Red Hat Virtualization 4.0

技术参考

Red Hat Virtualization 环境的技术架构

Last Updated: 2022-12-25

Red Hat Virtualization 4.0 技术参考

Red Hat Virtualization 环境的技术架构

Enter your first name here. Enter your surname here.
Enter your organisation's name here. Enter your organisational division here.
Enter your email address here.

法律通告

摘要

本参考记录了 Red Hat Virtualization 环境中使用的概念、组件和技术。

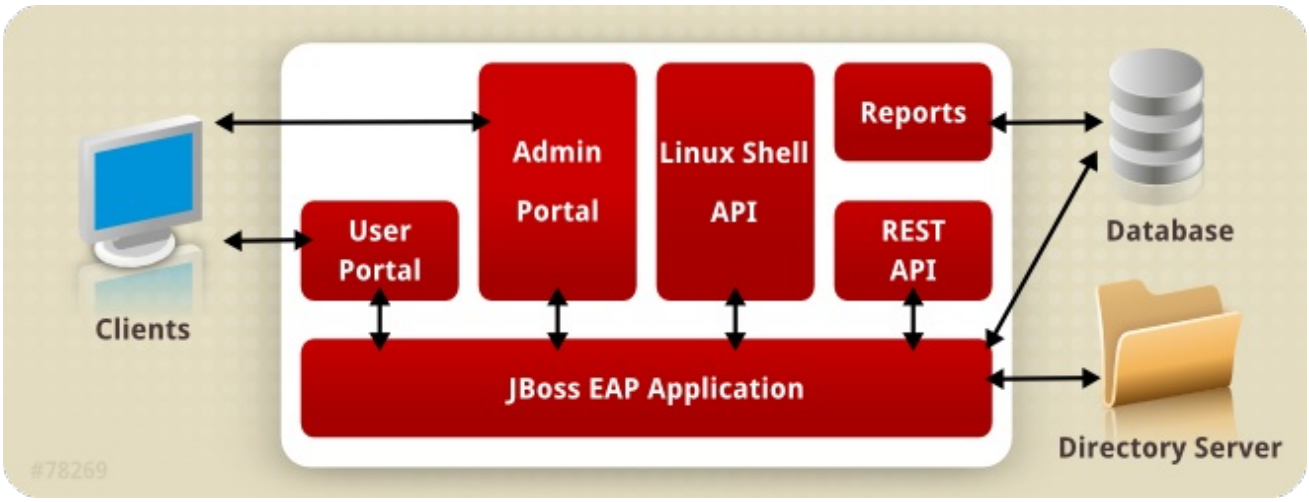
目录

第 1 章 简介

1.1. RED HAT VIRTUALIZATION MANAGER

Red Hat Virtualization Manager 为虚拟化环境提供集中管理。可以使用多个不同接口来访问 Red Hat Virtualization Manager。每个接口都有助于以不同的方式访问虚拟化环境。

图 1.1. Red Hat Virtualization Manager Architecture



Red Hat Virtualization Manager 提供图形界面和应用程序编程接口(API)。*每个接口都连接到管理器，这是一个由 Red Hat JBoss Enterprise Application Platform 的嵌入式实例提供的应用程序。除了 Red Hat JBoss Enterprise Application Platform 外，还有一些其他组件支持 Red Hat Virtualization Manager。*

1.2. RED HAT VIRTUALIZATION HOST

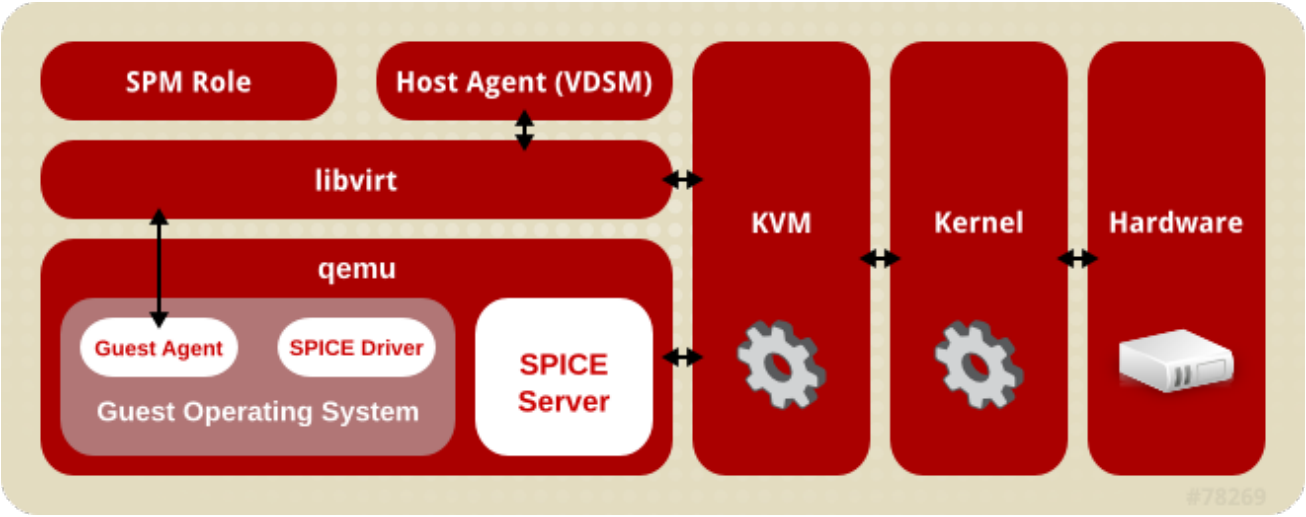
Red Hat Virtualization 环境附加有一个或多个主机。主机是提供虚拟机利用的物理硬件的服务器。

Red Hat Virtualization Host (RHVH) 使用特殊的自定义安装介质（专门用于创建虚拟化主机）运行优化的操作系统。

Red Hat Enterprise Linux 主机是运行标准 Red Hat Enterprise Linux 操作系统的服务器，并在安装后进行配置以允许用作主机。

这两种主机安装都会导致主机以相同的方式与虚拟化环境交互，等等，两者均称为 主机。

图 1.2. 主机架构



基于内核的虚拟机(KVM)

基于内核的虚拟机(KVM)是一种可加载内核模块，通过使用 Intel VT 或 AMD-V 硬件扩展来提供完整的虚拟化。虽然 KVM 本身在内核空间中运行，但其上运行的客户机作为用户空间中的单个 QEMU 进程运行。KVM 允许主机将其物理硬件提供给虚拟机。

QEMU

QEMU 是用于提供完整系统模拟的多平台模拟器。QEMU 可模拟完整的系统，如 PC，包括一个或多个处理器，以及外围设备。QEMU 可用于启动不同的操作系统或调试系统代码。QEMU 与 KVM 以及具有适当虚拟化扩展的处理器配合使用，提供完整的硬件辅助虚拟化。

Red Hat Virtualization Manager Host Agent, VDSM

*在 Red Hat Virtualization 中，**VDSM** 对虚拟机和存储发起操作。它还有助于主机间通信。VDSM 监控主机资源，如内存、存储和网络。此外，VDSM 负责管理诸如虚拟机创建、统计累计和日志收集等任务。VDSM 实例在每个主机上运行，并使用可重新配置的端口 **54321** 从 Red Hat Virtualization Manager 接收管理命令。*

VDSM-REG

***VDSM** 使用 **VDSM-REG** 将每个主机注册到 Red Hat Virtualization Manager。 **vdsm-REG** 使用端口 **80** 或端口 **443** 提供有关其自身及其主机的信息。*

libvirt

libvirt 有助于管理虚拟机及其关联的虚拟设备。当 Red Hat Virtualization Manager 启动虚拟机生命周期命令（启动、停止、重新启动）时，VDSM 会在相关主机上调用 libvirt 来执行它们。

存储池管理程序, SPM

存储池管理程序(SPM)是分配到数据中心内一个主机的角色。SPM 主机具有唯一权威，可以对数据中心进行所有存储域结构元数据更改。这包括创建、删除和操作虚拟磁盘镜像、快照和模板。它还包括为 存储区域网络(SAN)上的稀疏块设备分配存储。SPM 的角色可以迁移到数据中心中的任何主机。因此，数据中心中的所有主机都必须能够访问数据中心中定义的所有存储域。

Red Hat Virtualization Manager 确保 SPM 始终可用。如果出现存储连接性错误，管理器会将 SPM 角色重新分配到其他主机。

客户机操作系统

可以在不修改 Red Hat Virtualization 环境中的虚拟机的情况下安装客户机操作系统。客户机操作系统以及客户机中的任何应用程序都不知道虚拟化环境并正常运行。

红帽提供了增强的设备驱动程序，允许更快更高效地访问虚拟设备。您还可以在客户机上安装 Red Hat Virtualization 客户机代理，该代理为管理控制台提供了增强的客户机信息。

1.3. 用于访问 MANAGER 的接口

开发者门户

桌面虚拟化为用户提供与个人计算机桌面环境类似的桌面环境。客户门户是向用户提供 虚拟桌面基础架构。用户通过 Web 浏览器访问开发人员门户，以显示和访问其分配的虚拟桌面。用户访问中用户可用的操作由系统管理员设置。标准用户可以启动、停止和使用系统管理员为其分配的桌面。超级用户可以执行某些管理操作。两种类型的用户从同一 URL 访问虚拟机门户，并会看到适合他们在登录时权限级别的选项。

- **标准用户访问**

  标准用户可以打开和关闭其虚拟桌面，并通过客户门户网站连接到这些桌面。通过 独立计算环境(SPICE)或 虚拟网络计算(VNC) 客户端的简单协议来 直接连接到虚拟机。这两个协议都为用户提供类似于本地安装的桌面环境的环境。管理员指定创建虚拟机时用于连接到虚拟机的协议。

  有关用户界面中可用的操作的更多信息，以及受支持的浏览器和客户端的更多信息，请参阅《用户门户简介》。

- **超级用户访问权限**

  Red Hat Virtualization 开发人员门户为用户提供了一个图形用户界面来创建、使用和监控虚拟资源。系统管理员可以通过授予用户高级用户访问权限来委派某些管理任务。除了可由标准用户执行的任务外，高级用户可以：

  - 创建、编辑和删除虚拟机.
  - 管理虚拟磁盘和网络接口。
  - 为虚拟机分配用户权限。
  - 创建和使用模板来快速部署虚拟机。
  - 监控资源使用和高严重性事件。
  - 创建和使用快照将虚拟机恢复到以前的状态。

  高级用户可以执行虚拟机管理任务来委派。为环境管理员保存数据中心和集群级别管理任务。

管理门户

管理门户是 Red Hat Virtualization Manager 服务器的图形化管理界面。通过利用它，管理员可以从 Web 浏览器监控、创建和维护虚拟化环境的所有元素。可以从管理门户执行的任务包括：

- 创建和管理虚拟基础架构（子网、存储域）。
- 安装和管理主机。
- 创建和管理逻辑实体（数据中心、集群）。
- 创建和管理虚拟机。

- Red Hat Virtualization 用户和权限管理。

管理门户使用 JavaScript 显示。

《红帽虚拟化管理指南》中进一步详细讨论管理门户功能。https://access.redhat.com/documentation/en/red-hat-virtualization/4.0/single/administration-guide/ 有关管理门户支持的浏览器和平台的信息，请参阅 Red Hat Virtualization 安装指南。

表述性状态转移(REST) API

Red Hat Virtualization REST API 提供了一个软件接口，用于对 Red Hat Virtualization 环境进行干预和控制。REST API 可供支持 HTTP 操作的任何编程语言使用。

使用 REST API 开发人员和管理员可以：

- 与企业 IT 系统集成.
- 与第三方虚拟化软件集成.
- 执行自动化维护和错误检查任务。
- 使用脚本在 Red Hat Virtualization 环境中自动执行重复性任务。

如需 API 规格和用法示例，请参阅 Red Hat Virtualization REST API 指南 。

1.4. 支持 MANAGER 的组件

Red Hat JBoss Enterprise Application Platform

红帽 JBoss 企业应用平台是一个 Java 应用程序服务器。它提供框架，以支持有效开发和交付跨平台 Java 应用程序。Red Hat Virtualization Manager 使用红帽 JBoss Enterprise Application Platform 提供。

重要

与 Red Hat Virtualization Manager 捆绑的 Red Hat JBoss Enterprise Application Platform **的版本不** 用于为其他应用程序提供服务。它为 Red Hat Virtualization Manager 提供特定目的进行了定制。使用 Manager 中包含的红帽 JBoss Enterprise Application Platform 来获取其他目的，不会影响其为 Red Hat Virtualization 环境提供服务的能力。

收集报告和历史数据

Red Hat Virtualization Manager 包括一个数据仓库，用于收集有关主机、虚拟机和存储的监控数据。有多个预定义的报告可用。客户可以使用支持 SQL 的任何查询工具分析其环境并创建报告。

Red Hat Virtualization Manager 安装过程创建两个数据库。这些数据库在 Postgres 实例上创建，后者是在安装过程中选择的。

- engine 数据库是 Red Hat Virtualization Manager 使用的主要数据存储。有关虚拟化环境的信息，如状态、配置和性能会存储在此数据库中。
- ovirt_engine_history 数据库包含来自 engine 操作数据库的配置信息和统计指标。engine 数据库中的配置数据会每分钟检查，更改会复制到 ovirt_engine_history 数据库。跟踪对数据库的更改提供了有关数据库中对象的信息。这可让您分析和提高 Red Hat Virtualization 环境的性能并解决困难。

有关根据 ovirt_engine_history 数据库生成报告的更多信息，请参阅 Red Hat Virtualization Data Warehouse 指南中的 History 数据库。

重要

ovirt_engine_history 数据库中数据的复制由 **RHEVM History Service**、ovirt-engine-dwhd 执行。

目录服务

目录服务提供基于网络的集中存储和组织信息存储。存储的信息类型包括应用程序设置、用户配置文件、组数据、策略和访问控制。Red Hat Virtualization Manager 支持 Active Directory、身份管理 (IdM)、OpenLDAP 和 Red Hat Directory Server 9。另外，还有一个本地内部域进行管理目的。此内部域只有一个用户：admin 用户。

1.5. 存储

Red Hat Virtualization 使用集中存储系统用于虚拟磁盘镜像、模板、快照和 ISO 文件。存储以逻辑方式分组为存储池，它由存储域组成。存储域是存储容量和元数据的组合，用于描述存储的内部结构。存储域有三种类型：数据、导出和 ISO。

数据存储域是每个数据中心唯一需要的。数据存储域专用于单个数据中心。导出和导入 ISO 域是可选的。存储域是共享资源，必须可以被数据中心中的所有主机访问。

存储网络可以使用网络文件系统(NFS)、互联网小型计算机系统接口(iSCSI)、GlusterFS、光纤通道协议 (FCP)或任何 POSIX 兼容网络文件系统来实施。

在 NFS （及其他 POSIX 兼容文件系统）域上，所有虚拟磁盘、模板和快照都是简单文件。

在 SAN (iSCSI/FCP)域中，块设备由逻辑卷管理器(LVM)聚合至卷组(VG)。每个虚拟磁盘、模板和快照都是 VG 上的逻辑卷(LV)。有关 LVM 的详情，请查看 Red Hat Enterprise Linux 逻辑卷管理器管理指南 。

数据存储域

数据域保存环境中运行的所有虚拟机的虚拟硬盘镜像。虚拟机的模板和快照也存储在数据域中。数据域无法在数据中心间共享。

导出存储域

导出域是用于在数据中心和 Red Hat Virtualization 环境之间复制和移动镜像的临时存储存储库。导出域可用于备份虚拟机和模板。导出域可以在数据中心之间移动，但一次只能在一个数据中心内处于活动状态。

