 的个性需要手动创建引导参数并重新启动节点。请联系技术支持以获得更多详细信息。

ONTAP Select 存储效率配置

下表总结了可用的各种存储效率选项，默认情况下启用，或默认情况下不启用，但建议使用，具体视媒体类型和软件许可证而定。

ONTAP Select 功能	**DAS SSD**（**premium** 或 **premium XL**[1]）	**DAS HDD**（所有许可证）	**vNAS**（所有许可证）
内联零检测	是（默认）	是 由用户按卷启用	是 由用户按卷启用
卷内联重复数据删除	是（默认）	不可用	不支持
32K 内联压缩（二次压缩）	是，由用户按卷启用。	是 由用户按卷启用	不支持
8K 内联压缩（自适应压缩）	是（默认）	是 用户按卷启用	不支持
后台压缩	不支持	是 用户按卷启用	是 由用户按卷启用
压缩扫描程序	是	是	是 由用户按卷启用
内联数据压缩	是（默认）	是 用户按卷启用	不支持
压缩扫描程序	是	是	不支持
聚合内联重复数据删除	是（默认）	不适用	不支持
卷后台重复数据删除	是（默认）	是 用户按卷启用	是 由用户按卷启用
聚合后台重复数据删除	是（默认）	不适用	不支持

[1]ONTAP Select 9.6 支持新的许可证（premium XL）和新的虚拟机大小（large）。但是，large 虚拟机仅支持使用软件 RAID 的 DAS 配置。硬件 RAID 和 vNAS 配置不支持 9.6 版本中的 large ONTAP Select VM。

关于 **DAS SSD** 配置升级行为的说明

升级到 ONTAP Select 9.6 或更高版本后，等待 `system node upgrade-revert show` 命令指示升级已完成，然后验证现有卷的存储效率值。

在升级到 ONTAP Select 9.6 或更高版本的系统上，在现有聚合或新创建的聚合上创建的新卷具有与在全新部署上创建的卷相同的行为。进行 ONTAP Select 代码升级的现有卷具有与新创建的卷相同的大部分存储效率策略，但有一些变化：

场景 **1**

如果在升级之前未在卷上启用存储效率策略，则：

- 具有 `space guarantee = volume`的卷未启用内联数据压缩、聚合内联重复数据删除和聚合后台重复数据删除。这些选项可在升级后启用。
- 具有 `space guarantee = none`的卷未启用后台压缩。此选项可在升级后启用。
- 升级后，现有卷上的存储效率策略设置为自动。

场景 **2**

如果在升级之前已在卷上启用了一些存储效率，则：

- 具有 `space guarantee = volume`的卷在升级后看不到任何差异。
- 具有 `space guarantee = none`的卷已打开聚合后台重复数据删除。
- 具有 `storage policy inline-only`的卷的策略设置为 auto。
- 具有用户定义的存储效率策略的卷不会更改策略，但具有 `space guarantee = none`的卷除外。这些卷已启用聚合后台重复数据删除。

网络连接

ONTAP Select 网络概念和特征

首先熟悉适用于 ONTAP Select 环境的一般网络概念。然后探索单节点和多节点集群的具体特征和可用选项。

物理网络

物理网络主要通过提供底层第二层交换基础设施来支持 ONTAP Select 集群部署。与物理网络相关的配置包括虚拟机管理程序主机和更广泛的交换网络环境。

主机 **NIC** 选项

每个 ONTAP Select 虚拟机监控程序主机必须配置两个或四个物理端口。您选择的确切配置取决于几个因素，包括：

- 集群是否包含一个或多个 ONTAP Select 主机
- 使用何种虚拟机管理程序操作系统

- 如何配置虚拟交换机
- 是否将 LACP 与链接一起使用

物理交换机配置

您必须确保物理交换机的配置支持 ONTAP Select 部署。物理交换机与基于虚拟机管理程序的虚拟交换机集成。您选择的确切配置取决于几个因素。主要考虑因素包括：

- 您将如何保持内部和外部网络之间的分离？
- 您是否会在数据和管理网络之间保持分离？
- 如何配置第二层 VLAN？

逻辑网络

ONTAP Select 使用两个不同的逻辑网络，根据类型分离流量。具体而言，流量可以在集群内的主机之间流动，也可以流向存储客户端和集群外的其他机器。虚拟机管理程序管理的虚拟交换机有助于支持逻辑网络。

内部网络

使用多节点集群部署，各个 ONTAP Select 节点使用隔离的"内部"网络进行通信。此网络未在 ONTAP Select 集群中的节点之外公开或可用。

内部网络仅存在于多节点集群中。

内部网络具有以下特点：

- 用于处理 ONTAP 集群内流量，包括：
  - 集群
  - 高可用性互连 (HA-IC)
  - RAID 同步镜像 (RSM)
- 基于 VLAN 的单层二层网络
- 静态 IP 地址由 ONTAP Select 分配：
  - 仅限 IPv4
  - 未使用 DHCP
  - 链路本地地址
- 默认情况下，MTU 大小为 9000 字节，可以在 7500-9000 范围内（含）进行调整

外部网络

外部网络处理 ONTAP Select 集群的节点与外部存储客户端以及其他机器之间的流量。外部网络是每个集群部署的一部分，具有以下特点：

- 用于处理 ONTAP 流量，包括：
  - 数据 (NFS、CIFS、iSCSI)

- 管理（集群和节点；可选 SVM）
- 集群间（可选）

- 可选地支持 VLAN：
  - 数据端口组
  - 管理端口组
- 根据管理员的配置选项分配的 IP 地址：
  - IPv4 或 IPv6
- 默认情况下，MTU 大小为 1500 字节（可调整）

外部网络存在于各种规模的集群中。

虚拟机网络环境

虚拟机管理程序主机提供多种网络功能。

ONTAP Select 依赖于通过虚拟机公开的以下功能：

虚拟机端口
: 有几个端口可供 ONTAP Select 使用。它们是根据几个因素分配和使用的，包括集群的大小。

虚拟交换机
: 虚拟机监控程序环境中的虚拟交换机软件，无论是 vSwitch（VMware）还是 Open vSwitch（KVM），都会将虚拟机暴露的端口与物理以太网 NIC 端口连接起来。您必须根据您的环境为每个 ONTAP Select 主机配置 vSwitch。

ONTAP Select 单节点和多节点网络配置

ONTAP Select 支持单节点和多节点网络配置。

单节点网络配置

单节点 ONTAP Select 配置不需要 ONTAP 内部网络，因为没有集群、HA 或镜像流量。

与 ONTAP Select 产品的多节点版本不同，每个 ONTAP Select VM 都包含三个虚拟网络适配器，分别提供给 ONTAP 网络端口 e0a、e0b 和 e0c。

这些端口用于提供以下服务：管理、数据和集群间 LIF。

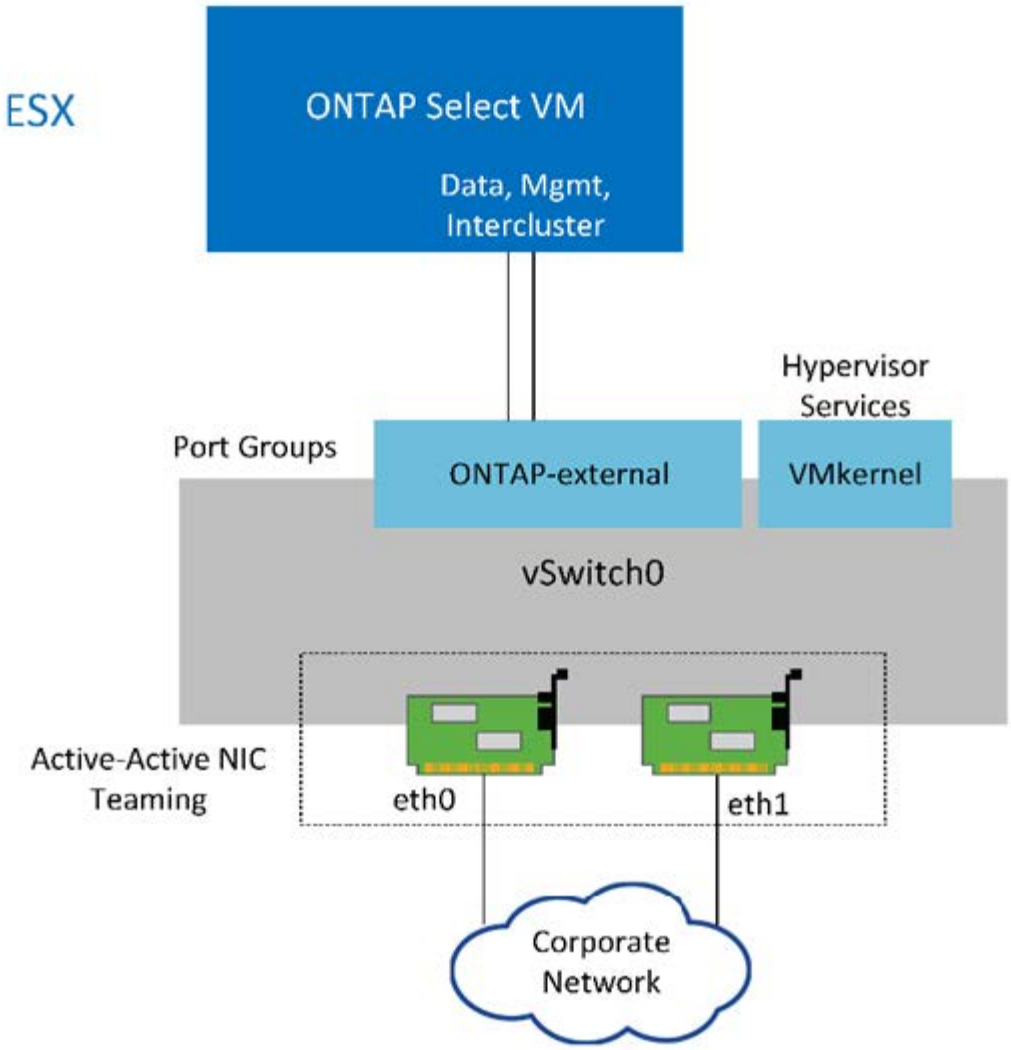
KVM

您可以将 ONTAP Select 部署为单节点集群。虚拟机监控程序主机包括一个提供对外部网络访问的虚拟交换机。

ESXi

下图显示了这些端口与底层物理适配器之间的关系。图中显示了 ESXi 虚拟机监控程序上的一个 ONTAP Select 群集节点。

单节点 **ONTAP Select** 集群的网络配置



即使两个适配器足以用于单节点集群，NIC 分组仍然是必需的。

LIF 分配

如本文档的多节点 LIF 分配部分所述，ONTAP 使用 IPspaces 将集群网络流量与数据和管理流量分开。此平台的单节点变体不包含集群网络。因此，集群 IPspace 中不存在任何端口。

集群和节点管理 LIF 在 ONTAP Select 集群设置期间自动创建。您可以在部署后创建剩余的 LIF。

管理和数据 LIF（e0a、e0b 和 e0c）

ONTAP 端口 e0a、e0b 和 e0c 被委派为传输以下类型流量的 LIF 的候选端口：

- SAN/NAS 协议流量（CIFS、NFS 和 iSCSI）
- 集群、节点和 SVM 管理流量
- 集群间流量（SnapMirror 和 SnapVault）

多节点网络配置

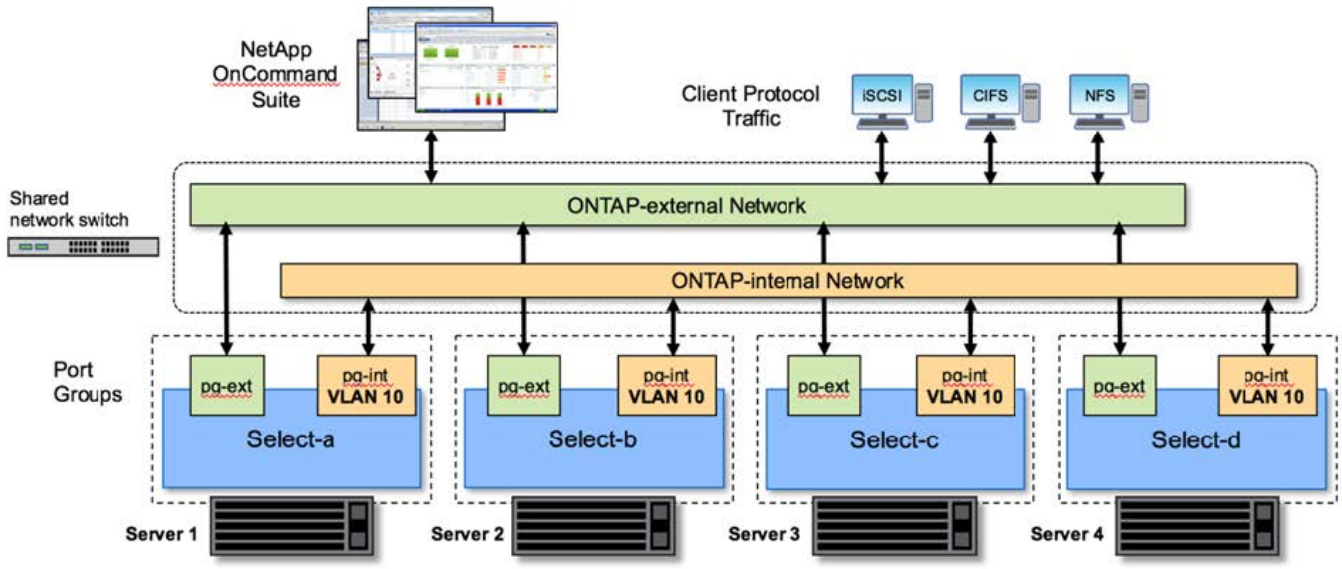
多节点 ONTAP Select 网络配置由两个网络组成。

它们是内部网络，负责提供集群和内部复制服务，以及外部网络，负责提供数据访问和管理服务。在这两个网络中流动的流量的端到端隔离对于构建适合集群弹性的环境极其重要。

下图显示了这些网络，其中显示了在 VMware vSphere 平台上运行的四节点 ONTAP Select 集群。六个、八个、十个和十二个节点集群具有相似的网络布局。

每个 ONTAP Select 实例驻留在单独的物理服务器上。内部和外部流量使用单独的网络端口组进行隔离，这些网络端口组分配给每个虚拟网络接口，并允许群集节点共享相同的物理交换机基础设施。

ONTAP Select 多节点集群网络配置概述



每个 ONTAP Select VM 包含七个虚拟网络适配器，作为一组七个网络端口（e0a 到 e0g）呈现给 ONTAP。虽然 ONTAP 将这些适配器视为物理 NIC，但它们实际上是虚拟的，并通过虚拟化网络层映射到一组物理接口。因此，每个托管服务器不需要六个物理网络端口。