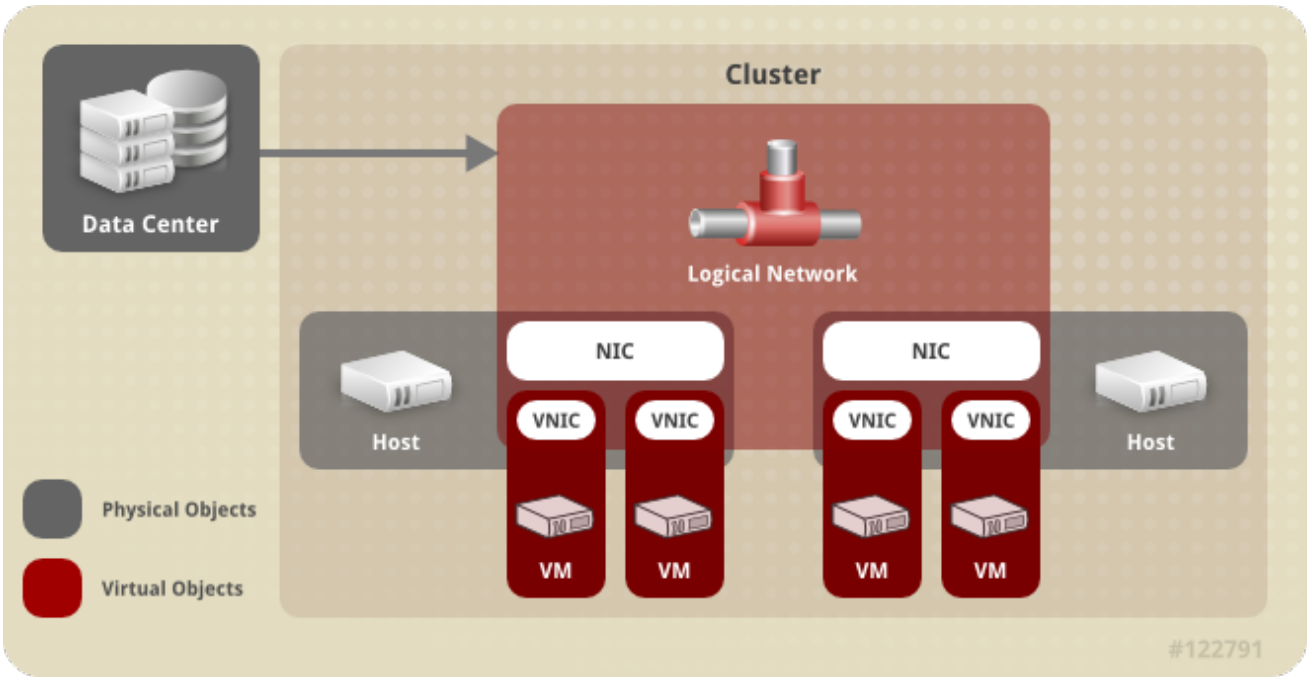
ISO 存储域

ISO 域存储 ISO 文件，这些文件是用于为虚拟机安装操作系统和应用程序的逻辑 CD-ROM。作为替换物理 CD-ROM 或 DVD 库的逻辑实体，ISO 域会删除数据中心对物理介质的需求。ISO 域可以在不同的数据中心之间共享。

1.6. NETWORK

Red Hat Virtualization 网络架构有助于在 Red Hat Virtualization 环境的不同元素之间进行连接。网络架构不仅支持网络连接，还允许网络隔离。

图 1.3. 网络架构



在 Red Hat Virtualization 在多个层中定义网络。底层物理网络基础架构必须就位，并且配置为允许 Red Hat Virtualization 环境的逻辑组件间的连接。

网络基础架构层

Red Hat Virtualization 网络架构依赖于一些通用的硬件和软件设备：

- *网络接口控制器(NIC)是将主机连接到网络的物理网络接口设备。*
- *虚拟 NIC (VNIC)是使用主机物理 NIC 操作的逻辑 NIC。它们为虚拟机提供网络连接。*
- *绑定将多个 NIC 绑定到一个接口。*
- *网桥是数据包交换网络的数据包转发技术。它们构成 虚拟机逻辑网络 的基础。*

逻辑网络

逻辑网络允许根据环境要求隔离网络流量。逻辑网络的类型有：

- *承载虚拟机网络流量的逻辑网络，*
- *不承载虚拟机网络流量的逻辑网络，*
- *可选的逻辑网络，*
- *和必需的网络。*

所有逻辑网络都可以是必需的或可选网络。

承载虚拟机网络流量的逻辑网络在主机级别作为软件网桥设备实施。默认情况下，在安装 Red Hat Virtualization Manager 时定义一个逻辑网络： ***ovirtmgmt*** *管理网络。*

管理员可以添加的其他逻辑网络有：专用存储逻辑网络，以及专用显示逻辑网络。不承载虚拟机流量的逻辑网络在主机上没有关联的网桥设备。它们直接与主机网络接口关联。

Red Hat Virtualization 将与管理相关的网络流量与迁移相关的网络流量隔离。这样，可以将专用网络（没有路由）用于实时迁移，并确保在迁移过程中无法丢失与 hypervisor 的连接。

不同层上的逻辑网络的说明

逻辑网络对虚拟化环境的每个层有不同的影响。

数据中心层

逻辑网络在数据中心级别上定义。默认情况下，每个数据中心都有 **ovirtmgmt** 管理网络。进一步的逻辑网络是可选的，但推荐使用。可以在数据中心级别上设置作为 **虚拟机网络** 和自定义 MTU。为数据中心定义的逻辑网络还必须添加到使用逻辑网络的集群中。

集群层

逻辑网络从数据中心提供，必须添加到将使用它们的集群中。默认情况下，每个集群都连接到管理网络。您可以选择添加到为集群父数据中心定义的集群逻辑网络。将必需的逻辑网络添加到集群后，必须为集群中的每个主机实施它。可选的逻辑网络可根据需要添加到主机中。

主机层

虚拟机逻辑网络是针对集群中每个主机实施的，作为与给定网络接口关联的软件网桥设备。非虚拟机逻辑网络没有关联的网桥，并直接与主机网络接口关联。由于 Red Hat Virtualization 环境中包含，每个主机已将管理网络作为桥接实施。添加至群集的进一步必需的逻辑网络必须与每个主机上的网络接口关联，才能变为正常运行状态。

虚拟机层

逻辑网络可用方式与网络对物理机可用的方式相同。虚拟机可以将其虚拟 NIC 连接到运行它的主机上的任何虚拟机逻辑网络。然后，虚拟机会获得到它所连接的逻辑网络上可用的任何其他设备或目的地的连接。

> **例 1.1. 管理网络**
>
> 安装 Red Hat Virtualization Manager 时，会自动创建名为 **ovirtmgmt** 的 management 逻辑网络。**ovirtmgmt** 网络专用于管理 Red Hat Virtualization Manager 和主机间的流量。如果没有设置其他用途的网桥，**ovirtmgmt** 是所有流量的默认网桥。

1.7. 数据中心

数据中心是 Red Hat Virtualization 中最高级别的抽象。数据中心是一个容器，它由三种类型的子容器组成：

- 存储容器 包含有关存储类型和存储域的信息，包括存储域的连接信息。存储是为数据中心定义的，并可用于数据中心中的所有集群。数据中心中的所有主机集群都能够访问相同的存储域。
- 网络容器 包含有关数据中心的逻辑网络的信息。这包括网络地址、VLAN 标签和 STP 支持等详情。逻辑网络是为数据中心定义的，也可选择性地在集群级别实施。
- 集群容器 包含集群。集群是具有兼容处理器内核的主机组，可以是 AMD 或 Intel 处理器。集群是迁移域；虚拟机可以实时迁移到集群中的任何主机，而不能实时迁移到其他集群。一个数据中心可以保存多个集群，每个集群可以包含多个主机。

第 2 章 存储

2.1. 存储域概述

存储域是一组具有通用存储接口的镜像集合。存储域包含模板和虚拟机（包括快照）、ISO 文件和自身元数据的完整镜像。存储域可以由块设备(SAN - iSCSI 或 FCP)或者文件系统(NAS - NFS、GlusterFS 或其他 POSIX 兼容文件系统)组成。

在 NAS 上，所有虚拟磁盘、模板和快照都是文件。

在 SAN (iSCSI/FCP)上，每个虚拟磁盘、模板或快照都是逻辑卷。块设备聚合到名为卷组的逻辑实体中，然后由 LVM （逻辑卷管理器）划分为逻辑卷，用作虚拟硬盘。有关 LVM 的详情，请查看 Red Hat Enterprise Linux 逻辑卷管理器管理指南。

虚拟磁盘可以具有两种格式之一，可以是 QCOW2 或 RAW。存储的类型可以是 Sparse 或 Preallocated。快照始终是稀疏的，但可以为作为 RAW 或稀疏创建的磁盘获取快照。

共享相同存储域的虚拟机可以在属于同一集群的主机之间迁移。

2.2. 存储备份存储域的类型

存储域可以使用基于块和基于文件的存储来实施。

基于文件的存储

Red Hat Virtualization 支持的基于文件的存储类型包括 NFS、GlusterFS、其他 POSIX 兼容文件系统，以及主机本地存储。

基于文件的存储在 Red Hat Virtualization 环境外部管理。

NFS 存储由 Red Hat Enterprise Linux NFS 服务器或其他第三方网络附加存储服务器管理。

主机可以管理自己的本地存储文件系统。

基于块的存储

块存储使用未格式化的块设备。逻辑卷管理器(LVM)将块设备聚合到卷组中。LVM 实例在所有主机上运行并不知道在其他主机上运行的实例。VDSM 通过扫描卷组来更改在 LVM 之上添加集群逻辑。当检测到更改时，VDSM 会通知单个主机刷新其卷组信息，从而更新各个主机。主机将卷组划分为逻辑卷，将逻辑卷元数据写入磁盘。如果在现有存储域中添加了更多存储容量，Red Hat Virtualization Manager 会导致每个主机中的 VDSM 刷新卷组信息。

逻辑单元号(LUN)是一个单独的块设备。支持的块存储协议(iSCSI, FCoE, 或 SAS)之一用于连接 LUN。Red Hat Virtualization Manager 管理到 LUN 的软件 iSCSI 连接。所有其他块存储连接都在 Red Hat Virtualization 环境外部管理。基于块的存储环境中的任何更改，如创建逻辑卷、扩展或删除逻辑卷，以及添加新 LUN，由称为存储池管理器的特殊选择的主机上由 LVM 处理。然后，VDSM 会同步更改，该存储元数据会在集群中的所有主机上刷新。

2.3. 存储域类型

Red Hat Virtualization 支持三种类型的存储域，包括每个存储域支持的存储类型：

- 数据存储域 存储在 Red Hat Virtualization 环境中所有虚拟机的硬盘镜像。磁盘映像可能包含已安装的操作系统或虚拟机存储或生成的数据。数据存储域支持 NFS、iSCSI、FCP、GlusterFS 和 POSIX 兼容存储。数据域无法在多个数据中心之间共享。

- 导出存储域为数据中心之间传输的硬盘镜像和虚拟机模板提供传输存储。另外，导出存储域存储虚拟机的备份副本。导出存储域支持 NFS 存储。多个数据中心可以访问单个导出存储域，但一次只能使用一个数据中心。

- ISO 存储域存储 ISO 文件，也称为 images。ISO 文件是物理 CD 或者 DVD 的表示。在 Red Hat Virtualization 环境中，常见的 ISO 文件类型是操作系统安装磁盘、应用程序安装磁盘和客户机代理安装磁盘。这些镜像可以附加到虚拟机，并以与将物理磁盘插入到磁盘驱动器和引导的方式相同。ISO 存储域允许数据中心中的所有主机共享 ISO，无需物理光盘介质。

2.4. 虚拟磁盘镜像的存储格式

QCOW2 格式虚拟机存储

QCOW2 是虚拟磁盘镜像的存储格式。QCOW 代表 在写入时的 QEMU 副本。QCOW2 格式通过在逻辑和物理块之间添加映射，将物理存储层与虚拟层分离。每个逻辑块都映射到其物理偏移，启用存储过量分配和虚拟机快照，其中每个 QCOW 卷只代表对底层磁盘镜像所做的更改。

初始映射将所有逻辑块指向备份文件或卷中的偏移量。当虚拟机在快照后将数据写入 QCOW2 卷时，从后备卷读取相关块，使用新信息修改并写入新快照 QCOW2 卷。然后，映射已被更新以指向新位置。

RAW

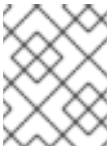
RAW 存储格式的性能优势与 QCOW2 相比，没有格式应用到以 RAW 格式存储的虚拟磁盘镜像。以 RAW 格式存储的磁盘镜像上的虚拟机数据操作不需要主机的额外工作。当虚拟机在其虚拟磁盘中将数据写入其虚拟磁盘中的给定偏移时，I/O 会写入后备文件或逻辑卷上的相同偏移。

原始格式要求预分配定义的映像的整个空间，除非使用来自存储阵列的外部管理的精简置备的 LUN。

2.5. 虚拟磁盘镜像存储分配策略

预分配存储

在创建虚拟机之前，分配了虚拟磁盘镜像所需的所有存储。如果为虚拟机创建了 20 GB 磁盘镜像，磁盘镜像将使用 20 GB 存储域容量。预分配的磁盘镜像无法放大。分配存储可能意味着加快写入时间，因为运行时没有进行存储分配，以灵活性为代价。以这种方式分配存储可减少 Red Hat Virtualization Manager 到过量使用存储的容量。建议将预分配存储用于高水平 I/O 任务，且对存储延迟的容错程度较低。通常，服务器虚拟机适合此描述。



注意

如果使用存储后端提供的精简配置功能，则虚拟机调配存储时仍应从管理门户中选择预分配存储。

稀疏分配存储

虚拟磁盘镜像的上限是在虚拟机创建时设置的。最初，磁盘镜像不使用任何存储域容量。当虚拟机将数据写入磁盘时，用量会增长，直到达到上限为止。删除磁盘镜像中的数据时，不会返回到存储域的容量。稀疏分配的存储适合具有低或中低或中度 I/O 任务的虚拟机，且对存储延迟进行一些容错。通常，桌面虚拟机适合此描述。

注意

如果您的存储后端提供了精简配置功能，则应将其用作精简配置的首选实施。存储应从图形用户界面置备，作为预分配，将精简配置留给后端解决方案。

2.6. RED HAT VIRTUALIZATION 中的存储元数据版本

Red Hat Virtualization 将存储域的信息存储为存储域本身的元数据。Red Hat Virtualization 的每个主发行版本都看到了存储元数据的改进实现。

- V1 元数据(Red Hat Virtualization 2.x 系列)

  每个存储域都包含描述其自身结构的元数据，以及用于备份虚拟磁盘镜像的所有物理卷名称。

  Master 域还包含存储池中所有域和物理卷名称的元数据。这个元数据的总大小限制为 2 KB，限制池中的存储域数量。

  模板和虚拟机基础镜像是只读的。

  V1 元数据适用于 NFS、iSCSI 和 FC 存储域。

- V2 元数据(Red Hat Enterprise Virtualization 3.0)

  所有存储域和池元数据都作为逻辑卷标签存储，而不是写入逻辑卷。有关虚拟磁盘卷的元数据仍然存储在域中的逻辑卷中。

  元数据中不再包含物理卷名称。

  模板和虚拟机基础镜像是只读的。

  V2 元数据适用于 iSCSI 和 FC 存储域。

- V3 元数据(Red Hat Enterprise Virtualization 3.1+)

  所有存储域和池元数据都作为逻辑卷标签存储，而不是写入逻辑卷。有关虚拟磁盘卷的元数据仍然存储在域中的逻辑卷中。

  虚拟机和模板基础镜像不再是只读的。这个更改启用了实时快照、实时存储迁移和从快照克隆。

  添加了对 unicode 元数据的支持，对于非英语卷名称。

  V3 元数据适用于 NFS、GlusterFS、POSIX、iSCSI 和 FC 存储域。

2.7. RED HAT VIRTUALIZATION 中的存储域自动恢复

Red Hat Virtualization 环境中的主机通过从每个域读取元数据来监控其数据中心中的存储域。当数据中心中的所有主机报告它们无法访问存储域时，存储域将变得不活动状态。