不支持向 ONTAP Select VM 中添加虚拟网络适配器。

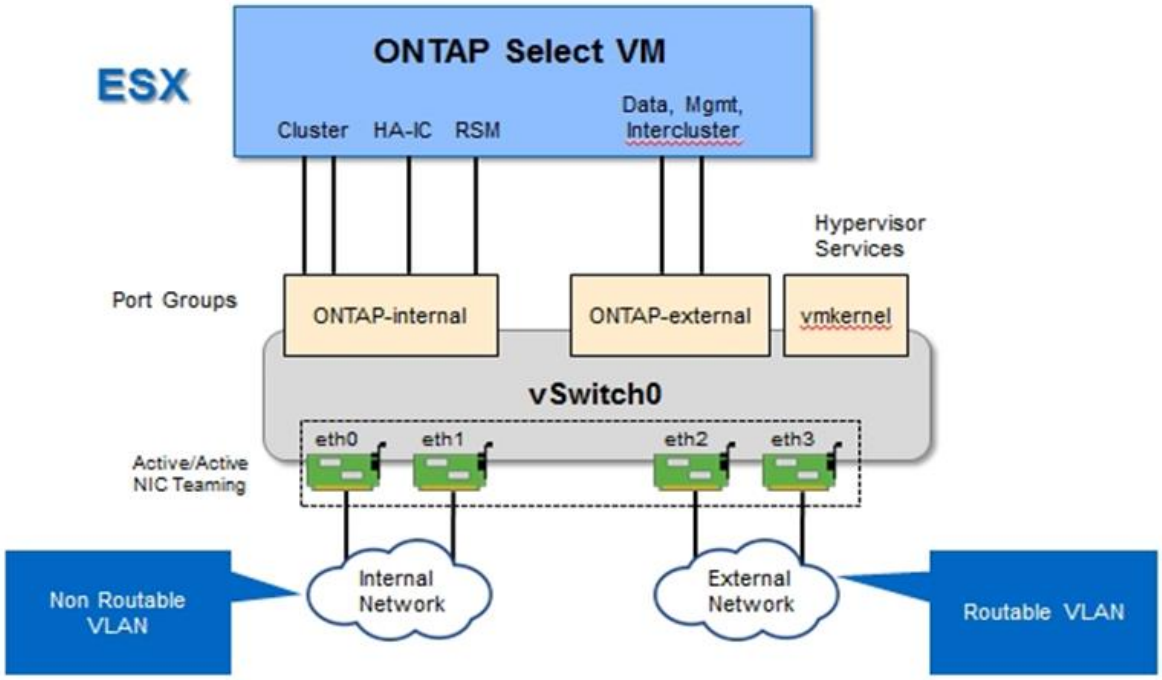
这些端口已预配置为提供以下服务：

- e0a、e0b 和 e0g。管理和数据 LIF
- e0c，e0d。集群网络 LIF
- e0e。RSM
- e0f。HA 互连

端口 e0a、e0b 和 e0g 位于外部网络上。虽然端口 e0c 到 e0f 执行几种不同的功能，但它们共同构成了内部 Select 网络。在做出网络设计决策时，应将这些端口放置在单个二层网络上。无需将这些虚拟适配器分离到不同的网络中。

下图显示了这些端口与底层物理适配器之间的关系，其中描绘了 ESXi 虚拟机监控程序上的一个 ONTAP Select 群集节点。

作为多节点 **ONTAP Select** 集群一部分的单个节点的网络配置



隔离不同物理网卡的内部和外部流量可防止对网络资源的访问不足，从而将延迟引入系统。此外，通过 NIC 组进行聚合允许 ONTAP Select 集群节点在单个网络适配器发生故障时继续访问网络。

> 外部网络和内部网络端口组均以对称方式包含所有四个 NIC 适配器。外部网络端口组中的活动端口是内部网络中的备用端口。相反，内部网络端口组中的活动端口是外部网络端口组中的备用端口。

LIF 分配

随着 IPspaces 的引入，ONTAP 端口角色已被弃用。与 FAS 阵列一样，ONTAP Select 集群同时包含默认 IPspace 和集群 IPspace。通过将网络端口 e0a、e0b 和 e0g 放入默认 IPspace，并将端口 e0c 和 e0d 放入集群 IPspace，这些端口基本上被隔离，无法托管不属于的 LIF。ONTAP Select 集群中的其余端口通过提供内部服务的接口的自动分配来消耗。它们不会通过 ONTAP shell 暴露，就像 RSM 和 HA 互连接口一样。

> 并非所有 LIF 都通过 ONTAP 命令外壳可见。HA 互连和 RSM 接口对 ONTAP 隐藏，在内部用来提供各自的服务。

以下各节将详细解释网络端口和 LIF。

管理和数据 **LIF**（**e0a**、**e0b** 和 **e0g**）

ONTAP 端口 e0a、e0b 和 e0g 被委派为承载以下类型流量的 LIF 的候选端口：

- SAN/NAS 协议流量（CIFS、NFS 和 iSCSI）
- 集群、节点和 SVM 管理流量

- 集群间流量（SnapMirror 和 SnapVault）

> 集群和节点管理 LIF 在 ONTAP Select 集群设置期间自动创建。您可以在部署后创建剩余的 LIF。

集群网络 **LIF**（**e0c**、**e0d**）

ONTAP 端口 e0c 和 e0d 被委派为集群接口的主端口。在每个 ONTAP Select 集群节点内，在 ONTAP 设置期间使用链接本地 IP 地址（169.254.x.x）自动生成两个集群接口。

> 您不能为这些接口分配静态 IP 地址，也不应创建其他集群接口。

集群网络流量必须通过低延迟、非路由的第 2 层网络。由于集群吞吐量和延迟要求，您应该将 ONTAP Select 集群物理定位在附近（例如，多包、单个数据中心）。不支持通过跨 WAN 或相当大的地理距离分离 HA 节点来构建四个、六个、八个、十个或十二个节点的扩展集群配置。支持带有中介器的拉伸双节点配置。

有关详细信息，请参阅部分 "双节点拉伸 HA（MetroCluster SDS）最佳实践"。

> 为确保群集网络流量的最大吞吐量，已将此网络端口配置为使用巨型帧（7500 到 9000 MTU）。要进行正确的群集操作，请验证是否在所有为 ONTAP Select 群集节点提供内部网络服务的上游虚拟和物理交换机上启用了巨型帧。

RAID SyncMirror 流量 **(e0e)**

跨 HA 伙伴节点的块同步复制使用驻留在网络端口 e0e 上的内部网络接口进行。此功能在集群设置期间使用 ONTAP 配置的网络接口自动进行，无需管理员进行配置。

> 端口 e0e 由 ONTAP 保留用于内部复制流量。因此，端口和托管 LIF 在 ONTAP CLI 或 System Manager 中都不可见。此接口配置为使用自动生成的链路本地 IP 地址，您无法分配备用 IP 地址。此网络端口需要使用巨型帧（7500 至 9000 MTU）。

HA 互连 **(e0f)**

NetApp FAS 阵列使用专用硬件在 ONTAP 集群中的 HA 对之间传递信息。然而，软件定义的环境往往没有这种类型的设备（例如 InfiniBand 或 iWARP 设备），因此需要替代解决方案。虽然考虑了几种可能性，但对互连传输的 ONTAP 要求要求在软件中模拟此功能。因此，在 ONTAP Select 集群中，HA 互连的功能（传统上由硬件提供）已被设计到操作系统中，使用以太网作为传输机制。

每个 ONTAP Select 节点配置有 HA 互连端口 e0f。此端口托管 HA 互连网络接口，该接口负责两个主要功能：

- 在 HA 对之间镜像 NVRAM 的内容
- 在 HA 对之间发送/接收 HA 状态信息和网络心跳消息

HA 互连流量通过在以太网数据包中分层远程直接内存访问（RDMA）帧，使用单个网络接口流经此网络端口。

> 与 RSM 端口（e0e）一样，无论是从 ONTAP CLI 还是从 System Manager，用户都不会看到物理端口或托管网络接口。因此，您无法修改此接口的 IP 地址，也无法更改端口的状态。此网络端口需要使用巨型帧（7500 至 9000 MTU）。

ONTAP Select 内部和外部网络

ONTAP Select 内部和外部网络的特征。

ONTAP Select 内部网络

内部 ONTAP Select 网络仅存在于产品的多节点变体中，负责为 ONTAP Select 集群提供集群通信、HA 互连和同步复制服务。此网络包括以下端口和接口：

- *e0c，e0d。*承载集群网络 LIF
- *e0e.*承载 RSM LIF
- *e0f.*承载 HA 互连 LIF

该网络的吞吐量和延迟对于确定 ONTAP Select 集群的性能和弹性至关重要。集群安全需要网络隔离，并确保系统接口与其他网络流量保持独立。因此，此网络必须由 ONTAP Select 集群独家使用。

> 不支持将 Select 内部网络用于 Select 集群流量以外的流量，例如应用程序或管理流量。ONTAP 内部 VLAN 上不能有其他虚拟机或主机。

通过内部网络的网络数据包必须位于专用 VLAN 标记的第 2 层网络上。为此，可使用以下方法之一：

- 将 VLAN 标记的端口组分配给内部虚拟 NIC（e0c 到 e0f）（VST 模式）
- 使用上游交换机提供的本征 VLAN，其中本征 VLAN 不用于任何其他流量（分配一个没有 VLAN ID 的端口组，即 EST 模式）

在所有情况下，内部网络流量的 VLAN 标记都是在 ONTAP Select VM 之外完成的。

> 仅支持 ESXi 标准和分布式 vSwitches。不支持其他虚拟交换机或 ESXi 主机之间的直接连接。必须完全打开内部网络；不支持 NAT 或防火墙。

在 ONTAP Select 集群中，使用称为端口组的虚拟二层网络对象将内部流量和外部流量分开。正确的 vSwitch 分配这些端口组非常重要，特别是对于负责提供集群、HA 互连和镜像复制服务的内部网络。这些网络端口的网络带宽不足会导致性能下降，甚至影响集群节点的稳定性。因此，四个、六个、八个、十个和十二个节点集群要求内部 ONTAP Select 网络使用 10Gb 连接；不支持 1Gb NIC。但是，可以对外部网络进行权衡，因为限制传入 ONTAP Select 集群的数据流量不会影响其可靠运行的能力。

双节点集群可以使用四个 1Gb 端口进行内部流量，也可以使用一个 10Gb 端口，而不是四节点集群所需的两个 10Gb 端口。在条件阻止服务器配备四个 10Gb NIC 卡的环境中，两个 10Gb NIC 卡可用于内部网络，两个 1Gb NIC 可用于外部 ONTAP 网络。

内部网络验证和故障排除

可以使用网络连接检查器功能验证多节点群集中的内部网络。可以从运行 `network connectivity-check start` 命令的 Deploy CLI 调用此函数。

运行以下命令来查看测试的输出：

```
network connectivity-check show --run-id X (X is a number)
```

此工具仅适用于多节点 Select 集群中的内部网络故障排除。该工具不应用于对单节点集群（包括 vNAS 配置）、ONTAP Deploy 到 ONTAP Select 连接或客户端连接问题进行故障排除。

集群创建向导（ONTAP Deploy UI 的一部分）包括内部网络检查器，作为创建多节点集群期间可用的可选步骤。鉴于内部网络在多节点集群中扮演的重要角色，使此步骤成为集群创建工作流的一部分可提高集群创建操作的成功率。

从 ONTAP Deploy 2.10 开始，内部网络使用的 MTU 大小可以设置在 7,500 到 9,000 之间。网络连接检查器还可用于测试 7,500 至 9,000 之间的 MTU 大小。默认 MTU 值设置为虚拟网络交换机的值。如果环境中存在像 VXLAN 这样的网络覆盖，则必须用较小的值替换该默认值。

ONTAP Select 外部网络

ONTAP Select 外部网络负责集群的所有出站通信，因此存在于单节点和多节点配置中。虽然此网络没有内部网络严格定义的吞吐量要求，但管理员应注意不要在客户端和 ONTAP VM 之间造成网络瓶颈，因为性能问题可能被误认为是 ONTAP Select 问题。

> 与内部流量类似，外部流量可以在 vSwitch 层（VST）和外部交换层（EST）进行标记。此外，ONTAP Select VM 本身还可以在称为 VGT 的过程中对外部流量进行标记。有关更多详细信息，请参阅 "数据和管理流量分离" 部分。

下表重点介绍了 ONTAP Select 内部和外部网络之间的主要区别。

内部与外部网络快速参考

问题描述	内部网络	外部网络
网络服务	集群 HA/IC RAID SyncMirror (RSM)	数据管理集群间（SnapMirror 和 SnapVault）
网络隔离	必填项	可选
帧大小 (MTU)	7,500 至 9,000	1,500（默认）9,000（支持）
IP 地址分配	自动生成	用户定义
DHCP 支持	否	否

NIC 组队

为了确保内部和外部网络同时具有提供高性能和容错所需的必要带宽和弹性特性，建议使用物理网络适配器组合。支持具有单个 10Gb 链路的双节点集群配置。但是，NetApp 推荐的最佳做法是在 ONTAP Select 集群的内部和外部网络上使用 NIC 组合。

MAC 地址生成

分配给所有 ONTAP Select 网络端口的 MAC 地址由包含的部署实用程序自动生成。该实用程序使用特定于 NetApp 的平台特定组织唯一标识符 (OUI)，以确保与 FAS 系统没有冲突。然后将此地址的副本存储在 ONTAP Select 安装虚拟机 (ONTAP Deploy) 的内部数据库中，以防止在未来的节点部署过程中意外重新分配。在任何时候，管理员都不应修改网络端口的分配 MAC 地址。

支持的 ONTAP Select 网络配置

选择最佳硬件并配置网络，以优化性能和故障恢复能力。

服务器供应商了解客户有不同的需求，选择至关重要。因此，在购买物理服务器时，在做出网络连接决策时有许多选项可供选择。大多数商品系统附带各种 NIC 选项，提供单端口和多端口选项，具有不同的速度和吞吐量排列。这包括支持 VMware ESX 的 25Gb/s 和 40Gb/s NIC 适配器。

由于 ONTAP Select VM 的性能与底层硬件的特性直接相关，通过选择更高速的 NIC 来提高 VM 的吞吐量可以实现更高性能的集群和更好的整体用户体验。可以使用四个 10Gb NIC 或两个更高速的 NIC（25/40 Gb/s）来实现高性能网络布局。还支持许多其他配置。对于双节点集群，支持 4 x 1Gb 端口或 1 x 10Gb 端口。对于单节点集群，支持 2 x 1Gb 端口。

网络最低配置和建议配置

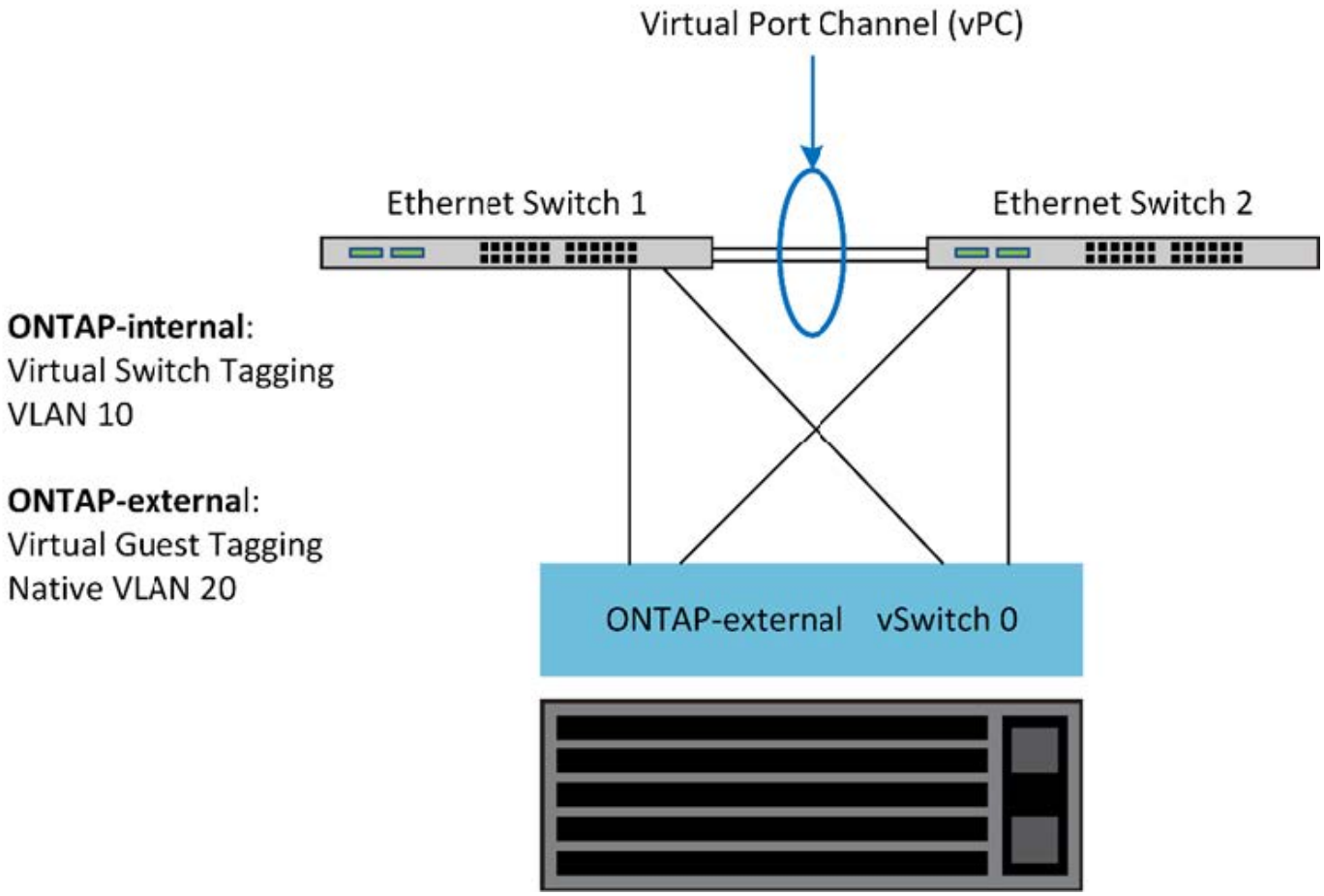
根据集群大小，支持多种以太网配置。

集群大小	最低要求	建议
单节点集群	2 x 1GbE	2 x 10GbE
双节点集群或 MetroCluster SDS	4 x 1GbE 或 1 x 10GbE	2 x 10GbE
四个、六个、八个、十个或十二个节点集群	2 x 10GbE	4 x 10GbE 或 2 x 25/40GbE