例如，由于临时网络中断，管理器会假设存储域暂时处于非活动状态，而不是断开不活跃的存储域。每 5 分钟后，管理器会尝试重新激活任何不活跃存储域。

可能需要管理员干预来补救存储连接中断的原因，但管理器会在恢复连接时处理重新激活存储域。

2.8. 存储池管理程序

Red Hat Virtualization 使用元数据来描述存储域的内部结构。structural 元数据被写入每个存储域的片段。主机使用基于单个写入器和多个读取器配置的存储域元数据。存储域结构元数据跟踪镜像和快照创建和删除，以及卷和域扩展。

可以对数据域的结构进行更改的主机称为存储池管理器(SPM)。SPM 协调数据中心中的所有元数据更改，如创建和删除磁盘镜像、创建和合并快照、在存储域之间复制镜像、为块设备创建模板和存储分配。每个数据中心都有一个 SPM。所有其他主机只能读取存储域结构元数据。

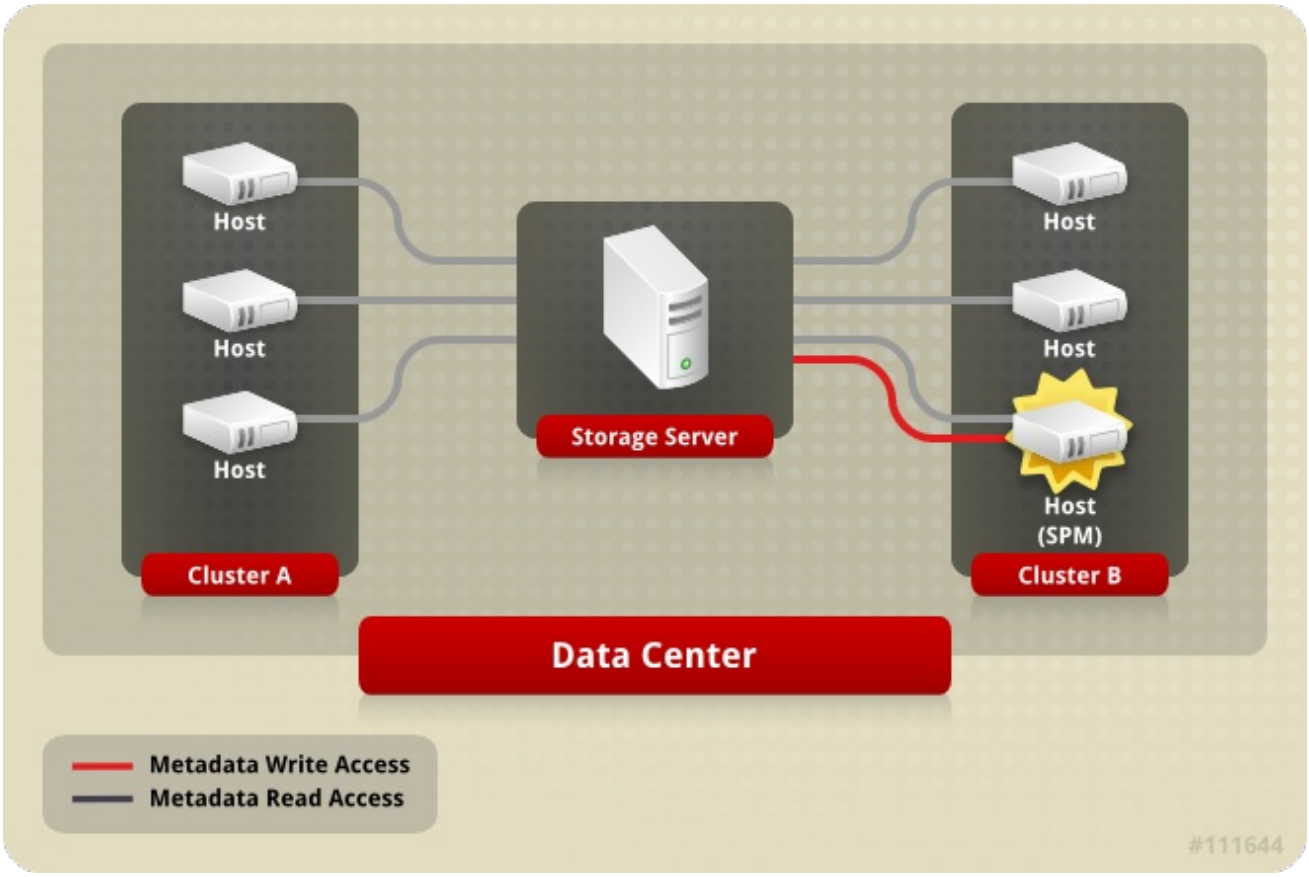
主机可以手动选择为 SPM，或者可由 Red Hat Virtualization Manager 分配。管理器通过导致潜在的 SPM 主机尝试假定 以存储为中心的租用 来分配 SPM 角色。租期允许 SPM 主机写入存储元数据。它是以存储为中心的，因为它被写入存储域，而不是由 Manager 或主机跟踪。以存储为中心的租用将写入主存储域中的特殊逻辑卷，名为 **leases**。有关数据域结构的元数据被写入一个名为 **metadata** 的特殊逻辑卷。对 **元数据** 逻辑卷的更改会受到 **leases** 逻辑卷的影响。

管理器使用 VDSM 向主机发出 **spmStart** 命令，从而导致该主机上的 VDSM 尝试假定以存储为中心的租用。如果主机成功成为 SPM，并保留以存储为中心的租用，直到 Red Hat Virtualization Manager 请求新主机假定 SPM 角色。

如果出现以下情况，管理器会将 SPM 角色移到另一主机：

- SPM 主机无法访问所有存储域，但可以访问主存储域。
- 由于存储连接丢失或租期卷已满且不执行写入操作，因此 SPM 主机无法续订租期。
- SPM 主机崩溃。

图 2.1. 存储池管理程序 Exclusively Writes Structural Metadata。



2.9. 存储池管理器选择过程

如果主机尚未手动分配存储池管理器(SPM)角色，则 Red Hat Virtualization Manager 启动和管理 SPM 选择过程。

首先，Red Hat Virtualization Manager 请求 VDSM 确认哪个主机具有以存储为中心的租用。

Red Hat Virtualization Manager 从最初创建存储域后跟踪 SPM 分配历史记录。SPM 角色的可用性通过三种方式确认：

- "getSPMstatus"命令：管理器使用 VDSM 与具有 SPM 状态的主机检查，并接收"SPM"之一、"Contending"或"Free"。
- 存储域的元数据卷包含具有 SPM 状态的最后一个主机。
- 存储域的元数据卷包含 SPM 状态的最后一个主机的版本。

如果正常运行，则响应的主机会保留以存储为中心的租用，Red Hat Virtualization Manager 会在管理员门户中标记该主机 SPM。不做进一步操作。

如果 SPM 主机没有响应，它被视为不可访问。如果为主机配置了电源管理，则会自动隔离。如果没有，则需要手动隔离。在隔离了以前的存储池管理程序之前，无法将存储池管理器角色分配给新主机。

当 SPM 角色和存储以存储为中心的租期空闲时，Red Hat Virtualization Manager 会将它们分配给数据中心中随机选择的操作主机。

如果新主机上 SPM 角色分配失败，Red Hat Virtualization Manager 会将主机添加到包含操作失败的主机列表中，将这些主机标记为 SPM 角色。在下一个 SPM 选择流程开始时清除此列表，以便所有主机都再次符合条件。

Red Hat Virtualization Manager 继续请求存储池管理器角色和存储中心租期被一个随机选择的主机假定，直到 SPM 选择成功为止。

每次当前 SPM 没有响应或无法履行其职责时，Red Hat Virtualization Manager 都会启动存储池管理器选择过程。

2.10. RED HAT VIRTUALIZATION 中的专用资源和 SANLOCK

Red Hat Virtualization 环境中的某些资源必须专门访问。

SPM 角色是一个这样的资源。如果有多个主机成为 SPM，则数据崩溃的风险可能同时从两个位置更改同一数据。

在 Red Hat Enterprise Virtualization 3.1 之前，使用名为 safelease 的 VDSM 功能维护并跟踪 SPM。该租期被写入数据中心所有存储域的特殊区域。环境中的所有主机都可以以网络独立的方式跟踪 SPM 状态。VDSM 的安全租用仅维护一个资源的exclusivity：SPM 角色。

sanlock 提供相同的功能，但将 SPM 角色视为可以锁定的资源之一。sanlock 更为灵活，因为它允许锁定其他资源。

需要资源锁定的应用程序可以使用 Sanlock 注册。然后注册的应用程序可以代表它请求 Sanlock 锁定资源，以便其他应用程序都无法访问它。例如，VDSM 现在不锁定 SPM 状态，而是请求 Sanlock 这样做。

在 锁定空间 的磁盘上跟踪锁定。每个存储域都有一个锁定空间。如果 SPM 资源上的锁定，每个主机的存活度由主机在锁定空间中跟踪，因为主机在它连接到存储时从 Manager 接收的 hostid，并以固定间隔将时间戳写入锁定间隔。ids 逻辑卷跟踪每个主机的唯一标识符，并在每次主机更新其 hostid 时更新。SPM 资源只能由实时主机持有。

在 租期 逻辑卷中在磁盘上跟踪资源。当其在磁盘上表示已使用已获取的进程的唯一标识符进行更新时，会采取资源。如果是 SPM 角色，则 SPM 资源使用已获取它的 hostid 进行更新。

每个主机上的 Sanlock 进程只需要检查一次资源，才能看到它们被使用。在进行初始检查后，Sanlock 可以监控锁定空间，直到具有锁定资源的主机的时间戳过时。

sanlock 监控使用资源的应用程序。例如，VDSM 是针对 SPM 状态和 hostid 的监控。如果主机无法从 Manager 续订它的主机，它会丢失锁定空间中所有资源的排除。sanlock 更新资源以显示不再被使用。

如果 SPM 主机无法在指定时间内将时间戳写入存储域上的锁定空间，则主机的 Sanlock 请求实例会释放其资源。如果 VDSM 进程响应，则释放其资源，并且锁定空间中的 SPM 资源可由其他主机执行。

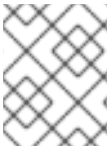
如果 SPM 主机上的 VDSM 没有响应释放资源的请求，则主机上的 Sanlock 将终止 VDSM 进程。如果 kill 命令失败，Sanlock 会试图使用 sigkill 来终止 VDSM。如果 sigkill 失败，Sanlock 依赖于 watchdog 守护进程来 重启主机。

每次主机上的 VDSM 续订其 hostid 并将时间戳写入锁定空间时，watchdog 守护进程都会收到 pet。当 VDSM 无法这样做时，watchdog 守护进程不再会被请求。在 watchdog 守护进程在给定时间内没有收到 pet 后，它会重启主机。达到的最终升级级别，保证 SPM 资源已被释放，并可以被另一主机执行。

2.11. 精简配置和存储覆盖提交

Red Hat Virtualization Manager 提供置备策略来优化虚拟化环境中的存储使用。精简配置策略允许您过度使用存储资源，根据虚拟化环境的实际存储使用情况置备存储。

存储过量分配是为虚拟机分配比存储池中物理可用的存储更多存储。通常，虚拟机使用的存储比已分配给它们的存储要少。精简配置允许虚拟机像为它定义的存储一样运行，如果实际上只分配了部分存储。



注意

虽然 Red Hat Virtualization Manager 提供了自己的精简配置功能，但如果提供存储后端，应使用存储后端的精简配置功能。

为了支持存储过量使用，在 VDSM 中定义阈值，该阈值将逻辑存储分配与实际存储使用情况进行比较。这个阈值用于确保写入磁盘镜像的数据小于支持它的逻辑卷。QEMU 识别逻辑卷中写入的最高偏移值，这表示最大存储使用点。VDSM 监控 QEMU 标记的最高偏移，以确保用量不会超过定义的阈值。因此，当 VDSM 继续表明最高偏移仍低于阈值时，Red Hat Virtualization Manager 知道问题中的逻辑卷有足够的存储才能继续操作。

当 QEMU 表示使用量超过阈值限制时，VDSM 与管理器通信时，磁盘镜像很快达到其逻辑卷的大小。Red Hat Virtualization Manager 请求 SPM 主机扩展逻辑卷。只要数据中心的数据存储域有可用空间，就可以重复此过程。当数据存储域耗尽可用空间时，您必须手动添加存储容量才能扩展它。

2.12. 逻辑卷扩展

Red Hat Virtualization Manager 使用精简配置来过量使用存储池中可用的存储，并分配比物理可用更多的存储。虚拟机在数据运行时写入数据。具有精简配置的磁盘镜像的虚拟机最终将写入比其磁盘镜像可以保存的逻辑卷更多的数据。当发生这种情况时，逻辑卷扩展用于提供额外的存储，并促进虚拟机的持续操作。

Red Hat Virtualization 通过 LVM 提供了一个精简配置机制。当使用 QCOW2 格式化的存储时，Red Hat Virtualization 依赖于主机系统处理 qemu-kvm，来以有序的方式将磁盘上的存储块映射到逻辑块。例如，这允许定义由 1 GB 逻辑卷支持的逻辑 100 GB 磁盘。当 qemu-kvm 超过 VDSM 设置的使用阈值时，本地 VDSM 实例会向 SPM 发出对 SPM 的请求，以便由另一个 1GB 扩展。在必须扩展卷扩展的主机上，运行虚拟机的 VDSM 会通知 SPM VDSM 需要更多空间。SPM 扩展逻辑卷和 SPM VDSM 实例会导致主机 VDSM 刷新卷组信息，并识别扩展操作已完成。主机可以继续操作。

逻辑卷扩展不要求主机知道哪个其他主机是 SPM；它甚至是 SPM 本身。存储扩展通信通过存储邮箱完成。存储邮箱是数据存储域中的专用逻辑卷。需要 SPM 扩展逻辑卷的主机会在指定给存储邮箱中特定主机的区域写入消息。SPM 定期读取传入邮件，执行请求的逻辑卷扩展，并在传出邮件中写入回复。发送

请求后，主机会监控其传入邮件的每两秒钟的响应。当主机收到成功回复其逻辑卷扩展请求时，它会在设备映射中刷新逻辑卷映射来识别新分配的存储。

当存储池可用的物理存储被接近用尽时，多个镜像可能会耗尽可用的存储，且无方法重新配置其资源。耗尽其存储的存储池会导致 QEMU 返回 **enospc** 错误，这表示该设备不再有可用的存储。此时，正在运行的虚拟机会自动暂停，并且需要手动干预，才能向卷组添加新的 LUN。

当向卷组添加新 LUN 时，存储池管理程序会自动将额外的存储分发到需要它的逻辑卷。自动分配额外资源可让相关虚拟机自动继续操作，如果停止，则恢复操作。

第 3 章 NETWORK

3.1. 网络架构

Red Hat Virtualization 网络可以在基本网络、集群内的网络和主机网络配置中讨论。基本网络术语涵盖有助于联网的基本硬件和软件元素。集群内的网络包括集群级别对象（如主机、逻辑网络和虚拟机）之间的网络交互。主机网络配置涵盖了主机中联网的支持的配置。

例如，一个精心设计并构建的网络可确保高带宽任务收到适当的带宽，用户交互不会因为延迟而产生，并且虚拟机可以在迁移域中成功迁移。构建较差的网络可能会导致，例如无法接受的延迟，以及网络激增的迁移和克隆故障。

3.2. 简介：基本网络术语

Red Hat Virtualization 使用以下内容提供虚拟机、虚拟化主机和更广泛的网络之间的网络功能：

- 网络接口控制器(NIC)
- Bridge
- Bond
- 虚拟 NIC
- 虚拟 LAN (VLAN)

NIC、网桥和 VNIC 允许主机、虚拟机、本地区域网络和互联网之间的网络通信。绑定和 VLAN 可以选择实施，以增强安全性、容错和网络容量。

3.3. NETWORK INTERFACE CONTROLLER

NIC （网络接口控制器） 是一个网络适配器或 LAN 适配器，用于将计算机连接到计算机网络。NIC 在机器的物理和虚拟数据链路层上运行，并允许网络连接。Red Hat Virtualization 环境中所有虚拟化主机都至少有一个 NIC，但主机更常见的是具有两个或多个 NIC。