不支持在运行的群集上单链路拓扑和多链路拓扑之间进行转换，因为可能需要在每个拓扑所需的不同 NIC 组配置之间进行转换。

使用多个物理交换机的网络配置

当有足够的硬件可用时，由于增加了对物理交换机故障的保护，NetApp 建议使用下图所示的多交换机配置。



ONTAP Select VMware vSphere vSwitch 在 ESXi 上的配置

ONTAP Select vSwitch 配置和双网卡及四网卡配置的负载平衡策略。

ONTAP Select 支持使用标准和分布式 vSwitch 配置。分布式 vSwitches 支持链路聚合构造（LACP）。链路聚合是一种常见的网络结构，用于跨多个物理适配器聚合带宽。LACP 是供应商中立的标准。它为将物理网络端口组捆绑到单个逻辑通道中的网络端点提供了开放协议。ONTAP Select 可以使用配置为链路聚合组 (LAG) 的端口组。但是，NetApp 建议使用单个物理端口作为简单的上行链路（中继）端口，以避免 LAG 配置。在这些情况下，标准和分布式 vSwitches 的最佳实践是相同的。

本节描述在双网卡和四网卡配置中应使用的 vSwitch 配置和负载平衡策略。

为 ONTAP Select 配置端口组时，请遵循以下最佳做法；端口组级别的负载平衡策略是基于发起虚拟端口 ID 的路由。VMware 建议在连接到 ESXi 主机的交换机端口上将 STP 设置为 Portfast。

所有 vSwitch 配置都需要将至少两个物理网络适配器捆绑到一个 NIC 团队中。ONTAP Select 支持双节点集群的单个 10Gb 链路。但是，NetApp 建议使用 NIC 聚合来确保硬件冗余。

在 vSphere 服务器上，NIC 团队是用于将多个物理网络适配器捆绑到单个逻辑通道的聚合结构，允许在所有成员端口上共享网络负载。请务必记住，无需物理交换机的支持，即可创建 NIC 团队。负载平衡和故障转移策略可以直接应用于 NIC 团队，而 NIC 团队并不知道上游交换机的配置。在此情况下，策略仅适用于出站流量。

ONTAP Select 不支持静态端口通道。分布式 vSwitches 支持启用 LACP 的通道，但使用 LACP LAG 可能会导致 LAG 成员之间的负载分布不均匀。

对于单节点集群，ONTAP Deploy 将 ONTAP Select VM 配置为使用用于外部网络的端口组，以及用于集群和节点管理流量的相同端口组或可选的不同端口组。对于单节点集群，您可以将所需数量的物理端口作为活动适配器添加到外部端口组。

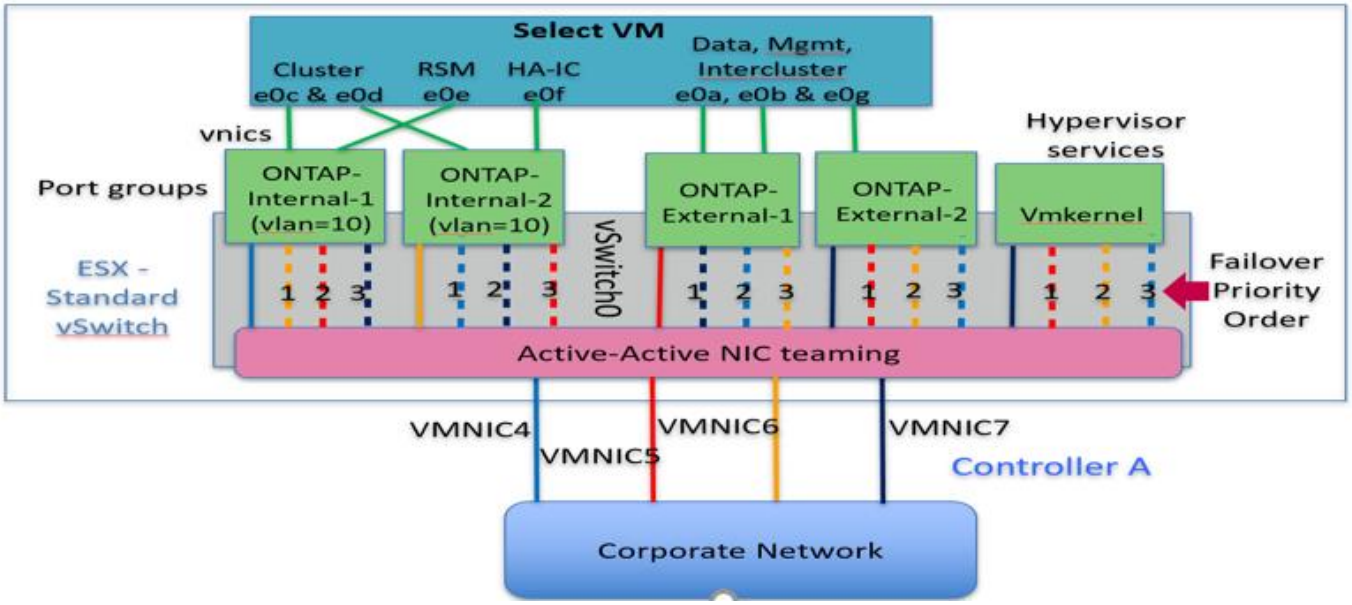
对于多节点集群，ONTAP Deploy 配置每个 ONTAP Select VM 为内部网络使用一个或两个端口组，外部网络单独使用一个或两个端口组。集群和节点管理流量可以使用与外部流量相同的端口组，也可以选择单独的端口组。集群和节点管理流量不能与内部流量共享同一个端口组。

ONTAP Select 支持最多四个 VMNIC。

标准或分布式 **vSwitch** 和每个节点四个物理端口

可以为多节点集群中的每个节点分配四个端口组。每个端口组具有一个活动物理端口和三个备用物理端口，如下图所示。

每个节点有四个物理端口的 **vSwitch**



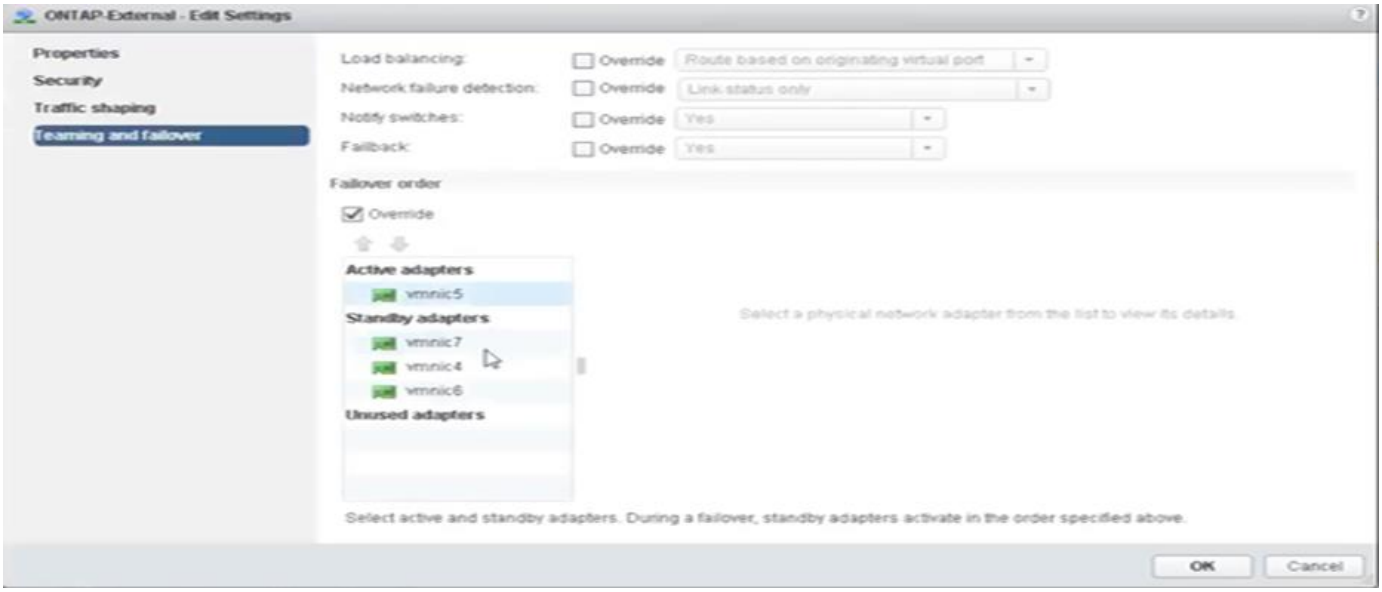
备用列表中端口的顺序很重要。下表提供了四个端口组之间的物理端口分布示例。

网络最低配置和建议配置

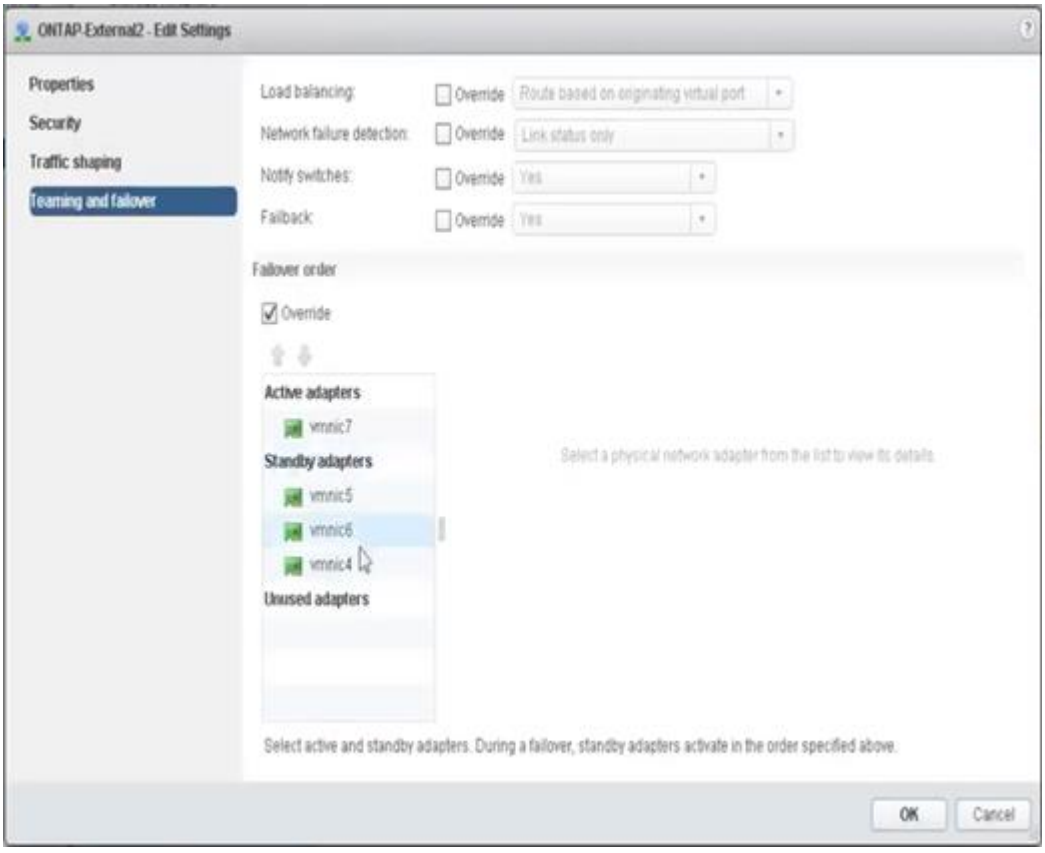
端口组	外部 **1**	外部 **2**	内部 **1**	内部 **2**
活动	vmnic0	vmnic1	vmnic2	vmnic3
待机 1	vmnic1	vmnic0	vmnic3	vmnic2
待机 2	vmnic2	vmnic3	vmnic0	vmnic1
待机 3	vmnic3	vmnic2	vmnic1	vmnic0

下图显示了来自 vCenter UI（ONTAP-External 和 ONTAP-External2）的外部网络端口组的配置。请注意，活动适配器来自不同的网卡。在此设置中，vmnic 4 和 vmnic 5 是同一物理网卡上的双端口，而 vmnic 6 和 vminc 7 同样是单独网卡上的双端口（本示例中不使用 vnmics 0 到 3）。备用适配器的顺序提供了一个分层故障转移，来自内部网络的端口是最后一个。备用列表中内部端口的顺序类似地在两个外部端口组之间交换。