一个物理 NIC 可以逻辑地连接多个虚拟 NIC (VNIC)。虚拟 NIC 充当虚拟机的物理网络接口。为了区分 VNIC 和支持它的 NIC，Red Hat Virtualization Manager 会为每个 VNIC 分配一个唯一的 MAC 地址。

3.4. BRIDGE

Bridge 是一个在数据包交换网络中使用数据包转发的软件设备。桥接允许多个网络接口设备共享一个 NIC 的连接，并作为单独的物理设备出现在网络中。网桥检查数据包的源地址，以确定相关的目标地址。确定目标地址后，网桥会将位置添加到表中以备将来参考。这允许主机将网络流量重定向到作为网桥成员的虚拟机关联的 VNIC。

在 Red Hat Virtualization 中，逻辑网络是使用网桥实现的。它是接收 IP 地址的主机上的网桥而不是物理接口。与网桥关联的 IP 地址不需要与使用该网桥的虚拟机位于同一个子网中。如果网桥被分配了与使用它的虚拟机在同一子网中的 IP 地址，则主机可以通过虚拟机在逻辑网络内寻址。作为规则，不建议在虚拟化主机上运行网络公开的服务。客户机通过其 VNIC 连接到逻辑网络，主机则通过其 NIC 连接到逻辑网络的远程元素。每个客户机都可以独立设置其 VNIC 的 IP 地址，具体由 DHCP 或静态设置。网桥可以连接到主机之外的对象，但这样的连接不强制使用。

可以为网桥和以太网连接定义自定义属性。VDSM 将网络定义和自定义属性传递给设置网络 hook 脚本。

3.5. BONDS

绑定是将多个网络接口卡聚合到单个软件定义设备中。因为绑定的网络接口结合了绑定中包含的网络接口卡的传输功能，以充当单个网络接口，所以它们可以提供比单个网络接口卡更高的传输速度。另外，因为绑定中的所有网络接口卡都必须失败，因此绑定本身都会导致容错功能增加。但是，一个限制是形成绑定网络接口的网络接口卡必须相同，以确保绑定中的所有网络接口卡都支持相同的选项和模式。

绑定的数据包分布算法由使用的绑定模式决定。

重要

模式 1、2、3 和 4 支持虚拟机（桥d）和非虚拟机（无桥）网络类型。模式 0、5 和 6 仅支持非虚拟机（无网桥）网络。

绑定模式

Red Hat Virtualization 默认使用模式 4，但支持以下通用绑定模式：

模式 0 （循环策略）

按顺序通过网络接口卡传输数据包。数据包在循环中传输，该循环以绑定中的第一个可用网络接口卡开始，并以绑定中最后一个可用网络接口卡结尾。然后，所有后续循环都从第一个可用的网络接口卡开始。模式 0 提供容错，并在绑定中的所有网络接口卡之间平衡负载。但是，模式 0 无法与网桥结合使用，因此与虚拟机逻辑网络不兼容。

模式 1 （主动备份策略）

在一个网络接口卡保持活动状态时，将所有网络接口卡设置为备份状态。如果活跃网卡中的故障，其中一个备份网卡将替换网卡作为绑定中唯一活跃的网卡。模式 1 中的绑定 MAC 地址只在一个端口上可见，以防止在绑定的 MAC 地址更改了活跃网络接口卡时导致的混淆。模式 1 提供容错功能，在 Red Hat Virtualization 中受到支持。

模式 2 (XOR 策略)

选择根据源和目标 MAC 地址 modulo 网络接口卡从计数来传输数据包结果的网络接口卡。此计算可确保为每个使用的目的地 MAC 地址选择相同的网卡。模式 2 提供容错和负载平衡，在 Red Hat Virtualization 中受到支持。

模式 3 （广播策略）

将所有数据包传输到所有网络接口卡。模式 3 提供容错功能，在 Red Hat Virtualization 中受到支持。

模式 4 (IEEE 802.3ad 策略)

创建聚合组，其中接口共享相同的速度和双工设置。模式 4 根据 IEEE 802.3ad 规格，使用活跃聚合组中的所有网络接口卡，并受 Red Hat Virtualization 支持。

模式 5 (adaptive 传输负载平衡策略)

确保传出流量帐户的发布，用于绑定中的每个网络接口卡上的负载，并且当前网络接口卡接收所有传入流量。如果分配给接收流量的网络接口卡失败，则会为接收传入流量的角色分配另一个网卡。模式 5 无法与网桥结合使用，因此它与虚拟机逻辑网络不兼容。

模式 6 (adaptive 负载均衡策略)

组合模式 5 （协调传输负载平衡策略）以及 IPv4 流量的接收负载平衡，而无需任何特殊的交换机要求。ARP 协商用于平衡接收负载。模式 6 无法与网桥结合使用，因此它与虚拟机逻辑网络不兼容。

3.6. 绑定的切换配置

交换机配置根据您的硬件要求而有所不同。请参阅您的操作系统的部署和网络配置指南。

重要

对于每种类型的交换机，务必要通过 链路聚合控制协议 (LACP) 协议设置交换机绑定，而不是 Cisco 端口聚合 协议 (PAgP) 协议。

3.7. 虚拟网络接口卡

虚拟网络接口卡是基于主机物理网络接口卡的虚拟网络接口。每个主机可以有多个网络接口卡，每个网卡可以是多个虚拟网络接口卡的基础。

当您将虚拟网卡附加到虚拟机时，Red Hat Virtualization Manager 会在要附加虚拟网络接口卡的虚拟机之间创建多个关联、虚拟网络接口卡本身、虚拟网络接口卡本身，以及虚拟网络接口卡所基于的物理主机网卡。特别是，当将虚拟网络接口卡附加到虚拟机时，会在基于虚拟网络接口卡的物理主机网络接口卡上创建一个新的虚拟网络接口卡和 MAC 地址。然后，虚拟机在附加了虚拟网络接口卡后第一次启动时，**libvirt** 会为虚拟网络接口卡分配 PCI 地址。然后，使用 MAC 地址和 PCI 地址来获取虚拟机中的虚拟网络接口卡（如 **eth**0）的名称。

分配 MAC 地址并将这些 MAC 地址与 PCI 地址相关联的过程在基于模板或快照创建虚拟机时略有不同。为模板或快照创建 PCI 地址后，根据按照该模板或快照创建的虚拟机上的虚拟网络接口卡按照该顺序分配的 PCI 地址和 MAC 地址排序。如果还没有为模板创建 PCI 地址，则基于该模板创建的虚拟机上的虚拟网络接口卡按照虚拟网络接口卡的命名顺序分配。如果尚未为快照创建 PCI 地址，Red Hat Virtualization Manager 会为基于该快照的虚拟机上的虚拟网络接口卡分配新的 MAC 地址。

创建后，虚拟网络接口卡将添加到网桥设备中。网桥设备是虚拟机如何连接到虚拟机逻辑网络。

在虚拟化主机上运行 **ip addr show** 命令可显示与该主机上虚拟机关联的所有虚拟网络接口卡。也可查看为支持逻辑网络创建的任何网桥，以及主机使用的任何网络接口卡。

```
[root@rhev-host-01 ~]# ip addr show
1: lo: <LOOPBACK,UP,LOWER_UP> mtu 16436 qdisc noqueue state UNKNOWN
    link/loopback 00:00:00:00:00:00 brd 00:00:00:00:00:00
    inet 127.0.0.1/8 scope host lo
    inet6 ::1/128 scope host
       valid_lft forever preferred_lft forever
2: eth0: <BROADCAST,MULTICAST,UP,LOWER_UP> mtu 1500 qdisc pfifo_fast state UP qlen 1000
    link/ether 00:21:86:a2:85:cd brd ff:ff:ff:ff:ff:ff
    inet6 fe80::221:86ff:fea2:85cd/64 scope link
       valid_lft forever preferred_lft forever
3: wlan0: <NO-CARRIER,BROADCAST,MULTICAST,UP> mtu 1500 qdisc mq state DOWN qlen
1000
    link/ether 00:21:6b:cc:14:6c brd ff:ff:ff:ff:ff:ff
5: ;vdsmdummy;: <BROADCAST,MULTICAST> mtu 1500 qdisc noop state DOWN
    link/ether 4a:d5:52:c2:7f:4b brd ff:ff:ff:ff:ff:ff
6: bond0: <BROADCAST,MULTICAST,MASTER> mtu 1500 qdisc noop state DOWN
    link/ether 00:00:00:00:00:00 brd ff:ff:ff:ff:ff:ff
7: bond4: <BROADCAST,MULTICAST,MASTER> mtu 1500 qdisc noop state DOWN
    link/ether 00:00:00:00:00:00 brd ff:ff:ff:ff:ff:ff
8: bond1: <BROADCAST,MULTICAST,MASTER> mtu 1500 qdisc noop state DOWN
    link/ether 00:00:00:00:00:00 brd ff:ff:ff:ff:ff:ff
9: bond2: <BROADCAST,MULTICAST,MASTER> mtu 1500 qdisc noop state DOWN
    link/ether 00:00:00:00:00:00 brd ff:ff:ff:ff:ff:ff
```

```
10: bond3: <BROADCAST,MULTICAST,MASTER> mtu 1500 qdisc noop state DOWN
    link/ether 00:00:00:00:00:00 brd ff:ff:ff:ff:ff:ff
11: ovirtmgmt: <BROADCAST,MULTICAST,UP,LOWER_UP> mtu 1500 qdisc noqueue state
UNKNOWN
    link/ether 00:21:86:a2:85:cd brd ff:ff:ff:ff:ff:ff
    inet 10.64.32.134/23 brd 10.64.33.255 scope global ovirtmgmt
    inet6 fe80::221:86ff:fea2:85cd/64 scope link
       valid_lft forever preferred_lft forever
```

*命令的控制台输出显示多个设备：一个循环设备(**lo**)、一个以太网设备(**eth0**)、一个无线设备(**wlan0**)、一个 VDSM dummy 设备(**;vdsmdummy;**)、五个绑定设备(**bond0**、**bond4**、**bond1**、**bond2**、**bond3)**和一个网桥(**ovirtmgmt**)。*

*虚拟网络接口卡是网桥设备和逻辑网络的所有成员。使用 **brctl show** 命令可以显示网桥成员资格：*

```
[root@rhev-host-01 ~]# brctl show
bridge name bridge id  STP enabled interfaces
ovirtmgmt  8000.e41f13b7fdd4 no  vnet002
       vnet001
       vnet000
       eth0
```

***brctl show** 命令的控制台输出显示 virtio 虚拟网络接口卡是 **ovirtmgmt** 网桥的成员。虚拟网络接口卡关联的所有虚拟机都连接到 **ovirtmgmt** 逻辑网络。**eth0** 网络接口卡也是 **ovirtmgmt** 网桥的成员。**eth0** 设备将电缆连接到提供主机外连接的交换机。*

3.8. 虚拟局域网(VLAN)

VLAN （虚拟 LAN） 是一个可应用于网络数据包的属性。网络数据包可以"标记"到数字的 VLAN 中。VLAN 是一种安全功能，用于在交换机级别上完全隔离网络流量。VLAN 是完全独立的，相互排斥。Red Hat Virtualization Manager 的 VLAN 感知并能够标记和重定向 VLAN 流量，但 VLAN 实现需要一个支持 VLAN 的交换机。

在交换机级别，为端口分配 VLAN 设计。交换机应用 VLAN 标签到源自特定端口的流量，将流量标记为 VLAN 的一部分，并确保响应执行相同的 VLAN 标签。VLAN 可以在多个交换机之间扩展。交换机上带有 VLAN 标记的网络流量完全无法检测到，但连接到使用正确 VLAN 指定的端口的计算机除外。给定端口可以标记到多个 VLAN 中，允许多个 VLAN 的流量发送到单个端口，以使用接收流量的软件来分离。

3.9. 网络标签

网络标签可用于大大简化与创建和管理逻辑网络关联的几个管理任务，并将这些逻辑网络与物理主机网络接口和绑定相关联。

网络标签是纯文本、人类可读的标签，可附加到逻辑网络或物理主机网络接口中。标签长度没有严格的限制，但您必须使用小写字母和大写字母、下划线和连字符的组合；不允许使用空格或特殊字符。

将标签附加到逻辑网络或物理主机网络接口可创建与其他逻辑网络或物理主机网络接口关联，如下所示：

网络标签关联

- *当您将标签附加到逻辑网络时，该逻辑网络将自动与具有给定标签的任何物理主机网络接口关联。*

- 当您将标签附加到物理主机网络接口时，任何具有给定标签的逻辑网络将自动与该物理主机网络接口关联。
- 更改附加到逻辑网络或物理主机网络接口的标签的方式与删除标签和添加新标签的方式相同。相关的逻辑网络或物理主机网络接口之间的关联已更新。

网络标签和集群

- 当将标记的逻辑网络添加到集群中并且集群中有具有相同标签的物理主机网络接口时，逻辑网络将自动添加到该物理主机网络接口中。
- 当标记的逻辑网络从集群分离并且集群中有具有相同标签的物理主机网络接口时，逻辑网络将自动从该主机网络接口分离。

使用角色的网络标签和逻辑网络

- 当将标记的逻辑网络分配为显示网络或迁移网络时，将使用 DHCP 在物理主机网络接口上配置该逻辑网络，以便可以分配逻辑网络。

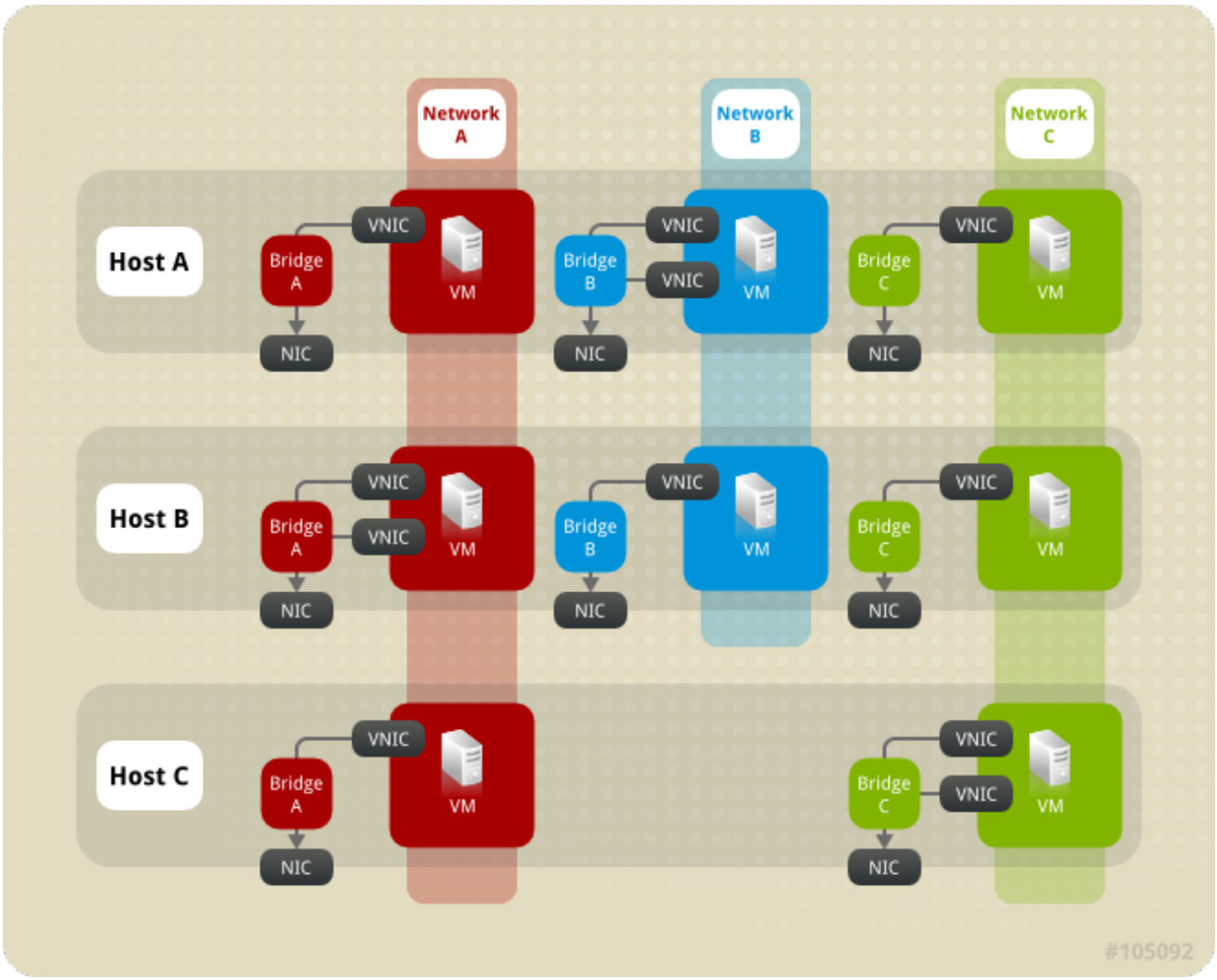
  在角色网络上设置标签（例如，"迁移网络"或"显示网络"）会导致在所有主机上大规模地部署该网络。这种大规模网络通过利用 DHCP 来实现。这种大规模部署方法是通过静态地址键入的方法进行选择，因为许多静态 IP 地址中键入的任务不可评分。

3.10. 集群网络

集群级别网络对象包括：

- 集群
- 逻辑网络

图 3.1. 集群中的网络



数据中心是多个集群的逻辑分组，每个集群都是多个主机的逻辑组。图 3.1 "集群中的网络" 描述单个集群的内容。

集群中的主机都可以访问同一存储域。集群中的主机也应用在集群级别的逻辑网络。要使虚拟机逻辑网络正常工作，必须使用 Red Hat Virtualization Manager 为集群中的每个主机定义和实施网络。其他逻辑网络类型只能在使用它们的主机上实施。

多主机网络配置会自动将任何更新的网络设置应用到分配到网络的所有主机。

3.11. 逻辑网络

*逻辑网络允许 Red Hat Virtualization 环境根据类型分隔网络流量。例如，在安装 Red Hat Virtualization 的过程中默认创建 **ovirtmgmt** 网络，用于管理 Manager 和主机之间的通信。逻辑网络的典型用途是将具有类似要求和用法类似的网络流量进行分组。在很多情况下，管理员会创建一个存储网络和显示网络，以隔离每种对应类型的流量进行优化和故障排除。*

逻辑网络的类型有：

- *承载虚拟机网络流量的逻辑网络，*
- *不承载虚拟机网络流量的逻辑网络，*
- *可选的逻辑网络，*

- 和必需的网络。

所有逻辑网络都可以是必需的或可选网络。

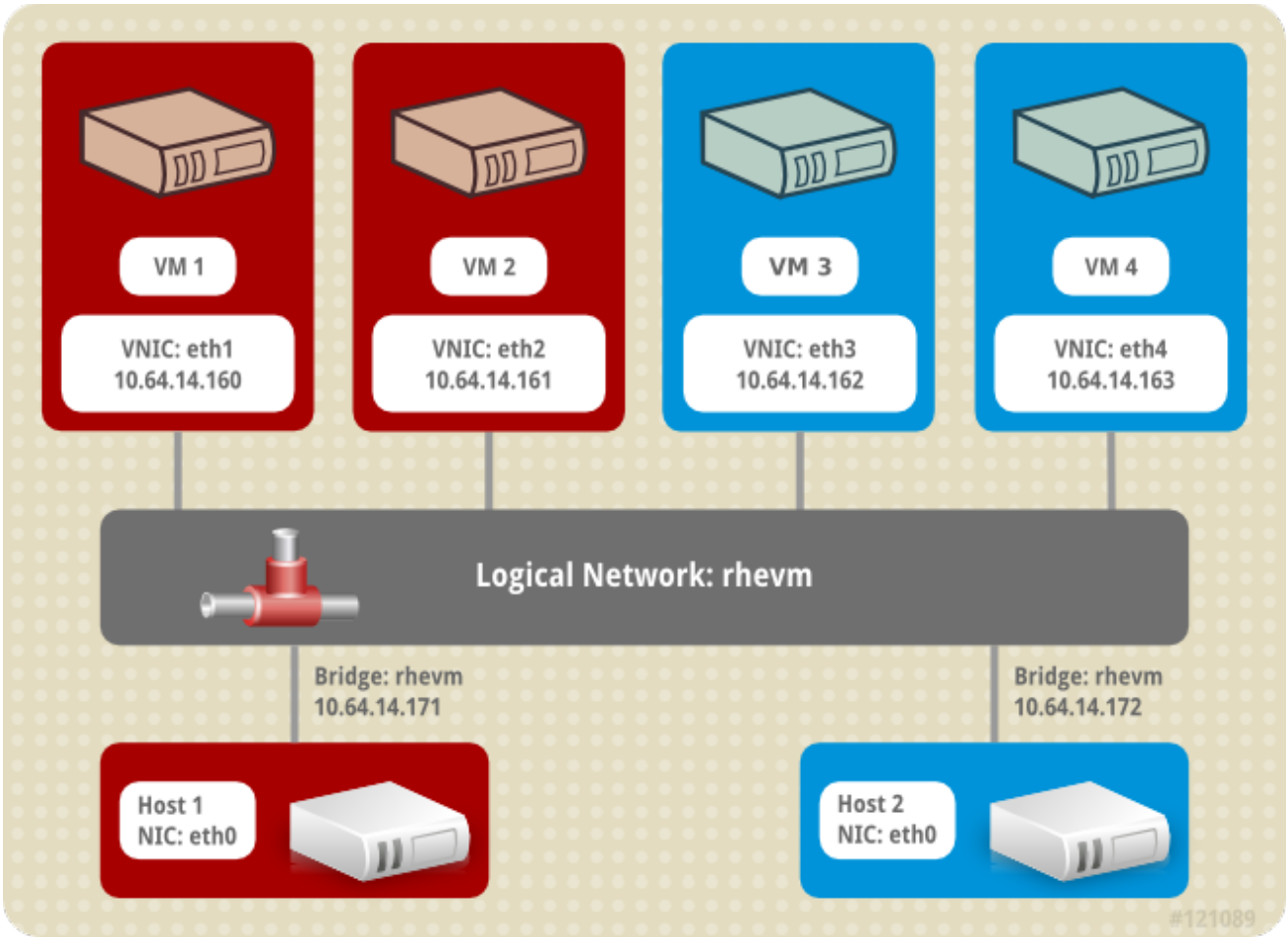
逻辑网络在数据中心级别上定义，并添加到主机。要使必需的逻辑网络正常运行，必须为给定集群中的每个主机实施。

Red Hat Virtualization 环境中的每个虚拟机逻辑网络都由主机上的网桥设备支持。因此，当为集群定义新的虚拟机逻辑网络时，必须在集群中的每个主机上创建一个匹配的网桥设备，然后才能使逻辑网络能够被虚拟机使用。Red Hat Virtualization Manager 会自动为虚拟机逻辑网络创建所需的网桥。

Red Hat Virtualization Manager 创建用来备份虚拟机逻辑网络的网桥设备与主机网络接口关联。如果作为网桥一部分的主机网络接口具有网络连接，则网桥中包含的任何网络接口共享网桥的网络连接。创建虚拟机并将其放置到特定的逻辑网络上时，其虚拟网卡将包含在该逻辑网络的网桥中。然后，这些虚拟机可以相互通信，以及与网桥连接的其他对象。

不用于虚拟机网络流量的逻辑网络直接与主机网络接口关联。

图 3.2. ovirtmgmt 逻辑网络。



例 3.1. 逻辑网络的示例用法.

在名为 Pink 的集群上，有两个称为 Red Hat 和 White 的主机，名为 Pink 的数据中心。Red Hat 和 White 都使用默认逻辑网络，**ovirtmgmt** 用于所有网络功能。负责 Pink 的系统管理员决定通过将 web 服务器和一些客户端虚拟机放在单独的逻辑网络上来隔离 Web 服务器的网络测试。她决定调用新的逻辑网络 **network_testing**。

首先，她为 Purple 数据中心定义逻辑网络。然后，她会将其应用到 Pink 集群。逻辑网络必须在处于维护模式的主机上实施。因此，管理员首先将所有正在运行的虚拟机迁移到红帽，并处于维护模式。然

> 后，她编辑与网桥中包含的物理网络接口关联的 **网络**。所选网络接口的 **Link Status** 将从 **Down** 更改为 **Non-Operational**。非工作状态是，因为必须通过将 Pink 集群中每一主机上的物理网络接口添加到 **network_testing** 网络，在集群中的所有主机中设置对应的网桥。接下来，她将激活 White，从红帽迁移所有正在运行的虚拟机，并为红帽重复该过程。
>
> 当 White 和 Red Hat 将 **network_testing** 逻辑网络桥接到物理网络接口时，**network_testing** 逻辑网络将变为 **Operational**，并可供虚拟机使用。

3.12. 所需的网络、可选网络和虚拟机网络

必需的网络是一个逻辑网络，必须可供集群中的所有主机使用。当主机的必需网络停止工作时，在该主机上运行的虚拟机将迁移到其他主机；此迁移的程度取决于所选的调度策略。如果您的虚拟机正在运行任务关键型工作负载，这将会很有用。

可选网络是一个逻辑网络，尚未明确声明 为**必需**。可选的网络只能在使用它们的主机上实施。可选网络的存在或不存在不会影响主机的 **Operational** 状态。当非必需网络停止工作时，网络上运行的虚拟机不会迁移到另一个主机。这可防止大量迁移导致不必要的 I/O 过载。请注意，当创建逻辑网络并添加到集群中时，默认会选中 **Required** 复选框。

要更改网络 **所需的** 指定，请从管理门户中选择一个网络，点 **Cluster** 选项卡，然后单击 **Manage Networks** 按钮。

虚拟机网络（在用户界面中称为 **虚拟机网络** ）是旨在仅承载虚拟机网络流量的逻辑网络。虚拟机网络可以是必需网络，也可以是可选的。使用可选虚拟机网络的虚拟机将仅在具有该网络的主机上启动。

3.13. 虚拟机连接

在 Red Hat Virtualization 中，虚拟机在创建虚拟机时将其 NIC 放置在逻辑网络上。此时，虚拟机可以与同一网络上的任何其他目标通信。

从主机的角度来看，当虚拟机放置在逻辑网络上时，支持虚拟机的 NIC 的 VNIC 将作为成员添加到逻辑网络的网桥设备中。例如，如果虚拟机位于 **ovirtmgmt** 逻辑网络上，其 VNIC 将添加为运行虚拟机的主机的 **ovirtmgmt** 网桥的成员。

3.14. 端口镜像

端口镜像会将给定逻辑网络和主机上的第 3 层网络流量复制到虚拟机上的虚拟接口。此虚拟机可用于网络调试和调优、入侵检测和监控同一主机和逻辑网络上其他虚拟机的行为。

复制的唯一流量是到一个主机上的一个逻辑网络的内部流量。主机外部网络上没有增加流量；但是，启用端口镜像的虚拟机使用比其他虚拟机更多的主机 CPU 和 RAM。

在逻辑网络的 vNIC 配置集中启用或禁用端口镜像，并有以下限制：

- 不支持使用启用了端口镜像的配置集热插 vNIC。
- 当 vNIC 配置集附加到虚拟机时，无法更改端口镜像。

鉴于上述限制，建议您在额外的专用 vNIC 配置文件中启用端口镜像。

重要

启用端口镜像可减少其他网络用户的隐私。

3.15. 主机网络配置

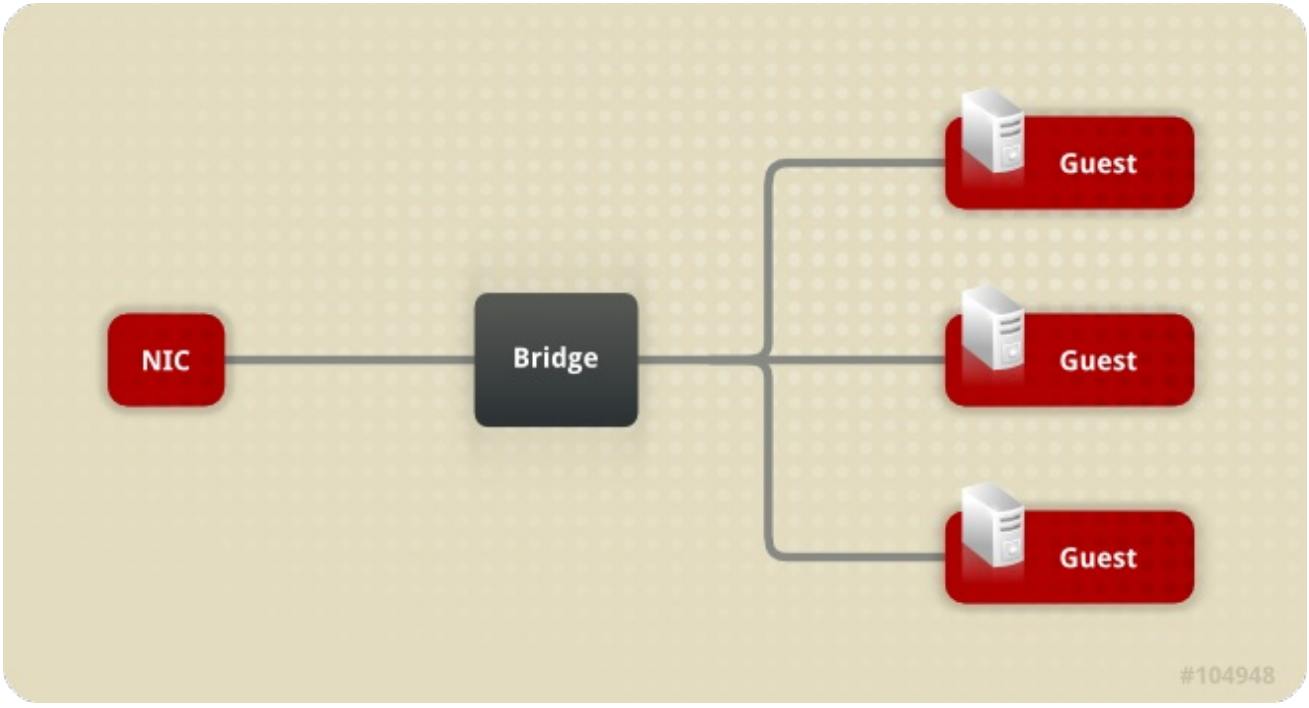
虚拟化主机的常见网络配置类型包括：

- 网桥和 NIC 配置.
- 网桥、VLAN 和 NIC 配置。
- 网桥、绑定和 VLAN 配置.
- 多个网桥、多个 VLAN 和 NIC 配置.