第 1 部分：**ONTAP Select** 外部端口组配置



第 2 部分：**ONTAP Select** 外部端口组配置

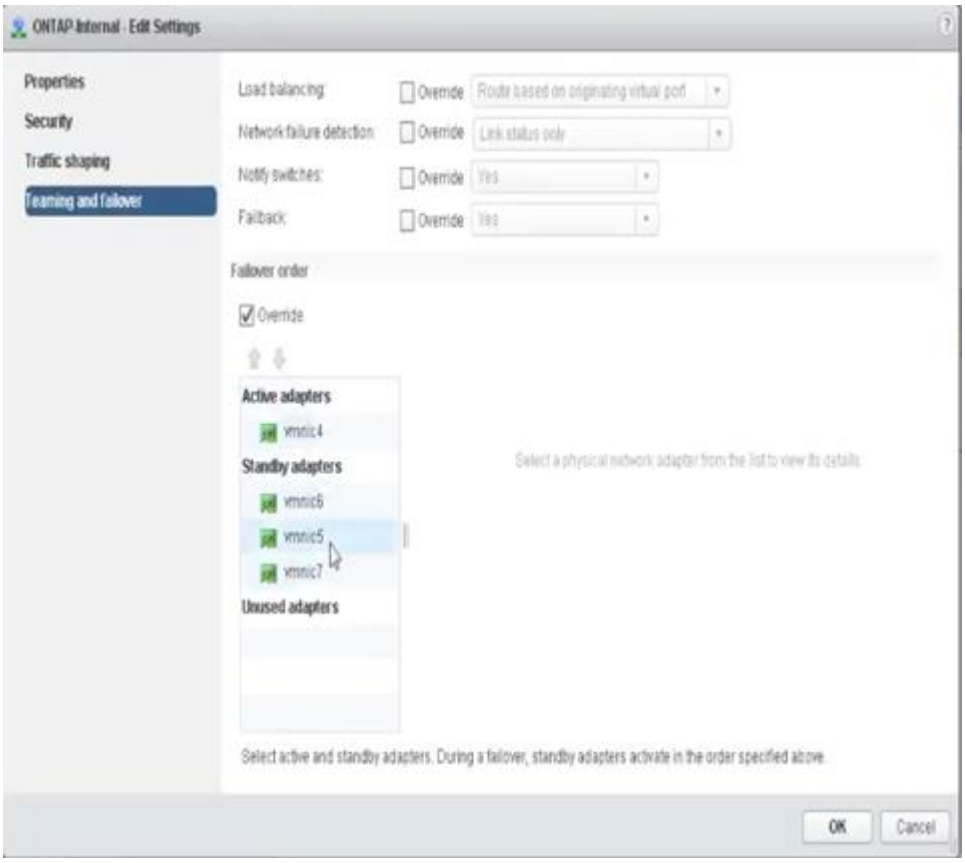


为了便于阅读，分配如下：

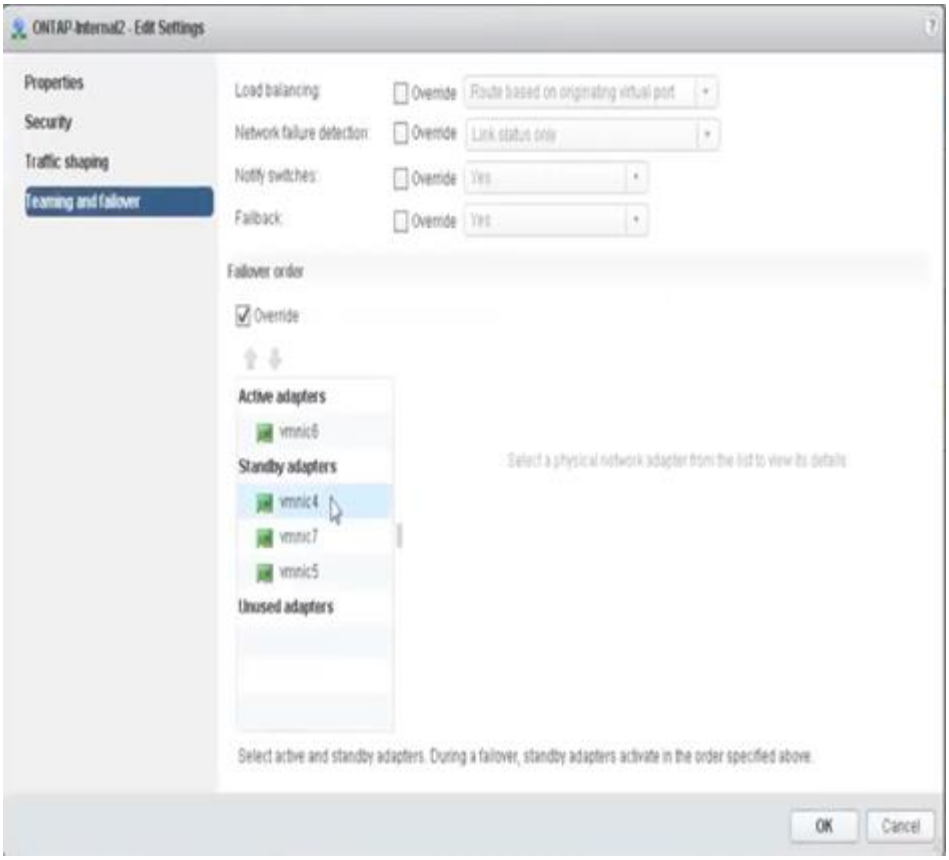
ONTAP-External	ONTAP-External2
活动适配器：vmnic5 备用适配器：vmnic7、vmnic4、vmnic6	活动适配器：vmnic7 备用适配器：vmnic5、vmnic6、vmnic4

下图显示了内部网络端口组（ONTAP-Internal 和 ONTAP-Internal2）的配置。请注意，活动适配器来自不同的网卡。在此设置中，vmnic 4 和 vmnic 5 是同一物理 ASIC 上的双端口，而 vmnic 6 和 vmnic 7 是独立 ASIC 上的类似双端口。备用适配器的顺序提供了分层故障切换，来自外部网络的端口是最后一个。备用列表中外部端口的顺序类似地在两个内部端口组之间交换。

第 **1** 部分：**ONTAP Select** 内部端口组配置



第 **2** 部分：**ONTAP Select** 内部端口组



为了便于阅读，分配如下：

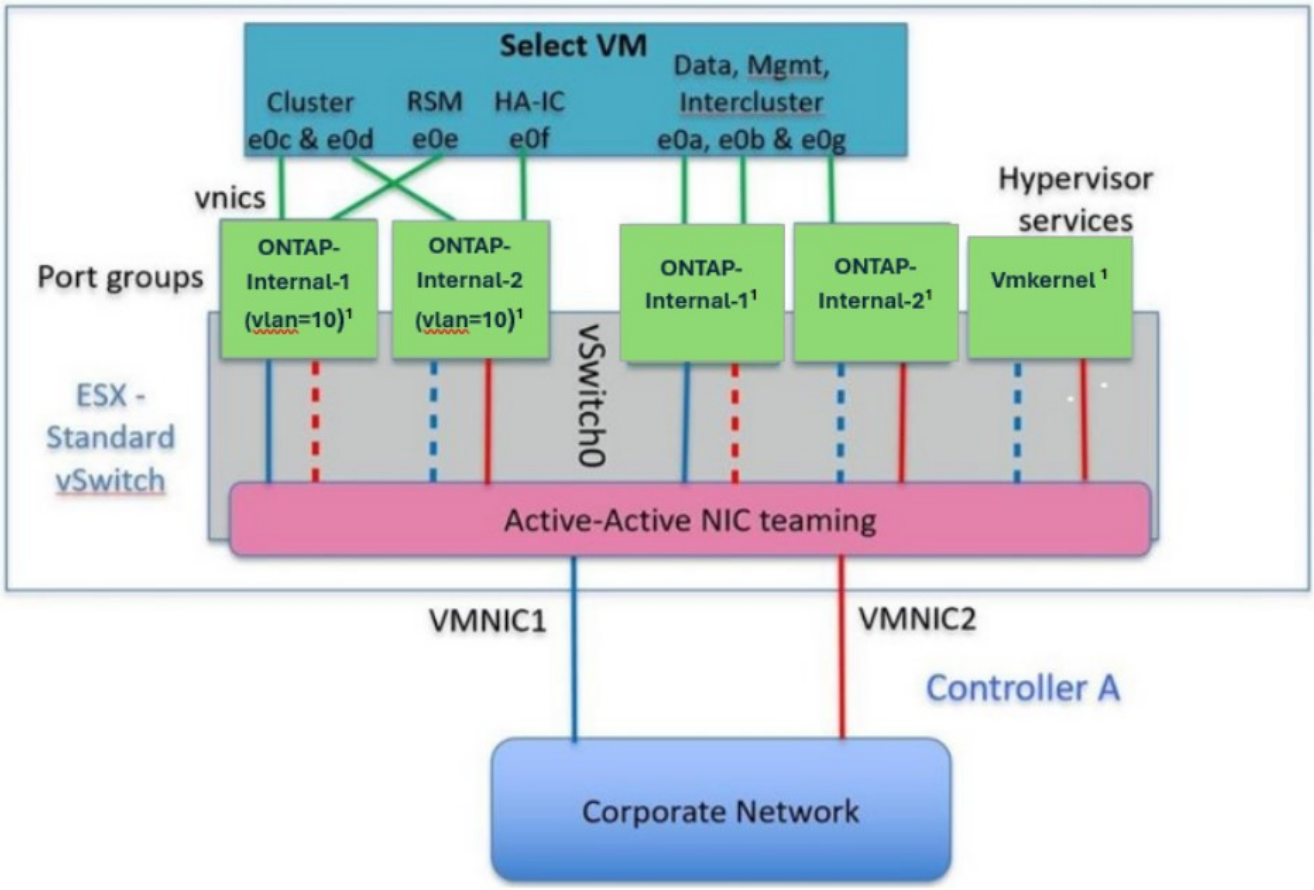
ONTAP-Internal	**ONTAP-Internal2**
活动适配器：vmnic4 备用适配器：vmnic6、vmnic5、vmnic7	活动适配器：vmnic6 备用适配器：vmnic4、vmnic7、vmnic5

标准或分布式 **vSwitch**，每个节点两个物理端口

当使用两个高速（25/40Gb）网卡时，建议的端口组配置在概念上与四个 10Gb 适配器的配置非常相似。即使仅使用两个物理适配器，也应使用四个端口组。端口组分配如下：

端口组	外部 **1 (e0a,e0b)**	内部 **1 (e0c,e0e)**	内部 **2 (e0d,e0f)**	外部 **2 (e0g)**
活动	vmnic0	vmnic0	vmnic1	vmnic1
备用	vmnic1	vmnic1	vmnic0	vmnic0

vSwitch 每个节点两个高速 **(25/40Gb)** 物理端口

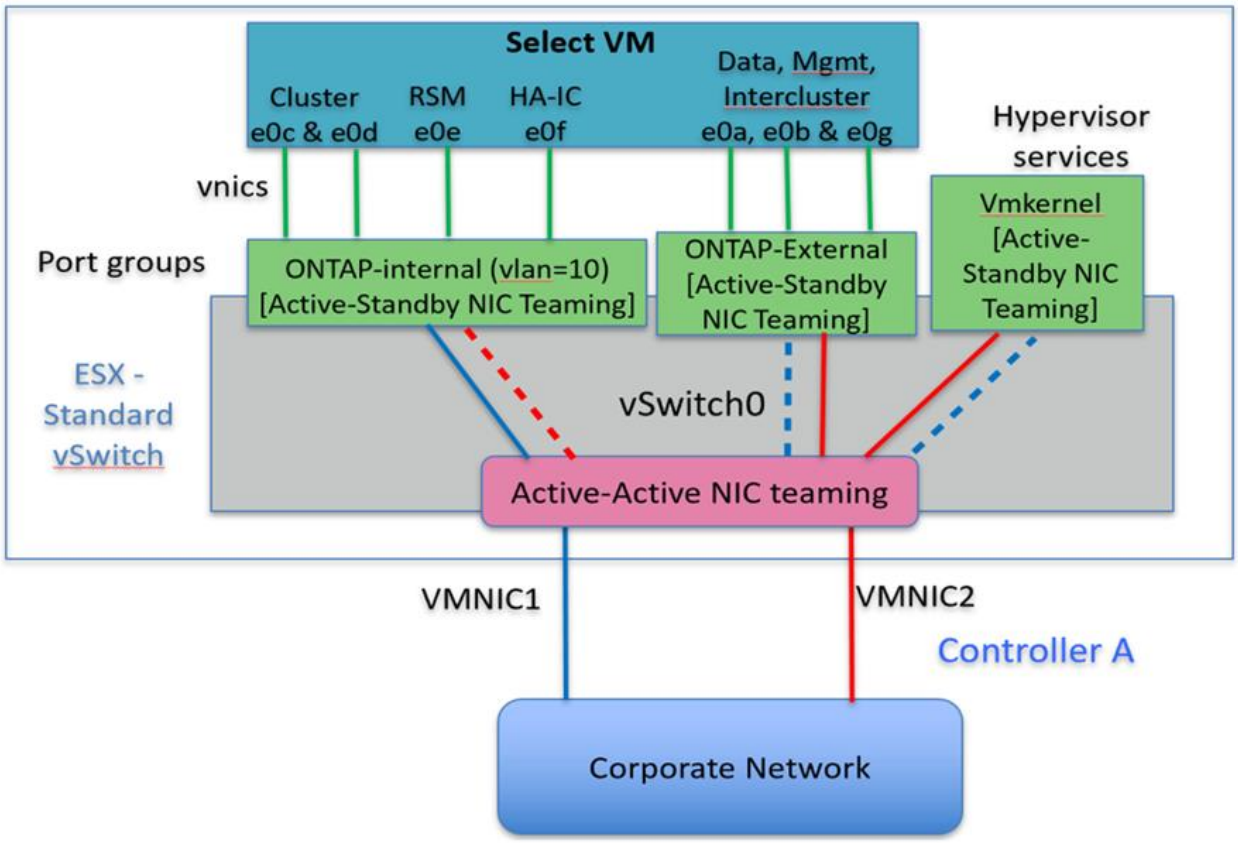


[1] The port groups attached to the virtual NICs are configured to use one NIC as active and the rest as standby.

当使用两个物理端口（10Gb 或更少）时，每个端口组应具有彼此相对配置的活动适配器和备用适配器。内部网络仅适用于多节点 ONTAP Select 群集。对于单节点群集，两个适配器都可以在外部端口组中配置为活动。

以下示例显示了 vSwitch 的配置以及负责处理多节点 ONTAP Select 群集的内部和外部通信服务的两个端口组。外部网络可以在网络中断时使用内部网络 VMNIC，因为内部网络 VMNIC 是此端口组的一部分并以待机模式配置。外部网络的情况正好相反。在两个端口组之间交替使用活动和备用 VMNIC 对于在网络中断期间正确故障转移 ONTAP Select VM 至关重要。