3.16. 网桥配置

Red Hat Virtualization 中最简单的主机配置是 Bridge 和 NIC 配置。如 图 3.3 “网桥和 NIC 配置” 描述，此配置使用网桥将一个或多个虚拟机（或客户机）连接到主机的 NIC。

图 3.3. 网桥和 NIC 配置

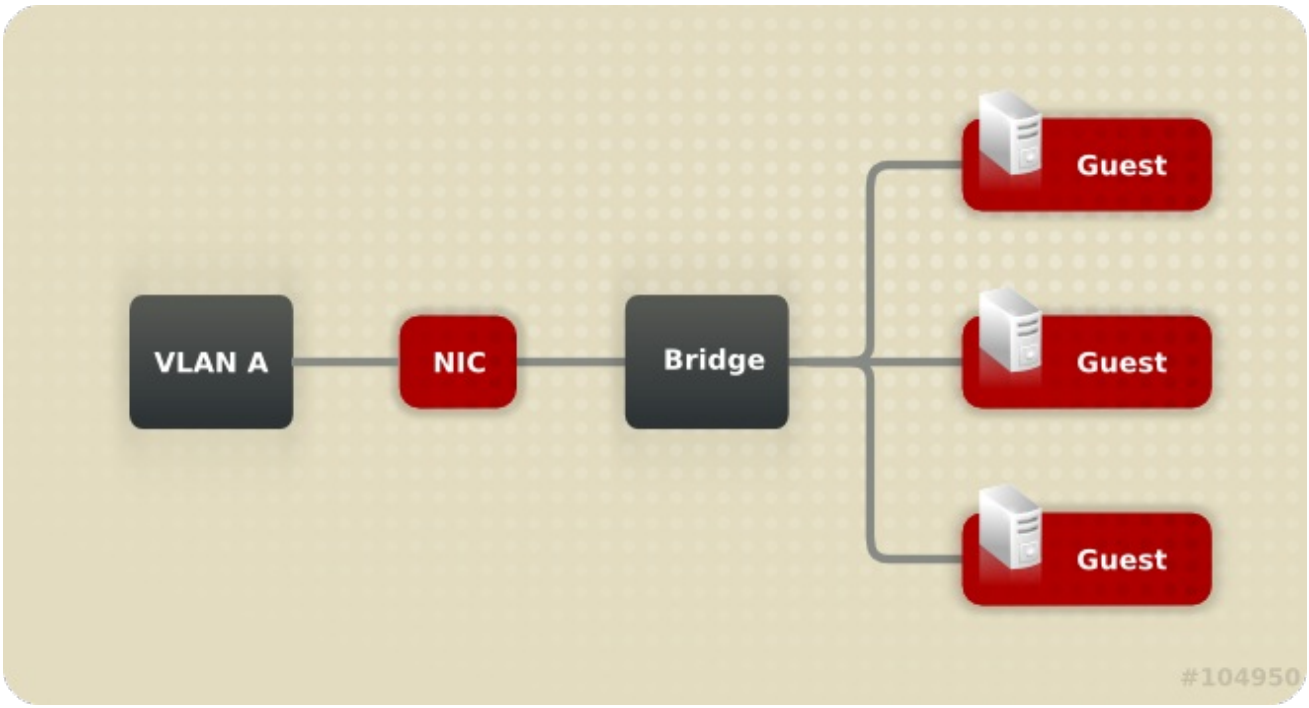


当安装 Red Hat Virtualization Manager 时，这个配置示例是自动创建 bridge **ovirtmgmt**。安装时，Red Hat Virtualization Manager 会在主机上安装 **VDSM**。**VDSM** 安装过程创建网桥 **ovirtmgmt**。然后，**ovirtmgmt** 网桥获取主机的 **IP** 地址，以启用主机的管理通信。

3.17. VLAN 配置

图 3.4 “网桥、VLAN 和 NIC 配置” 描述包含虚拟 LAN (VLAN)以连接主机 NIC 和网桥的替代配置。

图 3.4. 网桥、VLAN 和 NIC 配置

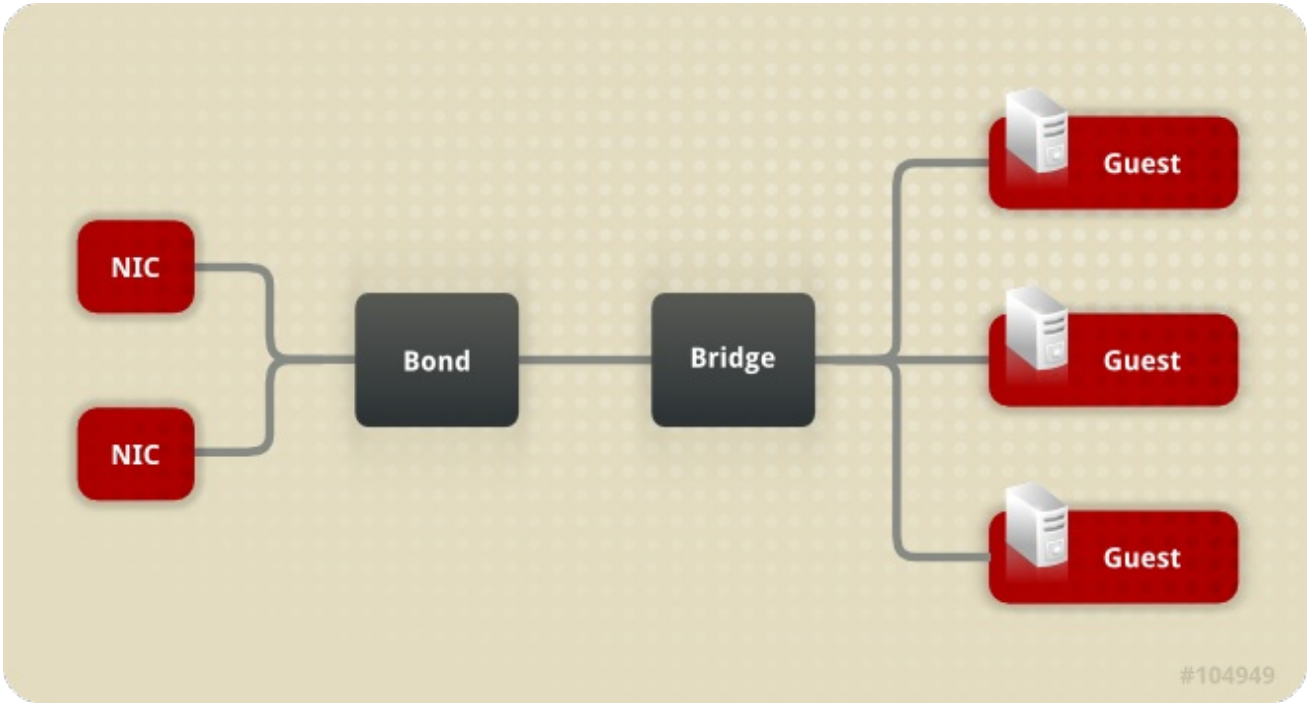


包含 VLAN，为通过此网络进行数据传输提供安全通道，同时还支持将多个网桥连接到使用多个 VLAN 的单个 NIC 的选项。

3.18. 网桥和绑定配置

图 3.5 "网桥、绑定和 NIC 配置" 显示包含绑定将多个主机 NIC 连接到同一网桥和网络的配置。

图 3.5. 网桥、绑定和 NIC 配置

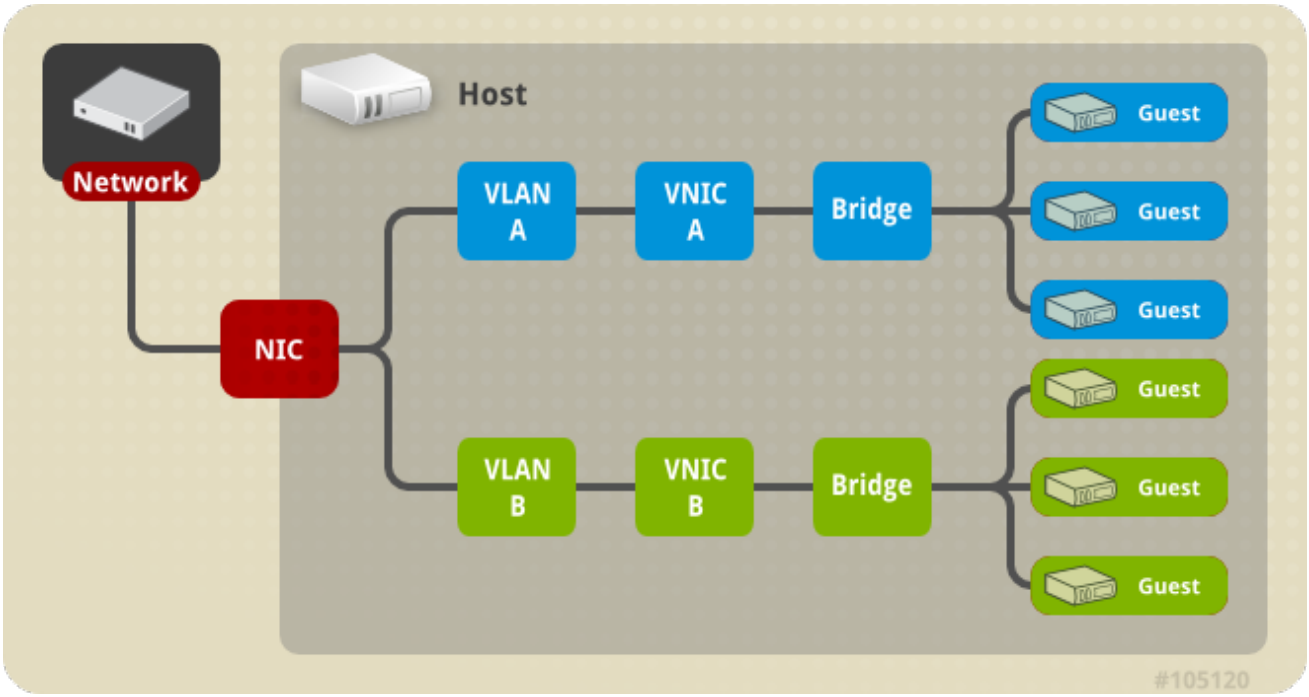


包括的绑定会创建一个组合两个（或更多）物理以太网链接的逻辑链接。结果的好处包括 NIC 容错和潜在的带宽扩展，具体取决于绑定模式。

3.19. 多个网桥、多个 VLAN 和 NIC 配置

图 3.6 “多个网桥、多个 VLAN 和 NIC 配置” 描述将单个 NIC 连接到两个 VLAN 的配置。这假设网络交换机已配置为将标记到两个 VLAN 之一的网络流量传递给主机上的一个 NIC。主机使用两个 VNIC 来分隔 VLAN 流量，每个 VLAN 都有一个。然后，标记为 VLAN 的流量通过将适当的 VNIC 作为网桥成员连接到单独的网桥。每个网桥由多个虚拟机连接到。

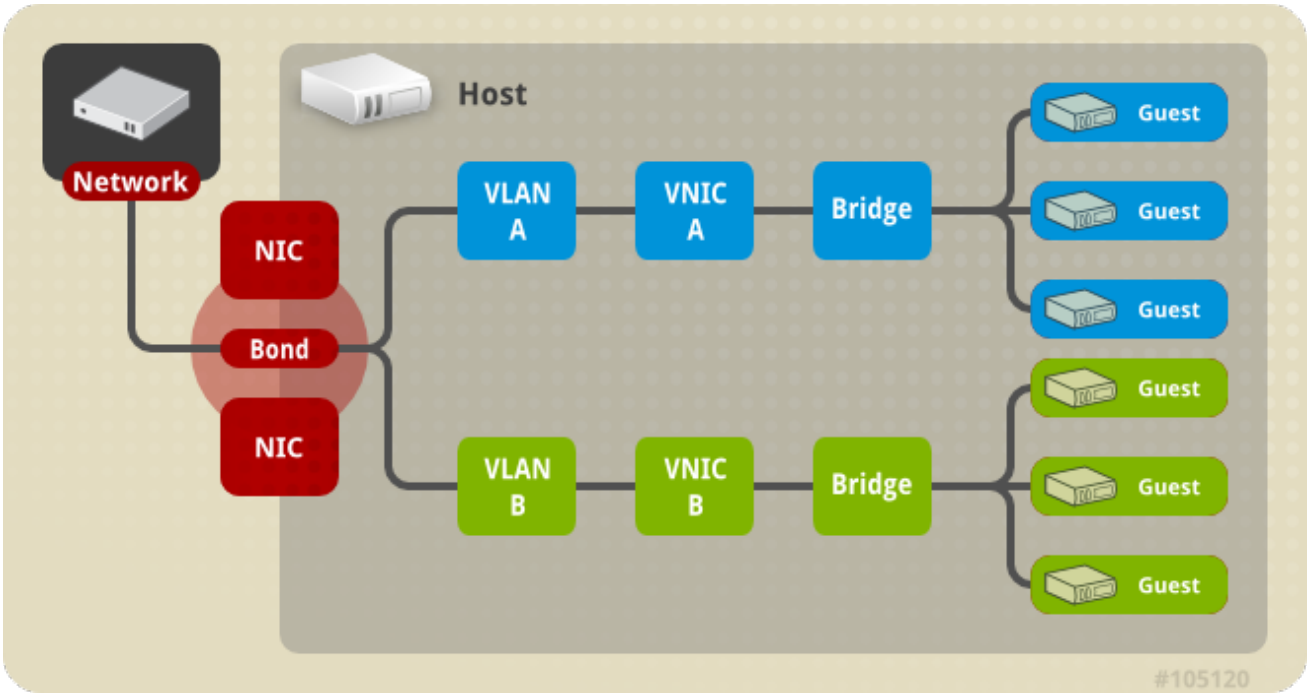
图 3.6. 多个网桥、多个 VLAN 和 NIC 配置



3.20. 多个网桥、多个 VLAN 和绑定配置

图 3.7 “使用 Bond 连接多个网桥、多个 VLAN 和多个 NIC” 显示绑定多个 NIC 的配置，以协助与多个 VLAN 的连接。

图 3.7. 使用 Bond 连接多个网桥、多个 VLAN 和多个 NIC



此配置中的每个 VLAN 通过连接 NIC 的绑定定义。每个 VLAN 连接到单个网桥，每个网桥连接到一个或多个客户机。

第 4 章 电源管理

4.1. 电源管理和隔离简介

当配置了电源管理和隔离时，Red Hat Virtualization 环境最灵活且具有弹性。电源管理允许 Red Hat Virtualization Manager 控制主机电源周期操作，最重要的是重启检测到问题的主机。隔离用于通过重新引导功能将问题主机与 Red Hat Virtualization 环境隔离，以防止性能下降。然后，可以通过管理员操作返回隔离的主机以响应响应状态，并将其整合到环境中。

电源管理和隔离利用特殊的专用硬件来独立于主机操作系统重新启动主机。Red Hat Virtualization Manager 使用网络 IP 地址或主机名连接到电源管理设备。在 Red Hat Virtualization 的上下文中，电源管理设备和隔离设备相同。

4.2. RED HAT VIRTUALIZATION 中的代理管理

Red Hat Virtualization Manager 不直接与隔离代理通信。相反，经理使用代理向主机电源管理设备发送电源管理命令。Manager 使用 VDSM 执行电源管理设备操作，因此环境中的另一台主机用作隔离代理。

您可以选择：

- 任何与需要隔离的主机相同的集群中的主机。
- 任何与需要隔离的主机在同一数据中心中的主机。

可行的隔离代理主机的状态为 UP 或 Maintenance。

4.3. 电源管理

Red Hat Virtualization Manager 能够重新引导进入非运行或不响应状态的主机，并准备关闭利用不足的主机来省电。这个功能取决于正确配置的电源管理设备。Red Hat Virtualization 环境支持以下电源管理设备：

- 美国动力转换 (pc)。
- BladeCenter.
- Cisco Unified Computing System（cisco_ucs).
- Dell Remote Access Card 5（drac5)。
- Dell Remote Access Card 7（drac7)。
- 电子电源交换机(eps)。
- HP BladeSystem（hpblade)。
- Integrated Lights Out（ilo,ilo2,ilo3,ilo4)。
- 智能平台管理接口 (ipmilan)。
- 远程 Supervisor Adapter (rsa)。
- rsb.
- West Telematic, Inc（wti)。

注意

apc 隔离代理不支持 APC 5.x 电源管理设备。使用 **apc_snmp** 隔离代理。

为了与列出的电源管理设备通信，Red Hat Virtualization Manager 使用 隔离代理。Red Hat Virtualization Manager 允许管理员在其环境中为电源管理设备配置隔离代理，其中设备将接受并响应。可以使用图形用户界面配置基本配置选项。也可以输入特殊配置选项，并将未解析的传给隔离设备。特殊配置选项特定于给定的隔离设备，而基本配置选项则用于所有支持的电源管理设备提供的功能。所有电源管理设备提供的基本功能是：