每个节点 **vSwitch** 有两个物理端口（**10Gb** 或更少）

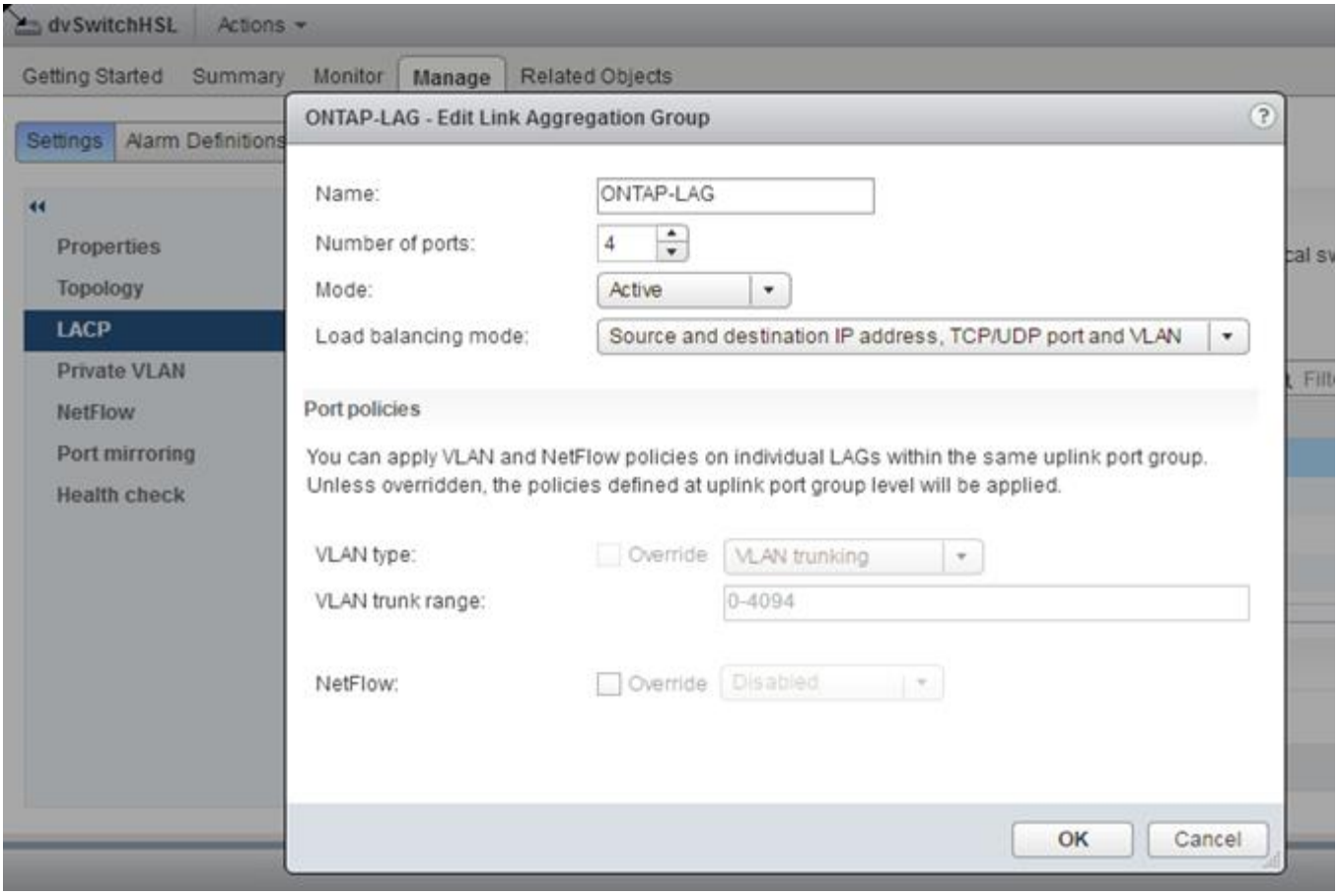


使用 **LACP** 的分布式 **vSwitch**

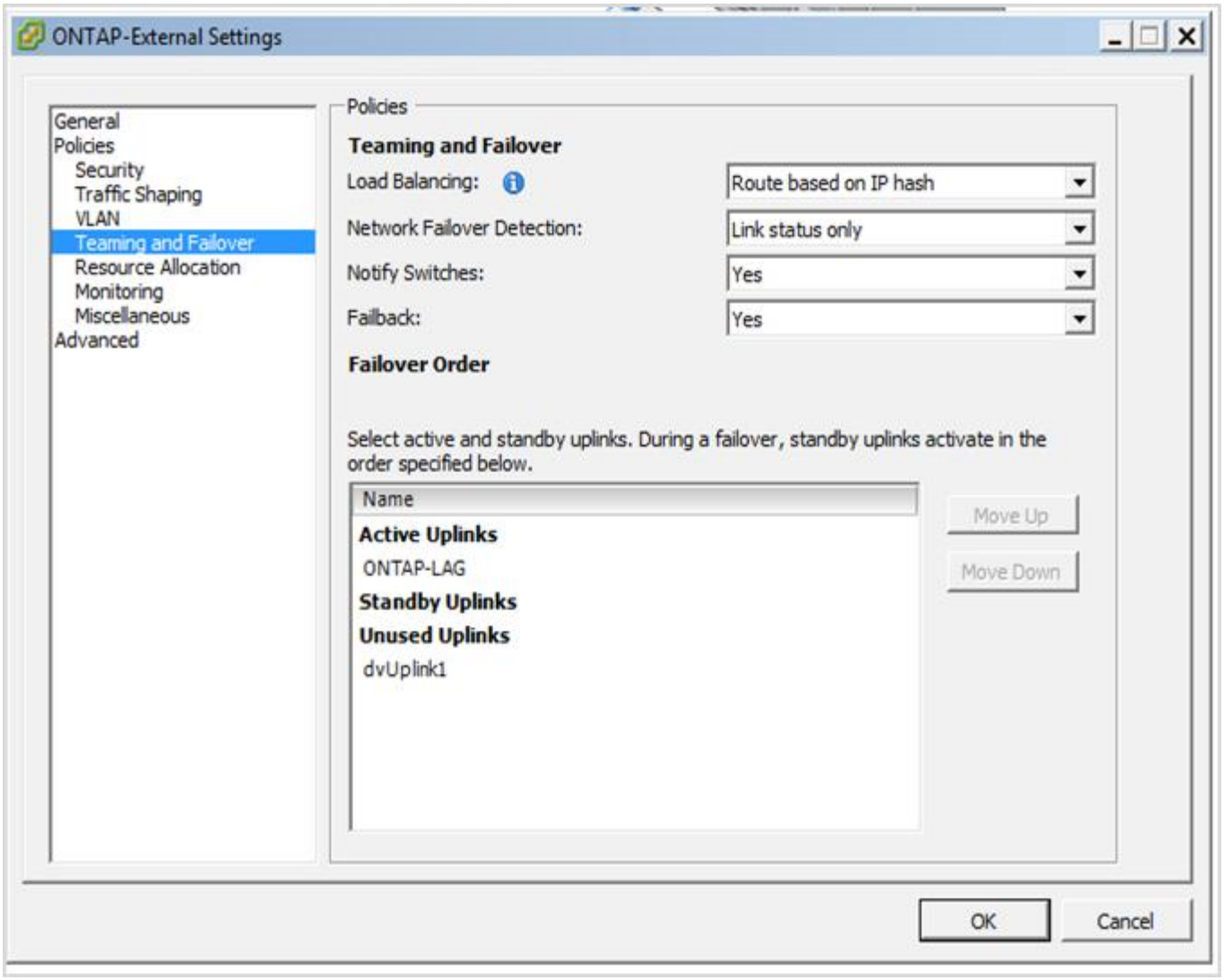
在配置中使用分布式 vSwitches 时，可以使用 LACP（尽管这不是最佳实践）来简化网络配置。唯一受支持的 LACP 配置要求所有 VMNIC 都在一个 LAG 中。上行链路物理交换机必须在通道中的所有端口上支持 7,500 到 9,000 之间的 MTU 大小。内部和外部 ONTAP Select 网络应在端口组级别隔离。内部网络应使用不可路由（隔离）的 VLAN。外部网络可以使用 VST、EST 或 VGT。

以下示例显示了使用 LACP 的分布式 vSwitch 配置。

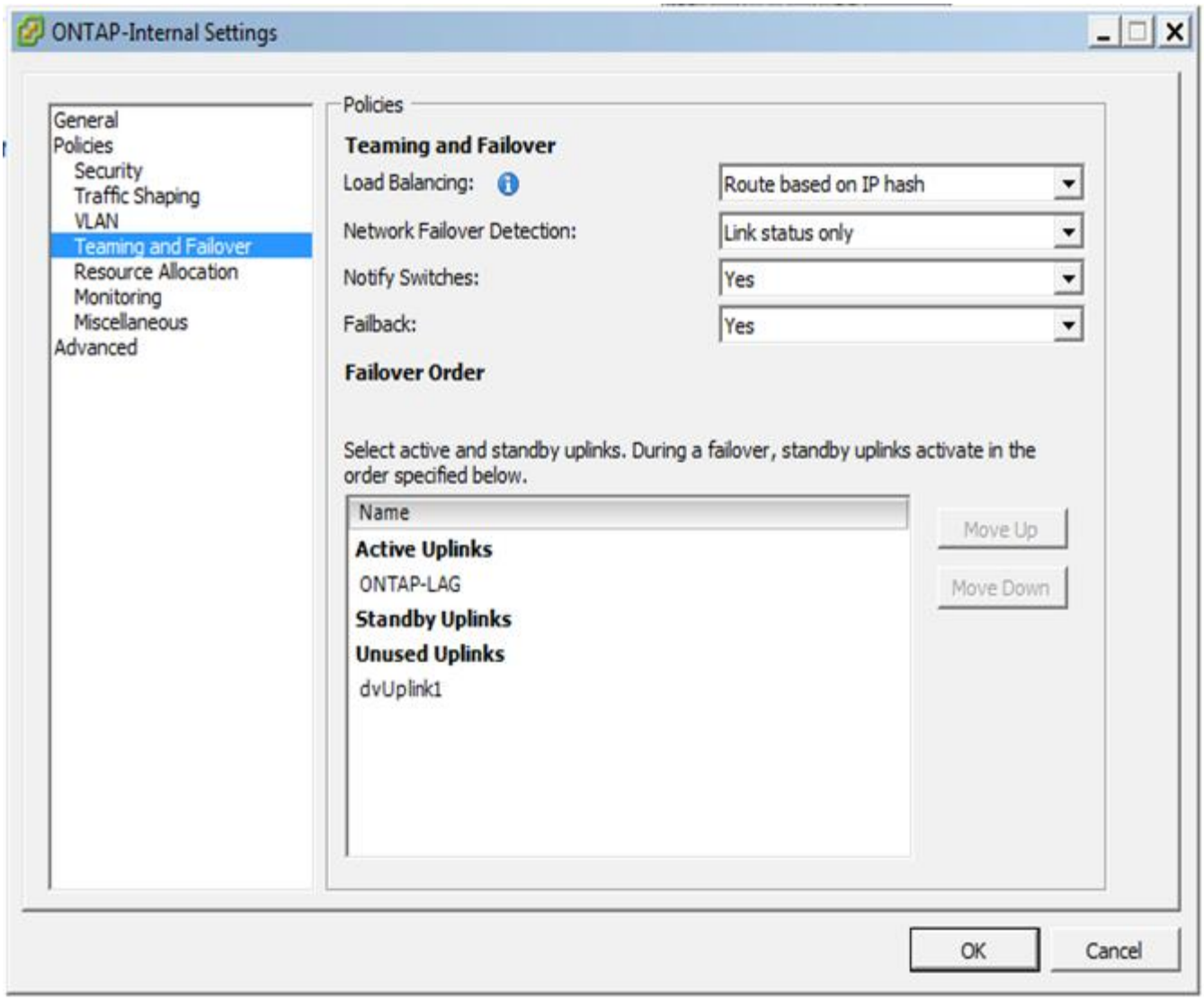
使用 **LACP** 时的 **LAG** 属性



使用启用了 **LACP** 的分布式 **vSwitch** 的外部端口组配置



使用启用了 **LACP** 的分布式 **vSwitch** 的内部端口组配置



LACP 要求您将上游交换机端口配置为端口通道。在分布式 vSwitch 上启用此配置之前，请确保已正确配置启用 LACP 的端口通道。

ONTAP Select 物理交换机配置

基于单交换机和多交换机环境的上游物理交换机配置详细信息。

在做出从虚拟交换机层到物理交换机的连接决策时，应谨慎考虑。内部集群流量与外部数据服务的分离应通过第二层 VLAN 提供的隔离延伸到上游物理组网层。

物理交换机端口应配置为中继端口。ONTAP Select 外部流量可以通过两种方式之一跨越多个二层网络进行分离。一种方法是将 ONTAP VLAN 标记的虚拟端口与单个端口组一起使用。另一种方法是在 VST 模式下将单独的端口组分配给管理端口 e0a。您还必须根据 ONTAP Select 版本和单节点或多节点配置将数据端口分配给 e0b 和 e0c/e0g。如果外部流量跨越多个二层网络分离，则上行物理交换机端口应在其允许的 VLAN 列表中具有这些 VLAN。

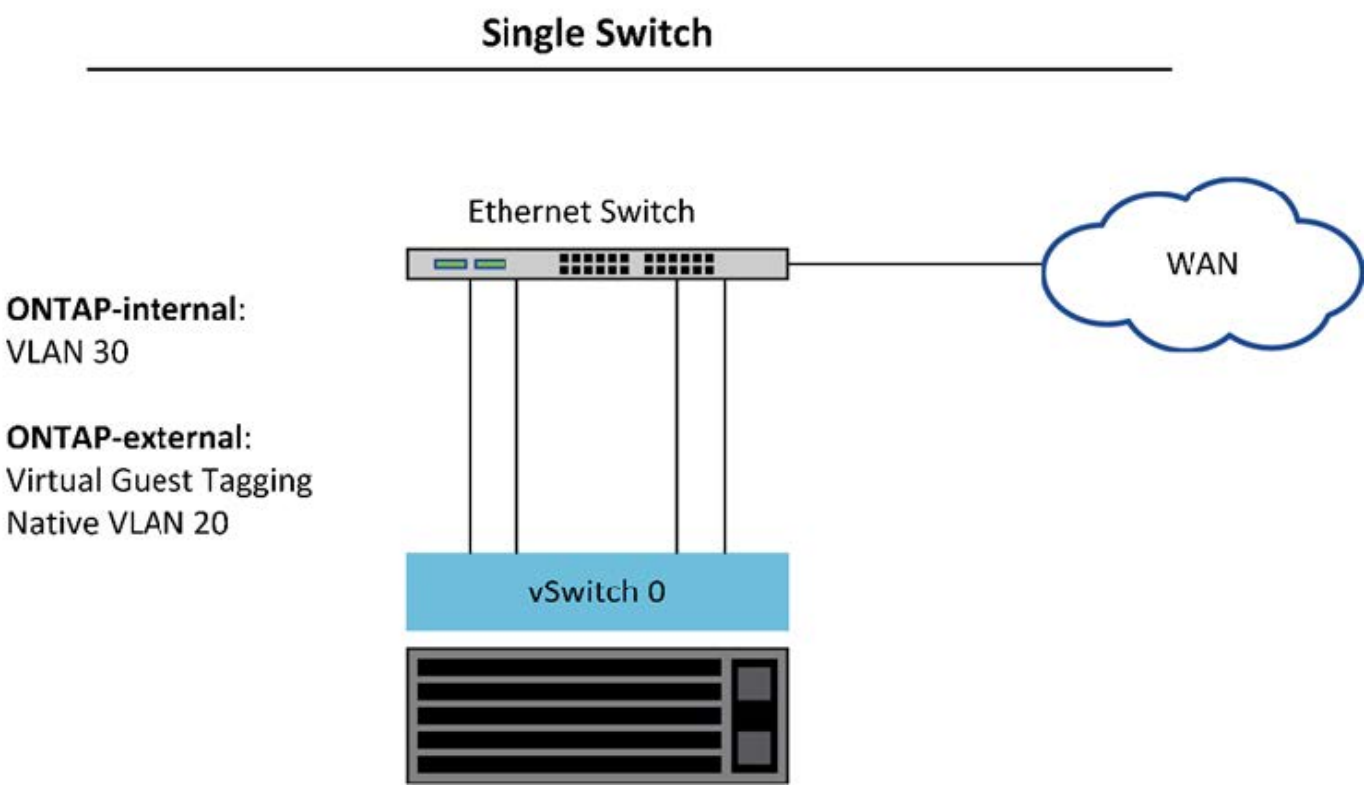
ONTAP Select 内部网络流量使用链路本地 IP 地址定义的虚拟接口进行。由于这些 IP 地址不可路由，因此集群节点之间的内部流量必须流经单个二层网络。不支持 ONTAP Select 集群节点之间的路由跃点。

共享物理交换机

下图显示了多节点 ONTAP Select 群集中一个节点可能使用的交换机配置。在此示例中，vSwitches 托管内部和外部网络端口组使用的物理 NIC 连接到同一个上游交换机。使用包含在单独 VLAN 中的广播域将交换机流量保持隔离。