- **状态** ：检查主机的状态。
- **启动** ：打开主机电源。
- **停止** ：关闭主机。
- **重新启动** ：重启主机。实际上将实施为 stop、wait、status、start、wait、status。

最佳实践是在第一次配置时一次测试电源管理配置，有时在这样才能确保继续功能。

通过在环境中所有主机中正确配置电源管理设备，提供弹性。隔离代理允许 Red Hat Virtualization Manager 与主机电源管理设备通信，以绕过问题主机上的操作系统，并通过重新引导主机将主机与环境隔离。然后，管理器可以重新分配 SPM 角色（如果由问题主机保存），并在其他主机上安全地重新启动任何高可用性虚拟机。

4.4. 隔离

在 Red Hat Virtualization 环境的情况下，隔离是由 Manager 使用隔离代理发起的主机重启，并由电源管理设备执行。隔离允许集群响应意外的主机故障，并强制进行节能、负载平衡和虚拟机可用性策略。

隔离可确保存储池管理程序(SPM)的角色始终被分配到功能主机。如果隔离的主机是 SPM，则 SPM 角色将被重新简化并重新分配给响应的主机。由于具有 SPM 角色的主机是唯一能够编写数据域结构元数据的主机，因此不响应的 un-fenced SPM 主机会导致其环境丢失创建和销毁虚拟磁盘的能力，执行快照、扩展逻辑卷以及需要更改数据域结构元数据的其他操作。

当主机变得不响应时，当前在该主机上运行的虚拟机的所有虚拟机也可以变得不响应。但是，不响应的主机会在虚拟机硬盘映像上保留其正在运行的虚拟机的锁定。尝试在第二个主机上启动虚拟机并为虚拟机硬盘镜像分配第二主机写入特权可能会导致数据崩溃。

隔离允许 Red Hat Virtualization Manager 假设虚拟机硬盘镜像上的锁定已被释放；管理器可以使用隔离代理确认问题主机已重启。收到此确认后，Red Hat Virtualization Manager 可以从另一主机上的问题主机上启动虚拟机，而不影响数据崩溃。隔离是高可用性虚拟机的基础。标记为高可用性的虚拟机在没有这样做的情况下无法在备用主机上安全地启动，这会导致数据崩溃。

当主机不响应时，Red Hat Virtualization Manager 允许在执行任何操作前传递 30 (30)秒，以允许主机从任何临时错误中恢复。如果主机在宽限期通过的时间没有响应，则管理器会自动开始缓解来自不响应的主机的任何负面影响。Manager 使用主机上的电源管理卡的隔离代理来停止主机，确认它已停止，启动主机，并确认主机已经启动。当主机完成引导后，它会尝试重新加入集群，该集群是被隔离前的一部分。如果导致主机变得不响应的问题已被重启解决，则主机会自动设置为 **Up** 状态，并且重新能够启动和托管虚拟机。

4.5. 软隔离主机

由于意外问题，主机有时可能会变得不响应，但 VDSM 无法响应请求，但依赖于 VDSM 的虚拟机仍保持有效并可访问。在这些情况下，重启 VDSM 会将 VDSM 返回到响应的状态，并解决这个问题。

"SSH 软隔离"是一个进程，管理器尝试在不响应的主机上通过 SSH 重新启动 VDSM。如果管理器无法通过 SSH 重新启动 VDSM，则隔离的责任将在配置了外部隔离代理时进入外部隔离代理。

通过 SSH 进行软隔离的工作方式如下。必须在主机上配置和启用隔离，并且数据中心必须存在有效的代理主机（第二个主机，处于 UP 状态）。当 Manager 和主机间的连接超时时，会出现以下情况：

1. 在第一个网络失败时，主机的状态将变为"连接"。
2. 然后，管理器会尝试询问 VDSM 的状态，或者等待由主机上的负载决定的时间间隔。确定间隔长度的公式由配置值 TimeoutToResetVdsInSeconds （默认为 60 秒）+ [DelayResetPerVmInSeconds （默认为 0.5 秒）]*（如果主机上运行虚拟机的数量） + [DelayResetForSpmInSeconds （默认为 20 秒）]1（如果主机运行为 SPM）或 0 （如果主机上运行为 SPM）。为了给 VDSM 给予响应的最大时间，经理可选择上述两个选项的更长时间（三个尝试检索 VDSM 的状态或以上公式决定的间隔）。
3. 如果该主机没有响应，则通过 SSH 执行 **vdsm 重启**。
4. 如果 **vdsm 重启** 在主机和管理器之间重新建立连接时没有成功，主机的状态将变为 **Non Responsive**，如果配置了电源管理，则隔离将移交给外部隔离代理。

注意

可以在没有配置电源管理的主机上执行 soft-fencing。这与 "fencing": 不同的是只能在配置了电源管理的主机上执行隔离。

4.6. 使用多个电源管理隔离代理

单个代理被视为主代理。当存在两个隔离代理时，二级代理有效，例如，每个电源交换机都有两个代理连接到同一电源交换机时。代理可以是相同或不同的类型。

在主机上有多个隔离代理会增加隔离过程的可靠性。例如，当主机上的唯一隔离代理出现故障时，主机将保持非运行状态，直到手动重启为止。以前在主机上运行的虚拟机将被暂停，且仅在手动隔离原始主机后切换到集群中的另一主机。有多个代理时，如果第一个代理失败，可以调用下一个代理。

当主机上定义了两个隔离代理时，可以将它们配置为使用 并发 或 顺序 流：

- **concurrent** ：主代理和次要代理都必须响应 Stop 命令，以便主机停止。如果一个代理响应 Start 命令，则主机将启动。
- **顺序** ：要停止或启动主机，首先使用主代理，如果失败，则使用二级代理。

第 5 章 负载均衡、调度和迁移

5.1. 负载均衡、调度和迁移

单个主机具有有限的硬件资源，且容易出现故障。为了减少故障和资源耗尽的问题，主机分组为集群，这基本上是共享资源的分组。Red Hat Virtualization 环境使用负载平衡策略、调度和迁移响应主机资源需求的变化。Manager 无法确保集群中的任何单一主机都负责该集群中的所有虚拟机。相反，管理器能够识别使用率不足的主机，并将所有虚拟机从其中迁移，从而让管理员关闭该主机以省电。

可用资源会因为三个事件的结果进行检查：

- 虚拟机启动 - 检查资源以确定虚拟机在哪个主机上启动。
- 虚拟机迁移 - 检查资源以确定适当的目标主机。
- time elapses - 定期检查资源，以确定单个主机负载是否与集群负载平衡策略相符。

Manager 通过使用集群的负载平衡策略来响应可用资源的更改，以将虚拟机从集群中的一个主机调度到另一个主机。以下部分将讨论负载平衡策略、调度和虚拟机迁移之间的关系。

5.2. 负载均衡策略

为集群设置负载平衡策略，其中包括一个或多个可能具有不同的硬件参数和可用内存的主机。Red Hat Virtualization Manager 使用负载平衡策略来决定集群中的哪些主机启动虚拟机。负载平衡策略还允许 Manager 决定何时将虚拟机从过度使用的主机移动到使用不足的主机。

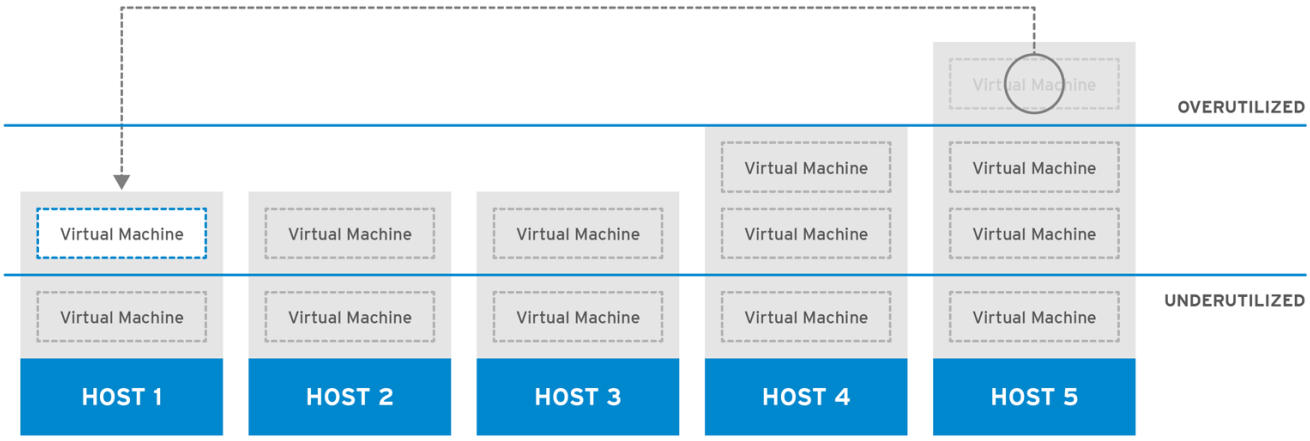
负载平衡过程每分钟为数据中心中每个集群运行一次。它决定了哪些主机被过度使用，它们是使用不足的主机，以及虚拟机迁移的有效目标。确定取决于管理员为给定集群设置的负载平衡策略。负载平衡策略的选项有 **VM_Evenly_Distributed**、**Evenly_Distributed**、**Power_Saving** 和 **None**。

5.3. 负载均衡策略：VM_EVENLY_DISTRIBUTED

虚拟机均匀分布式负载平衡策略根据虚拟机数量在主机之间均匀分配虚拟机。高虚拟机数是每个主机上可以运行的虚拟机的最大数量，这还达到了超量化主机过载。VM_Evenly_Distributed 策略允许管理员为主机设置较高的虚拟机数。管理员还可设置最高利用的主机和最低利用率主机之间虚拟机数量的最大差值。当集群中的每个主机都有不属于此迁移阈值的虚拟机数时，集群处于平衡状态。管理员还设置要在 SPM 主机上保留的虚拟机的插槽数。SPM 主机的负载比其他主机更低的负载，因此此变量定义其可以运行的其他主机的虚拟机数量要少。如果任何主机运行的虚拟机数量超过高虚拟机数，并且至少有一个主机具有超出迁移阈值的虚拟机数，则虚拟机将逐个迁移到集群中具有最低 CPU 使用率的主机。一个虚拟机一次迁移，直到集群中的每个主机都有不属于迁移阈值的虚拟机计数。

5.4. 负载均衡策略：EVENLY_DISTRIBUTED

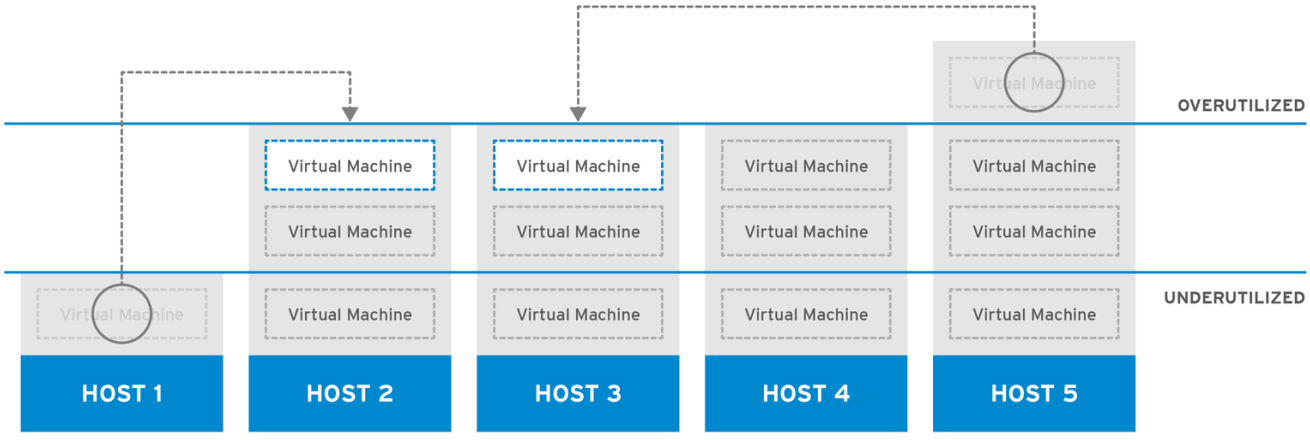
图 5.1. 平均分布式调度策略



平均分布式负载平衡策略根据最低 CPU 负载或可用内存选择新虚拟机的主机。集群中主机允许的最大 CPU 负载和最小可用内存由平均分布式调度策略的参数定义。除这些限制外，环境的性能会降低。平均分布式策略允许管理员为运行中的虚拟机设置这些级别。如果主机达到定义的最大 CPU 负载或最小可用内存，并且主机保留了多个设定时间，则该主机上的虚拟机会逐个迁移到集群中具有最低 CPU 或最高内存的主机，具体取决于使用的参数。主机资源会每分钟检查一次，一次迁移了一个虚拟机，直到主机 CPU 负载低于定义的限制或主机可用内存超过定义的限制。

5.5. 负载均衡策略：POWER_SAVING

图 5.2. 节能调度策略



节能负载平衡策略根据最低 CPU 或可用内存选择新虚拟机的主机。集群中主机允许的最大 CPU 负载和最小可用内存由节能调度策略的参数定义。除这些限制外，环境的性能会降低。节能参数还定义了集群中主机允许的最小 CPU 负载和最大可用内存，在继续操作主机之前，主机持续使用效率较低。如果主机达到最大 CPU 负载或最小可用内存，并且保留了多个设定时间，则该主机上的虚拟机会逐个迁移到具有最低 CPU 或可用内存的主机，具体取决于使用了哪个参数。主机资源会每分钟检查一次，一次迁移了一个虚拟机，直到主机 CPU 负载低于定义的限制或主机可用内存超过定义的限制。如果主机的 CPU 负载低于定义的最小级别，或者主机的可用内存超过定义的最大级别，则该主机中的虚拟机将迁移到集群中的其他主机，只要集群中的其他主机仍低于集群的最大 CPU 负载并低于最小可用内存。当被利用不足的主机清除其剩余的虚拟机时，管理器将自动关闭主机，并在负载平衡需要或集群中没有足够的可用主机时再次重新启动。

5.6. 负载平衡策略：无

如果没有选择负载平衡策略，则在具有最低 CPU 使用率和可用内存的集群中的主机上启动虚拟机。要确定组合指标的 CPU 利用率，其考虑虚拟 CPU 计数和 CPU 用量百分比。此方法是最动态的，因为唯一主机选择点是新虚拟机启动时。虚拟机不会自动迁移，以反映主机上增加的需求。

管理员必须决定哪个主机是给定虚拟机的相应迁移目标。虚拟机也可以通过 固定 来与特定主机关联。固定会阻止虚拟机自动迁移到其他主机。对于大量消耗资源的环境，手动迁移是最佳方法。

5.7. 负载均衡策略：INCLUSTERUPGRADE

InClusterUpgrade 调度策略根据主机操作系统版本分发虚拟机。具有比当前运行的虚拟机较新操作系统的主机的优先级高于具有相同操作系统的主机。迁移到具有较新操作系统的主机的虚拟机不会迁移到较旧的操作系统。虚拟机可以在集群中的任何主机上重启。策略允许集群具有混合操作系统版本来升级集群中的主机。在启用策略前必须满足先决条件。请参阅 Red Hat Enterprise Virtualization 3.6 升级指南中的将集群中的主机从 Red Hat Enterprise Linux 6 升级到 Red Hat Enterprise Linux 7。

重要

InClusterUpgrade 调度策略仅用于在主要版本间进行升级。例如，从 Red Hat Enterprise Linux 6 升级到 Red Hat Enterprise Linux 7。

5.8. 高可用性虚拟机保留

高可用性(HA)虚拟机保留策略可让 Red Hat Virtualization Manager 监控高可用性虚拟机的集群容量。Manager 具有为高可用性标记单个虚拟机的功能，这意味着在主机出现故障时，这些虚拟机将在备用主机上重新启动。此策略会在集群中的主机之间平衡高可用性虚拟机。如果集群中的任何主机失败，则剩余的主机可以在不影响集群性能的情况下支持迁移高可用性虚拟机负载。启用高可用性虚拟机保留时，Manager 确保集群中存在适当的容量，以便 HA 虚拟机在其现有主机意外失败时迁移。