对于 ONTAP Select 内部网络，标记在端口组级别完成。虽然以下示例将 VGT 用于外部网络，但该端口组同时支持 VGT 和 VST。

使用共享物理交换机的网络配置

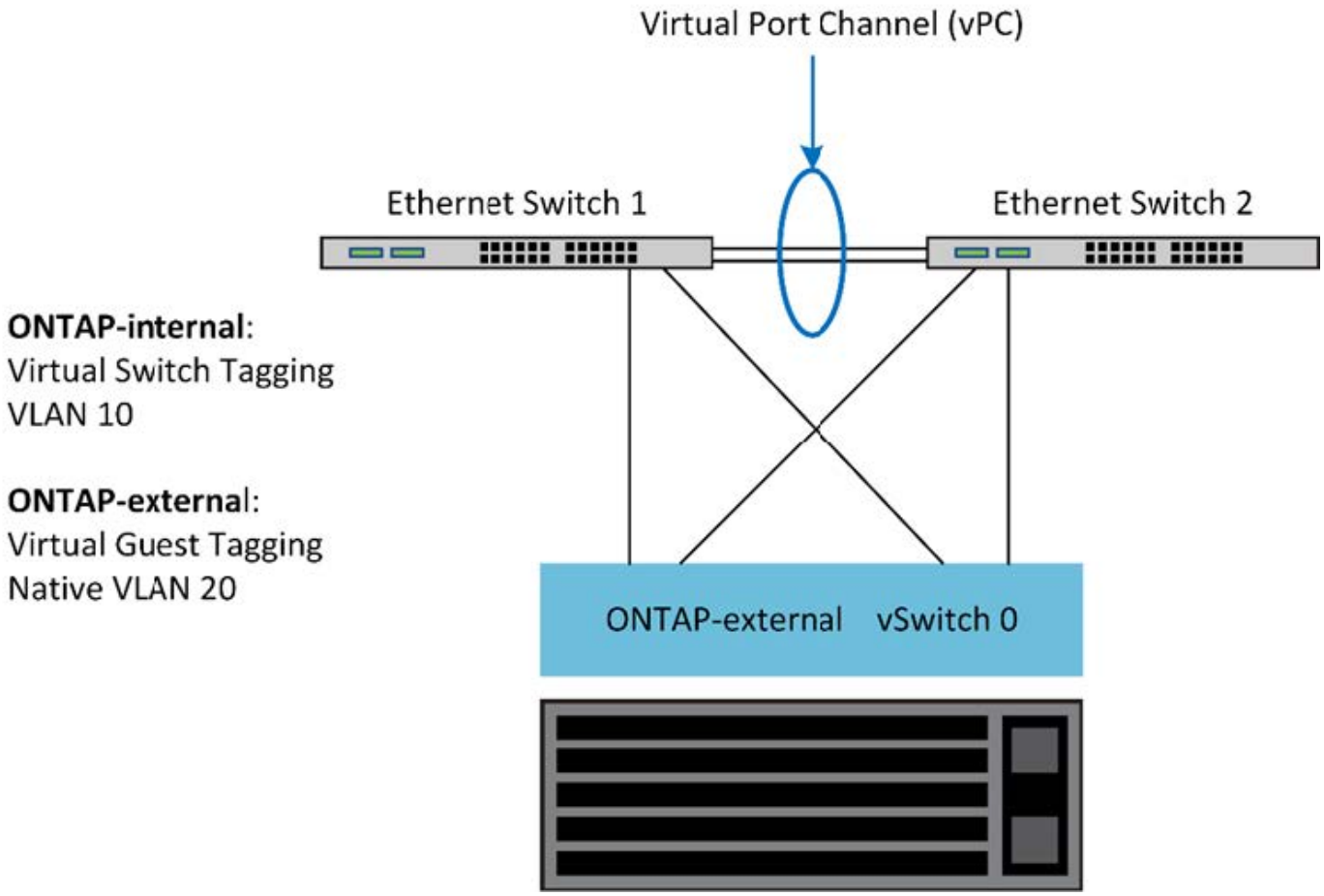


在此配置中，共享交换机将成为单点故障。如果可能，应使用多个交换机来防止物理硬件故障导致集群网络中断。

多个物理交换机

当需要冗余时，应使用多个物理网络交换机。下图显示了多节点 ONTAP Select 群集中一个节点使用的建议配置。来自内部和外部端口组的 NIC 都连接到不同的物理交换机，从而保护用户免受单个硬件交换机故障的影响。交换机之间配置了虚拟端口通道，以防止生成树问题。

使用多个物理交换机的网络配置



ONTAP Select 数据和管理流量分离

将数据流量和管理流量隔离到单独的二层网络中。

ONTAP Select 外部网络流量定义为数据（CIFS、NFS 和 iSCSI）、管理和复制（SnapMirror）流量。在 ONTAP 集群中，每种流量样式都使用必须托管在虚拟网络端口上的单独逻辑接口。在 ONTAP Select 的多节点配置中，这些被指定为端口 e0a 和 e0b/e0g。在单节点配置中，这些端口被指定为 e0a 和 e0b/e0c，而其余端口保留用于内部集群服务。

NetApp 建议将数据流量和管理流量隔离到单独的二层网络中。在 ONTAP Select 环境中，这是使用 VLAN 标签完成的。这可以通过将 VLAN 标记的端口组分配给网络适配器 1（端口 e0a）进行管理流量来实现。然后，您可以将单独的端口组分配给端口 e0b 和 e0c（单节点群集）以及 e0b 和 e0g（多节点群集）用于数据流量。

如果本文档前面描述的 VST 解决方案不够充分，则可能需要在同一虚拟端口上同时配置数据和管理 LIF。为此，请使用称为 VGT 的过程，其中虚拟机执行 VLAN 标记。

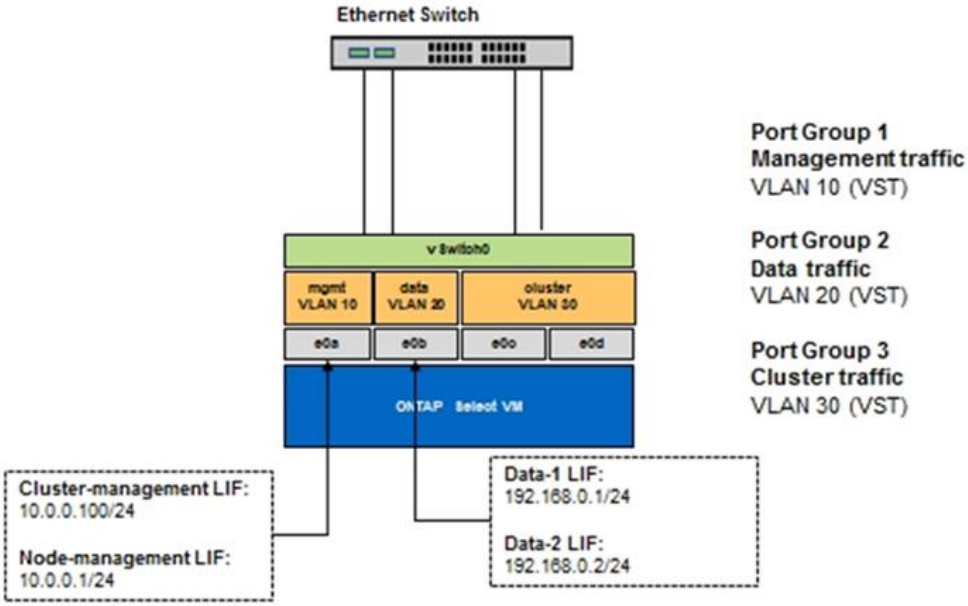
使用 ONTAP Deploy 实用程序时，无法通过 VGT 进行数据和管理网络分离。此过程必须在集群设置完成后执行。

使用 VGT 和双节点集群时还有一个额外的注意事项。在双节点集群配置中，节点管理 IP 地址用于在 ONTAP 完全可用之前建立与 mediator 的连接。因此，映射到节点管理 LIF（端口 e0a）的端口组仅支持 EST 和 VST 标记。此外，如果管理和数据流量都使用相同的端口组，则整个双节点集群仅支持 EST/VST。

支持 VST 和 VGT 这两个配置选项。下图显示了第一个场景 VST，其中流量通过分配的端口组在 vSwitch 层上进行标记。在此配置中，集群和节点管理 LIF 被分配给 ONTAP 端口 e0a，并通过分配的端口组使用 VLAN ID

10 进行标记。数据 LIF 分配给端口 e0b 和 e0c 或 e0g，并使用第二个端口组给定 VLAN ID 20。集群端口使用第三个端口组，并且位于 VLAN ID 30 上。

使用 **VST** 进行数据和管理分离

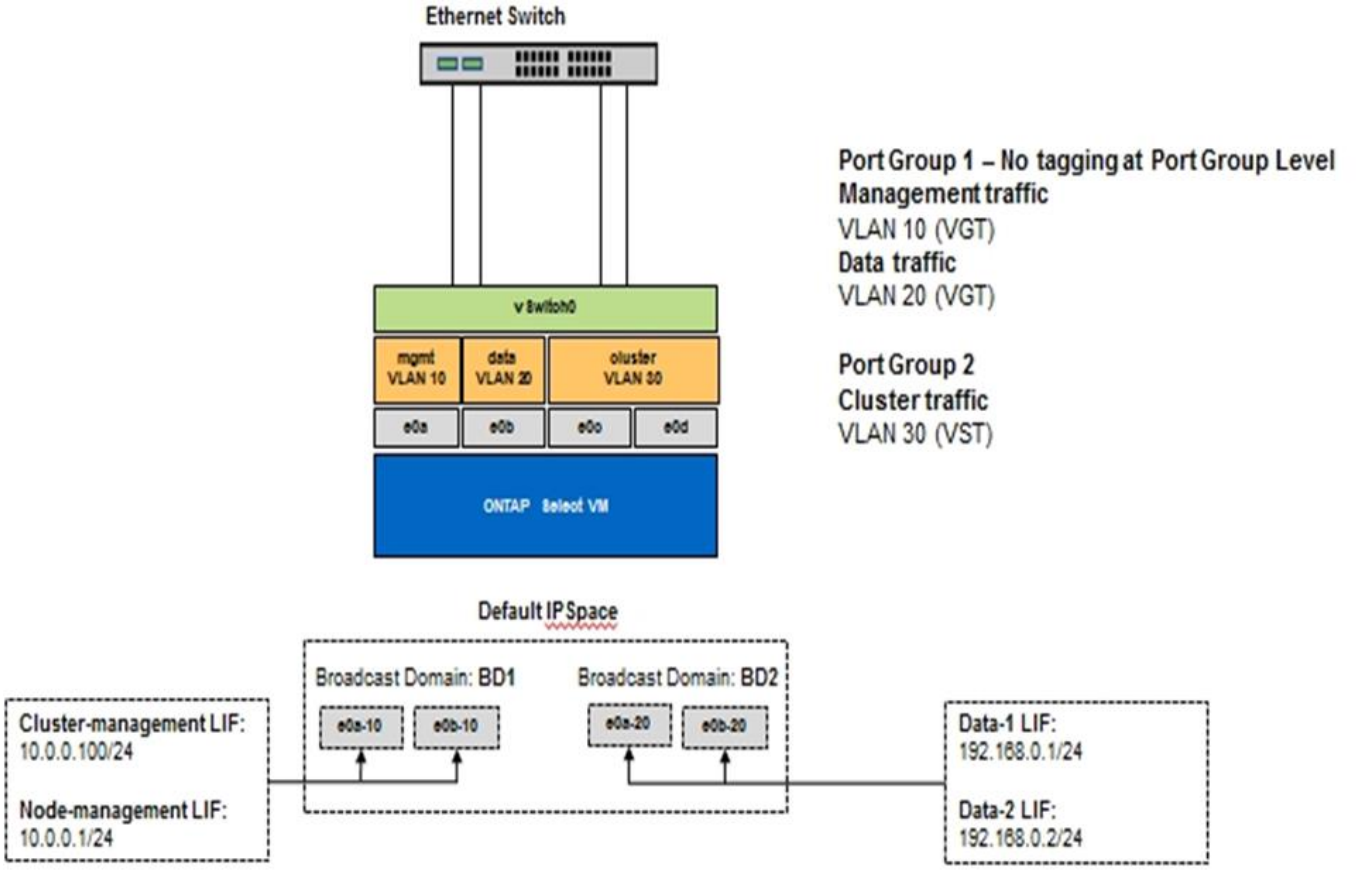


下图显示了第二种情况 VGT，其中 ONTAP VM 使用放置在单独广播域中的 VLAN 端口对流量进行标记。在此示例中，虚拟端口 e0a-10/e0b-10/(e0c 或 e0g)-10 和 e0a-20/e0b-20 放置在 VM 端口 e0a 和 e0b 的顶部。此配置允许网络标记直接在 ONTAP 内执行，而不是在 vSwitch 层执行。管理和数据 LIF 放置在这些虚拟端口上，允许在单个 VM 端口内进一步进行第 2 层细分。集群 VLAN（VLAN ID 30）仍在端口组中标记。

备注：

- 当使用多个 IPspace 时，这种配置风格尤其可取。如果需要进一步的逻辑隔离和多租户，将 VLAN 端口分组到单独的自定义 IPspace 中。
- 要支持 VGT，ESX 主机网络适配器必须连接到物理交换机上的中继端口。连接到虚拟交换机的端口组必须将其 VLAN ID 设置为 4095，才能在端口组上启用中继。

使用 **VGT** 进行数据和管理分离



高可用性架构

ONTAP Select 高可用性配置

发现高可用性选项，为您的环境选择最佳的 HA 配置。

尽管客户开始将应用程序工作负载从企业级存储设备转移到基于软件的商用硬件上运行的解决方案，但围绕弹性和容错的期望和需求并没有改变。提供零恢复点目标 (RPO) 的高可用性解决方案可保护客户免受基础设施堆栈中任何组件故障造成的数据丢失。

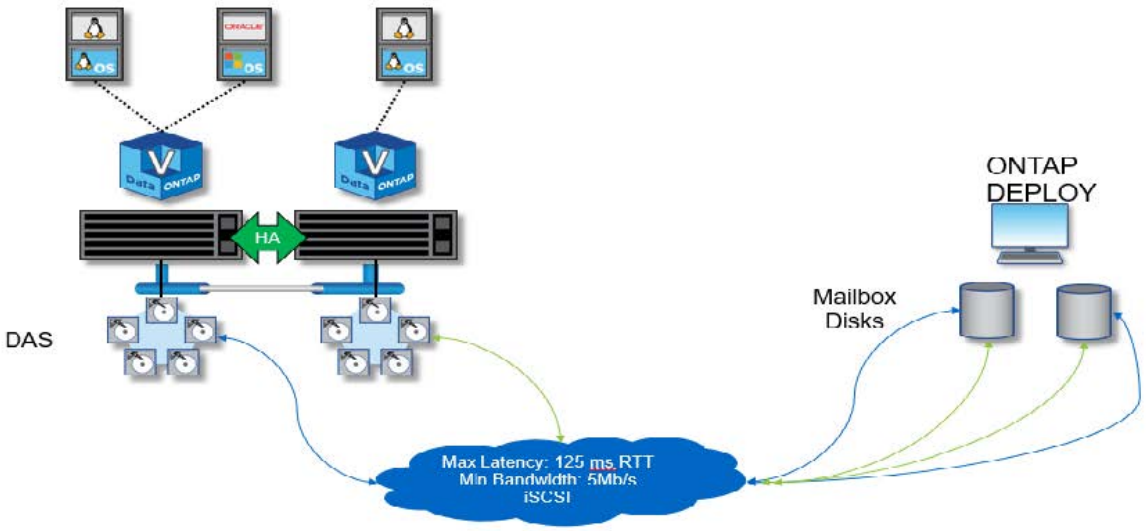
SDS 市场的很大一部分建立在无共享存储的概念之上，软件复制通过在不同的存储孤岛上存储多个用户数据副本来提供数据弹性。ONTAP Select 基于此前提，使用 ONTAP 提供的同步复制功能（RAID SyncMirror）在集群中存储用户数据的额外副本。这发生在 HA 对的上下文中。每个 HA 对存储两个用户数据副本：一个在本地节点提供的存储上，另一个在 HA 合作伙伴提供的存储上。在 ONTAP Select 集群中，HA 和同步复制联系在一起，两者的功能不能分离或独立使用。因此，同步复制功能仅在多节点产品中可用。

> 在 ONTAP Select 集群中，同步复制功能是 HA 实现的功能，而不是异步 SnapMirror 或 SnapVault 复制引擎的替代品。同步复制不能独立于 HA 使用。

有两种 ONTAP Select HA 部署模型：多节点集群（四个、六个、八个、十个或十二个节点）和双节点集群。双节点 ONTAP Select 集群的突出特点是使用外部调解器服务来解决脑裂场景。ONTAP Deploy VM 充当其配置的所有双节点 HA 对的默认调解器。

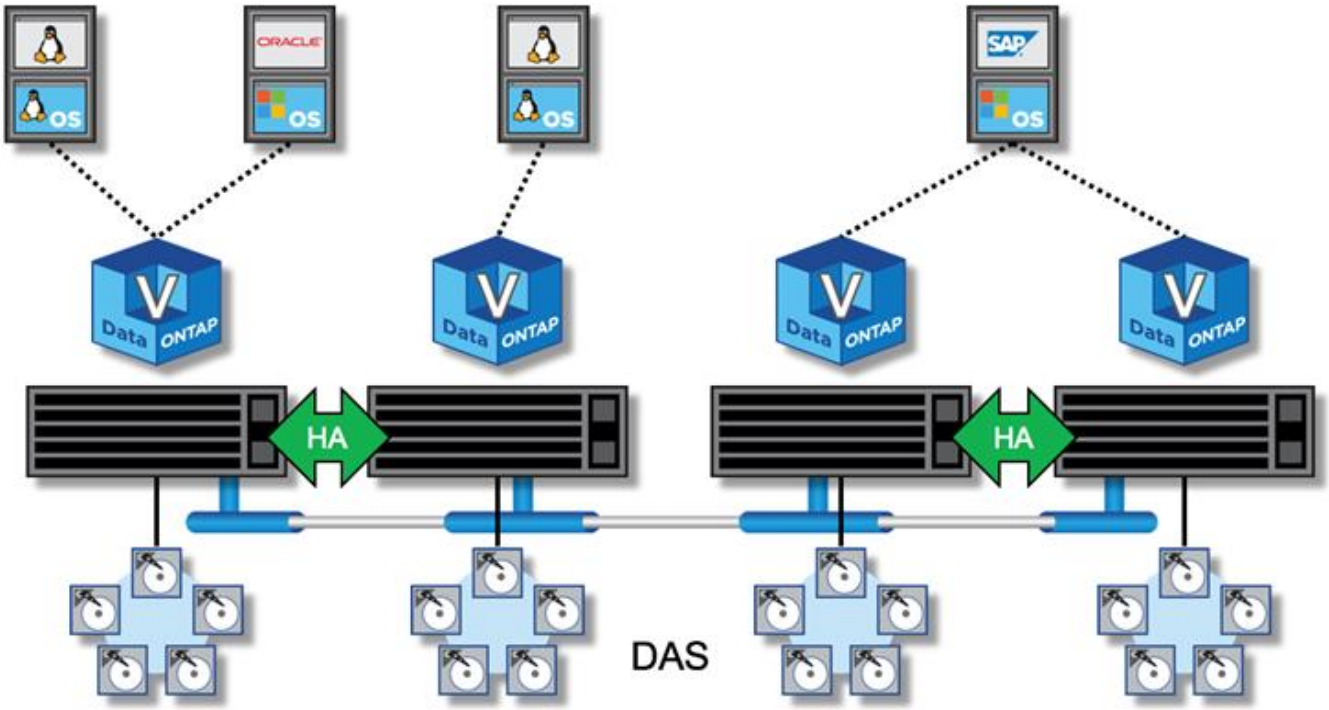
这两个架构如下图所示。

双节点 **ONTAP Select** 集群，带有远程介体并使用本地连接存储



双节点 ONTAP Select 集群由一个 HA 对和一个中介器组成。在 HA 对中，每个集群节点上的数据聚合会同步镜像，并且在发生故障转移时不会丢失数据。

四节点 **ONTAP Select** 集群使用本地连接存储



- 四节点 ONTAP Select 集群由两个 HA 对组成。六节点、八节点、十节点和十二节点集群分别由三个、四个、五个和六个 HA 对组成。在每个 HA 对中，每个集群节点上的数据聚合都会同步镜像，并且在发生故障转移时不会丢失数据。
- 使用 DAS 存储时，物理服务器上只能存在一个 ONTAP Select 实例。ONTAP Select 需要对系统的本地 RAID 控制器进行非共享访问，旨在管理本地连接的磁盘，如果没有对存储的物理连接，这是不可能的。

双节点 **HA** 与多节点 **HA**

与 FAS 阵列不同，HA 对中的 ONTAP Select 节点仅通过 IP 网络进行通信。这意味着 IP 网络是一个单点故障（SPOF），防止网络分区和分裂脑场景成为设计的一个重要方面。多节点集群可以承受单节点故障，因为集群仲裁可以由三个或多个幸存的节点建立。双节点集群依靠 ONTAP Deploy VM 托管的中介服务实现相同的结果。

ONTAP Select 节点和 ONTAP Deploy 中介服务之间的心跳网络流量最小且具有弹性，因此 ONTAP Deploy VM 可以托管在不同于 ONTAP Select 双节点集群的数据中心。

ONTAP Deploy VM 在充当双节点集群的中介时，成为该集群不可或缺的一部分。如果中介服务不可用，则双节点集群继续服务数据，但 ONTAP Select 集群的存储故障转移功能被禁用。因此，ONTAP Deploy 中介服务必须与 HA 对中的每个 ONTAP Select 节点保持恒定的通信。最小带宽为 5Mbps，最大往返时间（RTT）延迟为 125ms，才能使集群仲裁正常运行。

如果充当中介的 ONTAP Deploy VM 暂时或可能永久不可用，则可以使用辅助 ONTAP Deploy VM 来恢复双节点集群仲裁。这会导致新的 ONTAP Deploy VM 无法管理 ONTAP Select 节点，但它已成功参与集群仲裁算法的配置。ONTAP Select 节点与 ONTAP Deploy VM 之间的通信是通过 IPv4 上的 iSCSI 协议完成的。ONTAP Select 节点管理 IP 地址为发起方，ONTAP Deploy VM IP 地址为目标方。因此，在创建双节点集群时，不可能为节点管理 IP 地址支持 IPv6 地址。在创建双节点集群时，ONTAP Deploy 托管邮箱磁盘将自动创建并屏蔽为正确的 ONTAP Select 节点管理 IP 地址。整个配置在设置过程中自动执行，无需执行进一步的管理操作。创建此集群的 ONTAP Deploy 实例是此集群的默认中介。

如果必须更改原始调解员位置，则需要执行管理操作。即使原始 ONTAP Deploy VM 丢失，也可以恢复群集仲裁。但是，NetApp 建议在实例化每个双节点群集后备份 ONTAP Deploy 数据库。

双节点 **HA** 与双节点拉伸 **HA**（**MetroCluster SDS**）

可以将双节点、活动/活动 HA 集群扩展到更远的距离，并可能将每个节点放置在不同的数据中心。双节点集群和双节点延伸集群（也称为 MetroCluster SDS）之间的唯一区别是节点之间的网络连接距离。

双节点集群被定义为两个节点位于 300m 距离内的同一个数据中心的集群。一般来说，两个节点都有到同一网络交换机或一组交换机间链路（ISL）网络交换机的上行链路。

双节点 MetroCluster SDS 被定义为具有物理分隔（不同房间、不同建筑物和不同数据中心）超过 300 米的节点的集群。此外，每个节点的上行链路连接都连接到单独的网络交换机。MetroCluster SDS 不需要专用硬件。但是，环境应遵守延迟要求（RTT 最多 5ms，抖动最多 5ms，总计 10ms）。

MetroCluster SDS 是一项高级功能，需要 Premium 许可证或 Premium XL 许可证。Premium 许可证支持创建中小型虚拟机以及 HDD 和 SSD 介质。Premium XL 许可证还支持创建 NVMe 驱动器。

MetroCluster SDS 支持本地连接存储 (DAS) 和共享存储 (vNAS)。请注意，由于 ONTAP Select VM 和共享存储之间的网络，vNAS 配置通常具有更高的固有延迟。MetroCluster SDS 配置必须在节点之间提供最大 10ms 的延迟，包括共享存储延迟。换句话说，仅测量 Select VM 之间的延迟是不够的，因为共享存储延迟对于这些配置来说是不可忽视的。

ONTAP Select HA RSM 和镜像聚合

使用 RAID SyncMirror (RSM)、镜像聚合和写入路径防止数据丢失。

同步复制

ONTAP HA 模型建立在 HA 合作伙伴的概念之上。ONTAP Select 通过使用 ONTAP 中存在的 RAID SyncMirror (RSM) 功能在集群节点之间复制数据块，将此架构扩展到非共享商品服务器领域，提供分布在 HA 对中的用户数据的两个副本。

具有中介的双节点群集可以跨越两个数据中心。有关详细信息，请参阅部分"双节点拉伸 HA（MetroCluster SDS）最佳实践"。

镜像聚合

ONTAP Select 集群由 2 到 12 个节点组成。每个 HA 对包含两个用户数据副本，通过 IP 网络跨节点同步镜像。此镜像对用户是透明的，并且是数据聚合的属性，该属性在数据聚合创建过程中自动配置。

必须镜像 ONTAP Select 集群中的所有聚合，以便在发生节点故障转移时获得数据可用性，并在发生硬件故障时避免 SPOF。ONTAP Select 集群中的聚合是从 HA 对中每个节点提供的虚拟磁盘构建的，并使用以下磁盘：

- 一组本地磁盘（由当前 ONTAP Select 节点提供）
- 一组镜像磁盘（由当前节点的 HA 合作伙伴提供）

> 用于构建镜像聚合的本地磁盘和镜像磁盘必须大小相同。这些聚合被称为丛 0 和丛 1（分别表示本地和远程镜像对）。在您的安装中，实际的丛编号可能会有所不同。

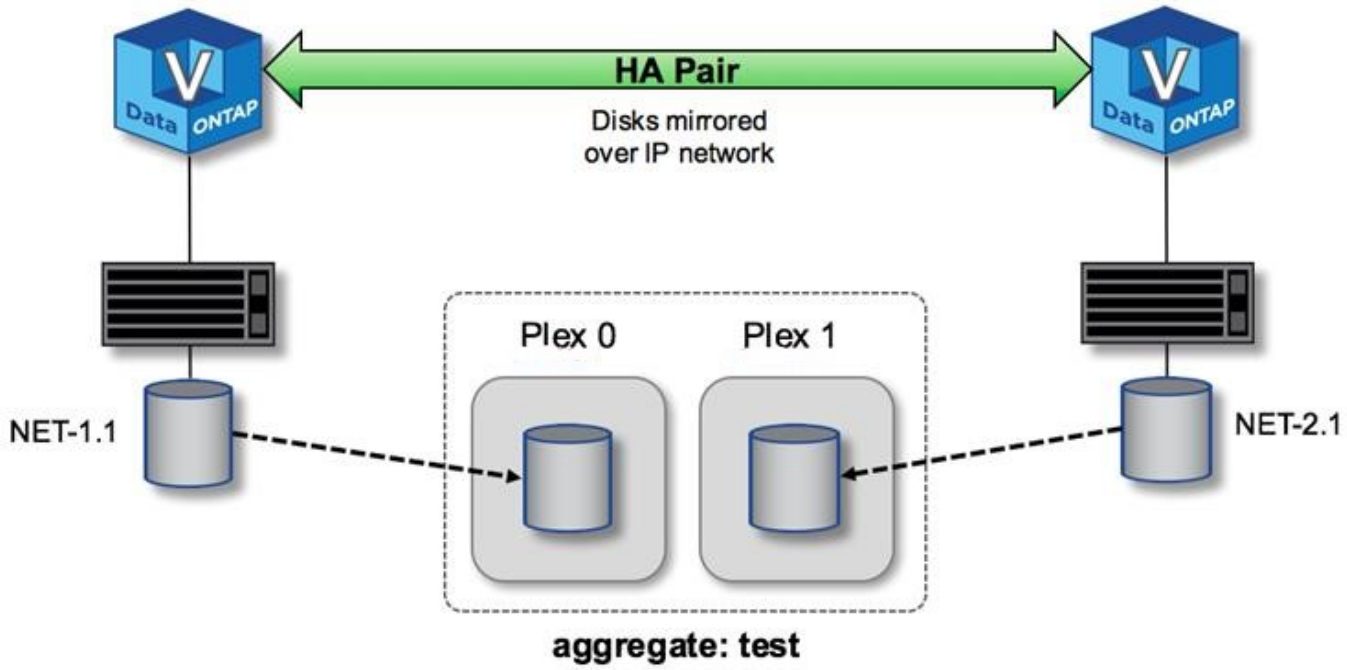
这种方法与标准 ONTAP 集群的工作方式截然不同。这适用于 ONTAP Select 集群中的所有根和数据磁盘。聚合包含数据的本地和镜像副本。因此，包含 N 个虚拟磁盘的聚合提供了 N/2 个磁盘的独特存储，因为数据的第二个副本驻留在其自己的独特磁盘上。

下图显示了四节点 ONTAP Select 集群中的 HA 对。在此集群中，有一个使用来自两个 HA 合作伙伴存储的单个聚合（test）。此数据聚合由两组虚拟磁盘组成：由拥有 ONTAP Select 集群节点贡献的本地集（Plex 0）和由故障转移合作伙伴贡献的远程集（Plex 1）。

Plex 0 是容纳所有本地磁盘的存储桶。Plex 1 是存储镜像磁盘的存储桶，或负责存储用户数据的第二个复制副本的磁盘。拥有聚合的节点向 Plex 0 提供磁盘，该节点的 HA 合作伙伴向 Plex 1 提供磁盘。

在下图中，有一个带有两个磁盘的镜像聚合。此聚合的内容在我们的两个集群节点上进行镜像，本地磁盘 NET-1.1 放置在 Plex 0 存储桶中，远程磁盘 NET-2.1 放置在 Plex 1 存储桶中。在此示例中，聚合 test 由左侧的集群节点所有，并使用本地磁盘 NET-1.1 和 HA 合作伙伴镜像磁盘 NET-2.1。

ONTAP Select 镜像聚合



部署 ONTAP Select 群集时，系统上存在的所有虚拟磁盘将自动分配给正确的丛，无需用户进行有关磁盘分配的额外步骤。这可以防止意外将磁盘分配给不正确的丛，并提供最佳的镜像磁盘配置。

写入路径

集群节点之间数据块的同步镜像以及系统故障时无数据丢失的要求对传入写入在 ONTAP Select 集群中传播时所采取的路径具有重大影响。这个过程包括两个阶段：

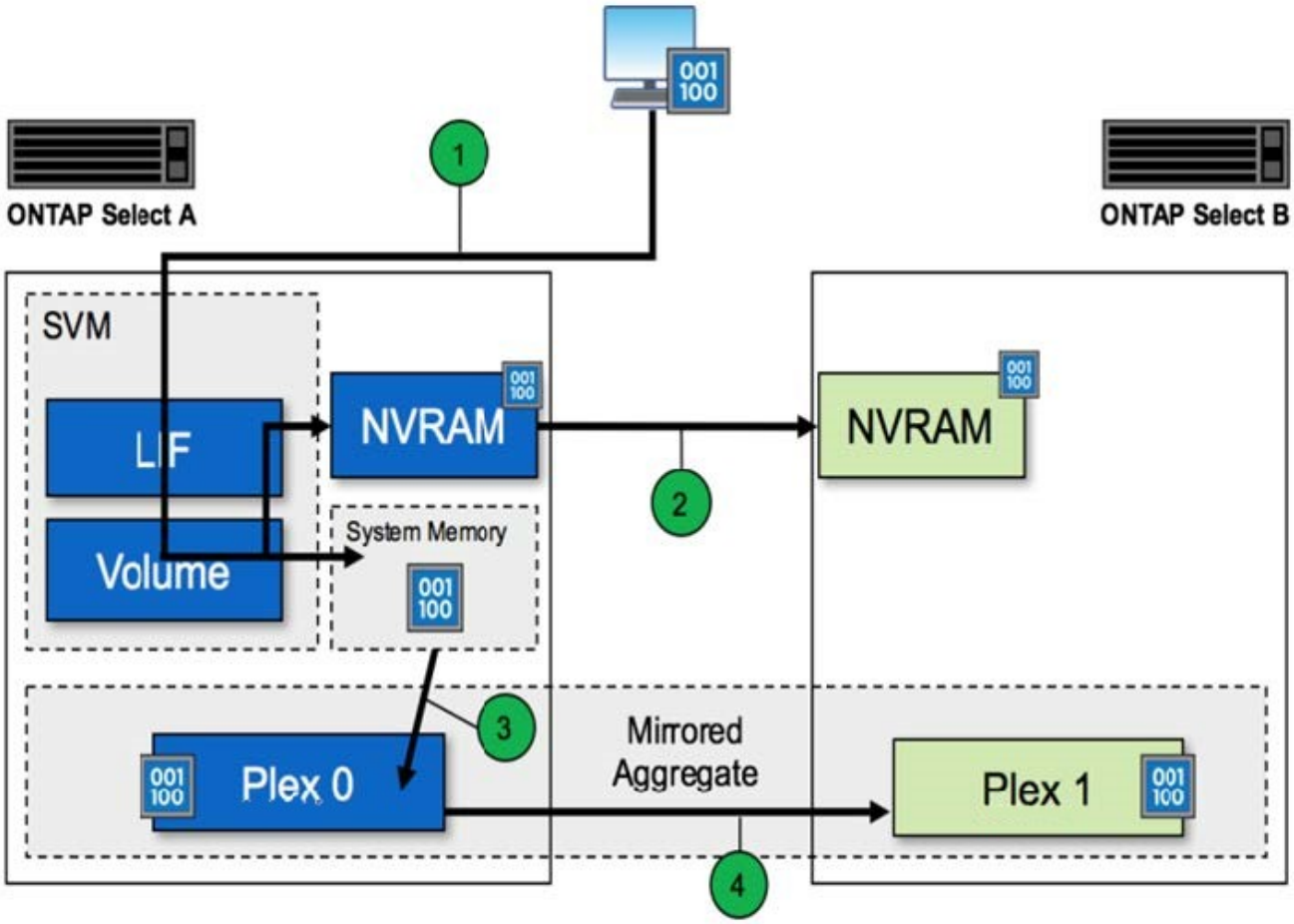
- 确认
- 分段

对目标卷的写入发生在数据 LIF 上，并提交到存在于 ONTAP Select 节点系统磁盘上的虚拟化 NVRAM 分区，然后才被确认返回到客户端。在 HA 配置上，会发生一个额外的步骤，因为这些 NVRAM 写入在被确认之前会立即镜像到目标卷所有者的 HA 合作伙伴。如果原始节点上存在硬件故障，此过程将确保 HA 合作伙伴节点上的文件系统一致性。

写入提交到 NVRAM 后，ONTAP 会定期将此分区的内容移动到适当的虚拟磁盘，这一过程称为转储。此过程仅在拥有目标卷的集群节点上发生一次，在 HA 合作伙伴上不会发生。

下图显示了传入 ONTAP Select 节点的写入请求的写入路径。

ONTAP Select 写入路径工作流程



传入写入确认包括以下步骤：

- 写操作通过 ONTAP Select 节点 A 拥有的逻辑接口进入系统。
- 写入提交到节点 A 的 NVRAM 并镜像到 HA 合作伙伴节点 B。
- 在两个 HA 节点上都存在 I/O 请求后，该请求将被确认回客户端。

ONTAP Select 从 NVRAM 到数据聚合（ONTAP CP）的转储包括以下步骤：

- 写入从虚拟 NVRAM 转储到虚拟数据聚合。
- 镜像引擎将块同步复制到两个丛。

ONTAP Select HA 增强了数据保护

高可用性 (HA) 磁盘心跳、HA 邮箱、HA 心跳、HA 故障转移和 Giveback 工作，以增强数据保护。

磁盘心跳

尽管 ONTAP Select HA 架构利用了传统 FAS 阵列使用的许多代码路径，但仍存在一些例外情况。其中一个例外是基于磁盘的心跳的实现，这是一种非基于网络的通信方法，由集群节点使用，以防止网络隔离导致脑裂行为。脑裂场景是集群分区的结果，通常是由网络故障引起的，其中双方都认为对方已关闭并试图接管集群资源。

企业级 HA 实现必须优雅地处理此类方案。ONTAP 通过自定义的基于磁盘的心跳方法来实现这一点。这是 HA 邮箱的工作，HA 邮箱是物理存储上的一个位置，群集节点使用它来传递心跳消息。这有助于群集确定连接，从而在故障转移时定义仲裁。

在使用共享存储 HA 架构的 FAS 阵列上，ONTAP 通过以下方式解决脑裂问题：

- SCSI 永久预留
- 持久 HA 元数据
- 通过 HA 互连发送的 HA 状态

但是，在 ONTAP Select 集群的无共享架构中，节点只能看到自己的本地存储，而不能看到 HA 合作伙伴的本地存储。因此，当网络分区隔离 HA 对的每一侧时，确定集群仲裁和故障转移行为的前述方法不可用。

虽然不能使用现有的脑裂检测和避免方法，但仍然需要一种仲裁方法，这种方法符合无共享环境的约束。ONTAP Select 扩展了现有的邮箱基础架构，使其能够在发生网络分区时充当仲裁方法。由于共享存储不可用，因此通过 NAS 访问邮箱磁盘来实现仲裁。这些磁盘使用 iSCSI 协议分布在整个集群中，包括双节点集群中的仲裁器。因此，集群节点可以根据对这些磁盘的访问情况做出智能故障转移决策。如果某个节点可以访问其 HA 合作伙伴之外的其他节点的邮箱磁盘，则该节点可能正常运行且运行状况良好。

> 解决集群仲裁和脑裂问题的邮箱架构和基于磁盘的心跳方法是 ONTAP Select 的多节点变体需要四个独立节点或双节点集群调解器的原因。

HA 邮箱发布

HA 邮箱架构使用消息发布模型。集群节点会定期向集群中的所有其他邮箱磁盘（包括调解器）发布消息，说明该节点已启动且正在运行。在健康集群中的任何时间点，集群节点上的单个邮箱磁盘都具有从所有其他集群节点发布的消息。

连接到每个 Select 群集节点的是专门用于共享邮箱访问的虚拟磁盘。此磁盘称为中介邮箱磁盘，因为其主要功能是在发生节点故障或网络分区时充当群集中介的方法。此邮箱磁盘包含每个集群节点的分区，并由其他 Select 集群节点通过 iSCSI 网络装载。这些节点定期将运行状况状态发布到邮箱磁盘的相应分区。通过使用遍布整个集群的网络可访问邮箱磁盘，您可以通过可访问性矩阵推断节点健康状况。例如，集群节点 A 和 B 可以发布到集群节点 D 的邮箱，但不能发布到节点 C 的邮箱。此外，集群节点 D 无法发布到节点 C 的邮箱，因此节点 C 很可能已关闭或网络隔离，应被接管。

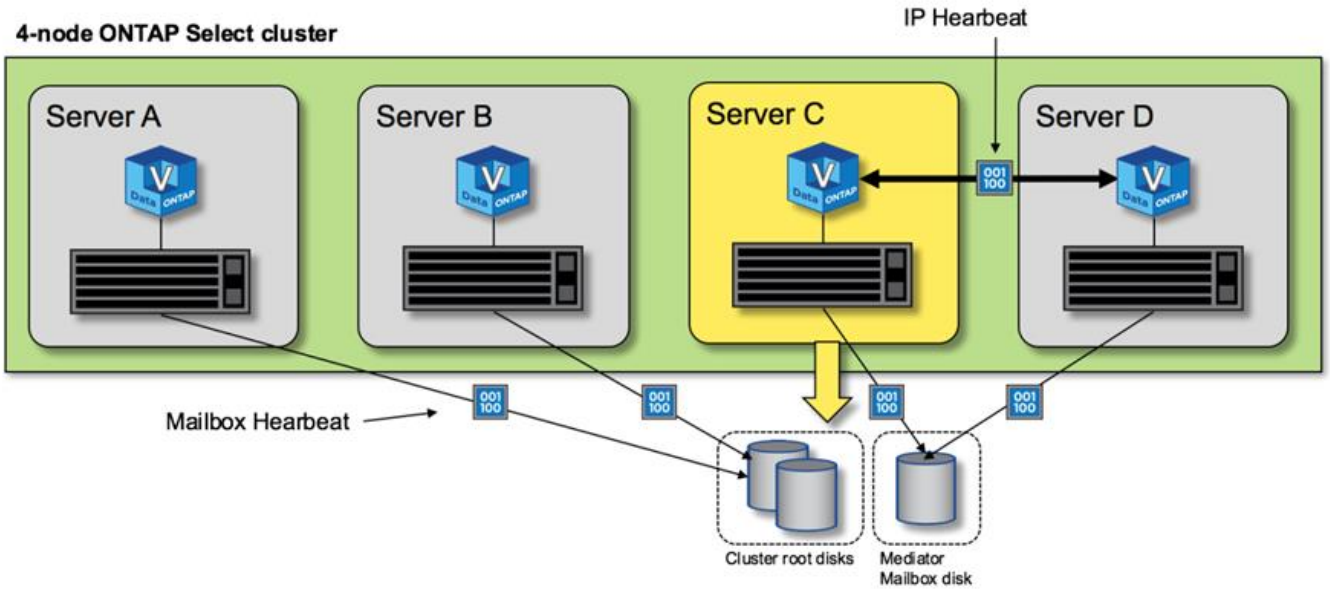
HA 心跳

与 NetApp FAS 平台一样，ONTAP Select 定期通过 HA 互连发送 HA 心跳消息。在 ONTAP Select 集群中，这是通过 HA 合作伙伴之间存在的 TCP/IP 网络连接执行的。此外，基于磁盘的心跳消息将传递到所有 HA 邮箱磁盘，包括中介邮箱磁盘。这些消息每隔几秒钟传递一次，并定期读回。发送和接收这些数据的频率允许 ONTAP Select 集群在大约 15 秒内检测到 HA 故障事件，这与 FAS 平台上可用的窗口相同。当心跳消息不再被读取时，将触发故障转移事件。

下图显示了从单个 ONTAP Select 集群节点（节点 C）的角度通过 HA 互连和调解器磁盘发送和接收心跳消息的过程。

> 网络心跳通过 HA 互连发送到 HA 合作伙伴节点 D，而磁盘心跳在所有群集节点 A、B、C 和 D 上使用邮箱磁盘。

四节点集群中的 **HA** 心跳：稳定状态



HA 故障转移和交还

在故障转移操作期间，幸存的节点使用其 HA 合作伙伴数据的本地副本为其对等节点承担数据服务责任。客户端 I/O 可以不间断地继续，但必须先复制对此数据的更改，然后才能进行回馈。请注意，ONTAP Select 不支持强制回馈，因为这会导致存储在幸存节点上的更改丢失。

当重新启动的节点重新加入集群时，会自动触发同步回操作。同步回所需的时间取决于几个因素。这些因素包括必须复制的更改数量、节点之间的网络延迟以及每个节点上磁盘子系统的速度。同步回所需的时间可能会超过 10 分钟的自动交还窗口。在这种情况下，需要在同步回之后进行手动交还。可以使用以下命令监控同步回的进度：

```
storage aggregate status -r -aggregate <aggregate name>
```

性能

ONTAP Select 性能概述

由于底层硬件和配置的特性，ONTAP Select 集群的性能可能会有很大差异。特定硬件配置是影响特定 ONTAP Select 实例性能的最大因素。以下是影响特定 ONTAP Select 实例性能的一些因素：

- 核心频率。一般来说，更高的频率是更好的。
- 单插槽与多插槽。ONTAP Select 不使用多插槽功能，但支持多插槽配置的虚拟机管理程序开销会导致总性能出现一定程度的偏差。
- **RAID** 卡配置和相关的虚拟机管理程序驱动程序。虚拟机管理程序提供的默认驱动程序可能需要由硬件供应商驱动程序替换。
- **RAID** 组中的驱动器类型和驱动器数量。
- 虚拟机监控程序版本和补丁级别。

ONTAP Select 9.6 性能：Premium HA 直连 SSD 存储

参考平台的性能信息。

参考平台

ONTAP Select (Premium XL) 硬件（每个节点）

- FUJITSU PRIMERGY RX2540 M4：
  - Intel® Xeon® Gold 6142b CPU，2.6 GHz
  - 32 个物理内核（16 x 2 插槽），64 个逻辑
  - 256 GB 内存
  - 每台主机的驱动器：24 个 960GB SSD
  - ESXi 6.5U1

客户端硬件

- 5 x NFSv3 IBM 3550m4 客户端

配置信息

- SW RAID 1 x 9 + 2 RAID-DP（11 个驱动器）
- 22+1 RAID-5（ONTAP 中的 RAID-0）/ RAID 缓存 NVRAM
- 未使用存储效率功能（压缩、重复数据删除、快照副本、SnapMirror 等）

下表列出了使用软件 RAID 和硬件 RAID 针对高可用性 (HA) 对 ONTAP Select 节点的读/写工作负载测量的吞吐量。使用 SIO 负载生成工具进行性能测量。

这些性能数字基于 ONTAP Select 9.6。

直连存储 **(DAS) SSD** 上单个节点（四节点中等实例的一部分）**ONTAP Select** 集群的性能结果，使用软件 **RAID** 和硬件 **RAID**

问题描述	顺序读取 **64KiB**	顺序写入 **64KiB**	随机读取 **8KiB**	随机写入 **8KiB**	随机 **WR/RD (50/50) 8KiB**
ONTAP Select 大型实例，带 DAS (SSD) 软件 RAID	2171 MiBps	559 MiBps	954 MiBps	394 MiBps	564 MiBps
ONTAP Select 中型实例，带 DAS (SSD) 软件 RAID	2090 MiBps	592 MiBps	677 MiBps	335 MiBps	441 3MiBps
ONTAP Select 中型实例，采用 DAS (SSD) 硬件 RAID	2038 MiBps	520 MiBps	578 MiBps	325 MiBps	399 MiBps

64K 顺序读取

详细信息:

- 已启用 SIO 直接 I/O
- 2 个节点
- 每个节点 2 个数据 NIC
- 每个节点 1 个数据聚合（2TB 硬件 RAID），（8TB 软件 RAID）
- 64 个 SIO 处理器，每个处理器 1 个线程
- 每个节点 32 个卷
- 每个进程 1 个文件；每个文件为 12000MB

64K 顺序写入

详细信息:

- 已启用 SIO 直接 I/O
- 2 个节点
- 每个节点 2 个数据网络接口卡 (NIC)
- 每个节点 1 个数据聚合（2TB 硬件 RAID），（4TB 软件 RAID）
- 128 个 SIO 进程，每个进程 1 个线程
- 每个节点的卷数：32（硬件 RAID），16（软件 RAID）
- 每个进程 1 个文件；每个文件为 30720MB

8K 随机读取

详细信息:

- 已启用 SIO 直接 I/O
- 2 个节点
- 每个节点 2 个数据网卡
- 每个节点 1 个数据聚合（2TB 硬件 RAID），（4TB 软件 RAID）
- 64 个 SIO 进程，每个进程 8 个线程
- 每个节点的卷数：32
- 每个进程 1 个文件；每个文件为 12228MB

8K 随机写入

详细信息:

- 已启用 SIO 直接 I/O
- 2 个节点

- 每个节点 2 个数据网卡
- 每个节点 1 个数据聚合（2TB 硬件 RAID），（4TB 软件 RAID）
- 64 个 SIO 进程，每个进程 8 个线程
- 每个节点的卷数：32
- 每个进程 1 个文件；每个文件为 8192MB

8K 随机 **50%** 写入 **50%** 读取

详细信息:

- 已启用 SIO 直接 I/O
- 2 个节点
- 每个节点 2 个数据网卡
- 每个节点 1 个数据聚合（2TB 硬件 RAID），（4TB 软件 RAID）
- 每个进程 64 个 SIO proc208 线程
- 每个节点的卷数：32
- 每个进程 1 个文件；每个文件为 12228MB