5.9. 调度

在 Red Hat Virtualization 中，调度指的是 Red Hat Virtualization Manager 中选择主机作为新或迁移虚拟机的目标的方式。

要使主机有资格启动虚拟机或接受来自另一主机的迁移虚拟机，它必须具有足够的可用内存和 CPU 来支持正在启动的虚拟机或迁移到它的要求。虚拟机将不会在 CPU 过载的主机上启动。默认情况下，如果主机的 CPU 的负载超过 80% 达到 5 分钟，则主机 CPU 被视为过载，但这些值可以使用调度策略来更改。如果多个主机符合条件的目标，将根据集群的负载平衡策略选择一个。例如，如果 Evenly_Distributed 策略生效，则管理器会选择 CPU 使用率最低的主机。如果 Power_Saving 策略生效，则会选择最大和最小服务级别之间具有最低 CPU 使用率的主机。给定主机的存储池管理器(SPM)状态也会影响作为启动虚拟机或虚拟机迁移的目标。如果 SPM 角色由集群中的主机保存，则非SPM 主机是首选的目标主机，例如，集群中启动的第一个虚拟机不会在 SPM 主机上运行。

5.10. MIGRATION（迁移）

Red Hat Virtualization Manager 使用迁移来强制实施集群的负载平衡策略。虚拟机迁移根据集群的负载平衡策略以及集群中主机的当前需求进行。迁移也可以配置为在主机被隔离或移到维护模式时自动进行。Red Hat Virtualization Manager 首先迁移 CPU 使用率最低的虚拟机。这计算为百分比，除了 I/O 操作影响 CPU 使用率外，不考虑 RAM 使用量或 I/O 操作。如果有多个虚拟机具有相同的 CPU 使用率，则首先迁移的虚拟机是 Red Hat Virtualization Manager 运行的数据库查询返回的第一个虚拟机，以确定虚拟机 CPU 用量。

虚拟机迁移默认有以下限制：

- 每个虚拟机迁移实施 52 MiBps （每秒兆字节）的带宽限制。

- 迁移将在每 GB 虚拟机内存的 64 秒后超时。
- 如果进行停滞了 240 秒，则迁移将中止。
- 并发传出迁移限制为每个主机的每个 CPU 内核有一个，或 2 个（以较小的值）。

有关调整迁移设置的详情，请参阅 https://access.redhat.com/solutions/744423。

第 6 章 目录服务

6.1. 目录服务

Red Hat Virtualization 平台依赖于目录服务进行用户身份验证和授权。与所有管理器接口（包括开发人员门户、管理门户和 REST API）的交互仅限于经过身份验证的授权用户。Red Hat Virtualization 环境中的虚拟机可以使用相同的目录服务来提供身份验证和授权，但它们必须配置为这样做。目前支持与 Red Hat Virtualization Manager 搭配使用的目录服务供应商是 Identity Management (IdM)、Red Hat Directory Server 9 (RHDS)、Active Directory (AD) 和 OpenLDAP。Red Hat Virtualization Manager 与目录服务器的接口：

- 门户登录（用户、Power 用户、管理员、REST API）。
- 查询以显示用户信息。
- 将管理器添加到域中。

身份验证是生成一些数据的方的验证和识别，以及生成的数据的完整性。主体是验证其身份的方。验证器是需要保证主体身份的方。对于 Red Hat Virtualization，Manager 是验证器，用户是主体。数据完整性是保证接收的数据与主体生成的数据相同。

保密性和授权与身份验证紧密相关。保密性可防止数据泄露到不打算接收数据。强大的身份验证方法可以选择提供保密性。授权决定是否允许主体执行操作。Red Hat Virtualization 使用目录服务将用户与角色关联，并相应地提供授权。授权通常在主体通过身份验证后执行，并可能基于对验证器的信息本地或远程。

在安装过程中，本地域会自动配置来管理 Red Hat Virtualization 环境。安装完成后，可以添加更多域。

6.2. 本地身份验证：内部域

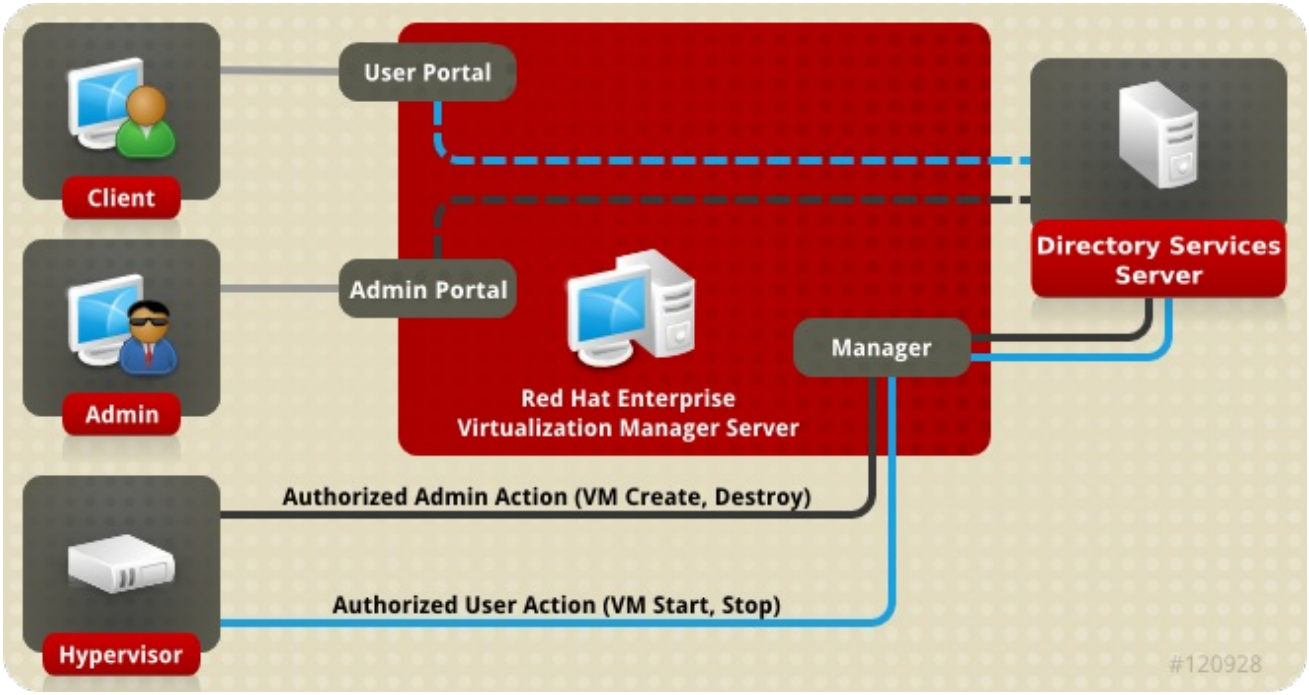
Red Hat Virtualization Manager 在安装过程中创建有限的内部管理域。这个域与 AD 或 IdM 域不同，因为它基于 Red Hat Virtualization PostgreSQL 数据库中的密钥而不是目录服务器中的目录服务用户存在。内部域也与外部域不同，因为内部域只有一个用户：**admin@internal** 用户。通过此方法进行初始身份验证，可以评估 Red Hat Virtualization 而无需一个完整的功能目录服务器，并确保可以使用管理帐户来排除外部目录服务的任何问题。

admin@internal 用户用于环境的初始配置。这包括安装和接受主机，添加外部 AD 或 IdM 身份验证域，以及从外部域向用户分配权限。

6.3. 使用 GSSAPI 进行远程身份验证

在 Red Hat Virtualization 上下文中，远程身份验证是指从 Red Hat Virtualization Manager 远程处理的身份验证。远程身份验证用于来自 AD、IdM 或 RHDS 域中的 Manager 的用户或 API 连接。Red Hat Virtualization Manager 必须通过管理员使用 **engine-manage-domains** 工具作为 RHDS、AD 或 IdM 域的一部分来配置。这要求 Manager 提供了来自 RHDS、AD 或 IdM 目录服务器的帐户凭据，以供有足够权限的域将系统加入域。添加了域后，Red Hat Virtualization Manager 可以使用密码对域用户进行身份验证。管理器使用名为 Simple Authentication and Security Layer (SASL) 的框架，它使用 通用安全服务应用程序接口 (GSSAPI) 安全地验证用户的身份，以及确定用户可用的授权级别。

图 6.1. GSSAPI 身份验证



第 7 章 模板和池

7.1. 模板和池

Red Hat Virtualization 环境为管理员提供了相应的工具，以简化虚拟机调配给用户。这些是 **模板和池** 。模板是一种快捷方式，管理员可以绕过操作系统安装和配置，根据现有的预配置虚拟机快速创建新虚拟机。这对将像设备一样使用的虚拟机（如 Web 服务器虚拟机）特别有用。如果组织使用特定 Web 服务器的多个实例，管理员可以创建用作模板的虚拟机，安装操作系统、Web 服务器、任何支持的软件包，并应用唯一的配置更改。然后，管理员可以基于正常工作的虚拟机创建模板，该虚拟机将用于在需要时创建新的相同虚拟机。

虚拟机池是基于给定模板的虚拟机组，可以快速调配给用户。在池级别授予使用虚拟机的权限；被赋予使用池权限的用户将被从池中分配任何虚拟机。虚拟机池的固有是虚拟机内虚拟机的传输性质。因为用户被分配了虚拟机，而不考虑过去使用的虚拟机，所以池不适用于需要数据持久性的目的。虚拟机池最适合在中央位置存储用户数据，并且虚拟机是一种访问和使用这些数据的方法，或者数据持久性并不重要。创建池会导致创建填充池的虚拟机，处于已停止状态。然后会在用户请求中启动它们。

7.2. 模板

要创建模板，管理员会创建和自定义虚拟机。安装所需软件包，应用自定义配置，虚拟机将针对其预期用途做好准备，以最大程度降低部署后必须对其进行的更改。从虚拟机创建模板前可选但推荐的步骤是常规化。规范化用于删除部署时将更改的系统用户名、密码和时区信息等详细信息。规范化不会影响自定义配置。虚拟机管理指南中的 *模板* 将进一步讨论 Red Hat Virtualization 环境中的 Windows 和 Linux 客户机。Red Hat Enterprise Linux 客户机使用 **sys-unconfig** 一般化。Windows 客户机使用 **sys-prep** 一般化。

当为模板提供基础的虚拟机时，最好地配置模板（如果需要），管理员可以从虚拟机创建模板。从虚拟机创建模板会导致创建特殊配置的虚拟磁盘镜像的只读副本。只读镜像将构成基于该模板的所有创建虚拟机的后备镜像。换句话说，模板基本上是一个带有相关虚拟硬件配置的自定义的只读磁盘映像。可以在从模板创建的虚拟机中更改硬件，例如，从具有 1GB RAM 的模板创建的虚拟机调配两个 GB RAM。但是，模板磁盘镜像无法更改，因为这样做会导致基于模板的所有虚拟机的更改。

创建模板后，它可以用作多个虚拟机的基础。虚拟机使用 **精简置备** 方法或 **克隆** 置备方法从给定模板创建。从模板克隆的虚拟机会占用模板基础镜像的完整可写副本，从而牺牲精简创建方法在交换过程中节省的空间，具体取决于模板的存在。使用 thin 方法从模板创建的虚拟机使用作为基础镜像的模板的只读镜像，这需要模板及其创建的所有虚拟机都存储在同一存储域中。对数据和新生成的数据的更改存储在写入镜像的副本中。基于模板的每个虚拟机都使用相同的基础读取镜像，以及对虚拟机独有的写镜像的副本。这通过限制保存相同数据的次数来节省存储。此外，频繁使用只读后备镜像可能会导致数据被缓存，从而提高了性能。

7.3. 池

虚拟机池允许对作为桌面的用户身份快速调配大量相同的虚拟机。被赋予访问和使用池中的虚拟机权限的用户根据其在请求队列中的位置接收可用虚拟机。池中的虚拟机不允许数据持久性；每次从池中分配虚拟机时，它都会在其基本状态下分配。非常适合在用户数据集中存储的情况下使用。

虚拟机池从模板创建。池中的每个虚拟机都使用相同的后备只读镜像，并使用写镜像的临时副本来保存更改和新生成的数据。池中的虚拟机与其他虚拟机不同，在写入层上保存用户生成和更改的数据的副本会在关闭时丢失。这一点的含义是，虚拟机池不需要比支持它的模板更多的存储，再加上一些使用时生成或更改的数据的空间。虚拟机池是一种高效的方法，可为用户提供一些任务的计算能力，而无需为每个用户提供一个专用虚拟桌面的存储成本。

> ***例 7.1. 池用法示例***

> *技术支持公司采用 10 个帮助台员工。但是，在任何给定时间，只有五个工作。可以创建 5 个虚拟机池，而不是为每个帮助台员工创建 10 个虚拟机。帮助台员工在转换开始时为自己分配一台虚拟机，并在结尾将其返回给池。*

第 8 章 虚拟机快照

8.1. 快照

快照是存储功能，允许管理员在某一时间点创建虚拟机的操作系统、应用程序和数据的恢复点。快照将当前存在于虚拟机硬盘镜像中的数据保存为 COW 卷，并允许在拍摄快照时恢复到数据。快照会导致在当前层上创建新的 COW 层。拍摄快照后执行的所有写入操作都会被写入新的 COW 层。

务必要了解虚拟机硬盘映像是一个或多个卷的链。从虚拟机的角度来看，这些卷显示为单个磁盘镜像。虚拟机会模糊处理其磁盘由多个卷组成的事实。

术语 COW 卷和 COW 层可互换使用，但层更清晰地识别快照的时序性质。每个快照都会创建，以便管理员在拍摄快照 后 丢弃对数据所做的不满意更改。快照提供与许多单词处理器中存在的 **Undo** 功能类似的功能。

注意

不支持标记为 **共享** 的虚拟机硬盘的快照，以及基于 **直接 LUN** 连接的用户，否则支持。

三个主要快照操作是：

- 创建，这涉及为虚拟机创建的第一个快照。
- 预览（包括预览快照）来确定是否将系统数据恢复到拍摄快照的时间点。
- 删除，这涉及删除不再需要的恢复点。

有关快照操作的任务信息，请参阅 Red Hat Virtualization 虚拟机管理指南中的 快照。

8.2. RED HAT VIRTUALIZATION 中的实时快照

不支持标记为 **共享** 的虚拟机硬盘的快照，以及基于 **直接 LUN** 连接的用户，否则支持。

任何未克隆或迁移的其他虚拟机都可以在运行、暂停或停止时执行快照。

当启动虚拟机的实时快照时，Manager 会请求 SPM 主机为要使用的虚拟机创建新卷。当新卷就绪时，管理器使用 VDSM 与运行虚拟机的虚拟机上的 libvirt 和 qemu 通信，以启动虚拟机写入操作。如果虚拟机能够写入新卷，则快照操作将被视为成功，并且虚拟机停止写入上一个卷。如果虚拟机无法写入新卷，则快照操作将被视为失败，并且新卷会被删除。

当启动实时快照后，虚拟机需要同时访问其当前卷和新卷，直到新卷就绪后，两个卷都会以读写权限打开。

具有已安装客户机代理的虚拟机支持静止功能可以确保快照间的文件系统一致性。注册的 Red Hat Enterprise Linux 客户机可以安装 **qemu-guest-agent**，以便在快照前启用静默。

如果在拍摄快照时在虚拟机上存在静默兼容的客户机代理，VDSM 使用 libvirt 与代理通信，以准备快照。完成未完成的写入操作，然后在拍摄快照前冻结文件系统。当快照完成后，libvirt 已将虚拟机切换到新卷以进行磁盘写入操作，文件系统会被解封，写入磁盘恢复。

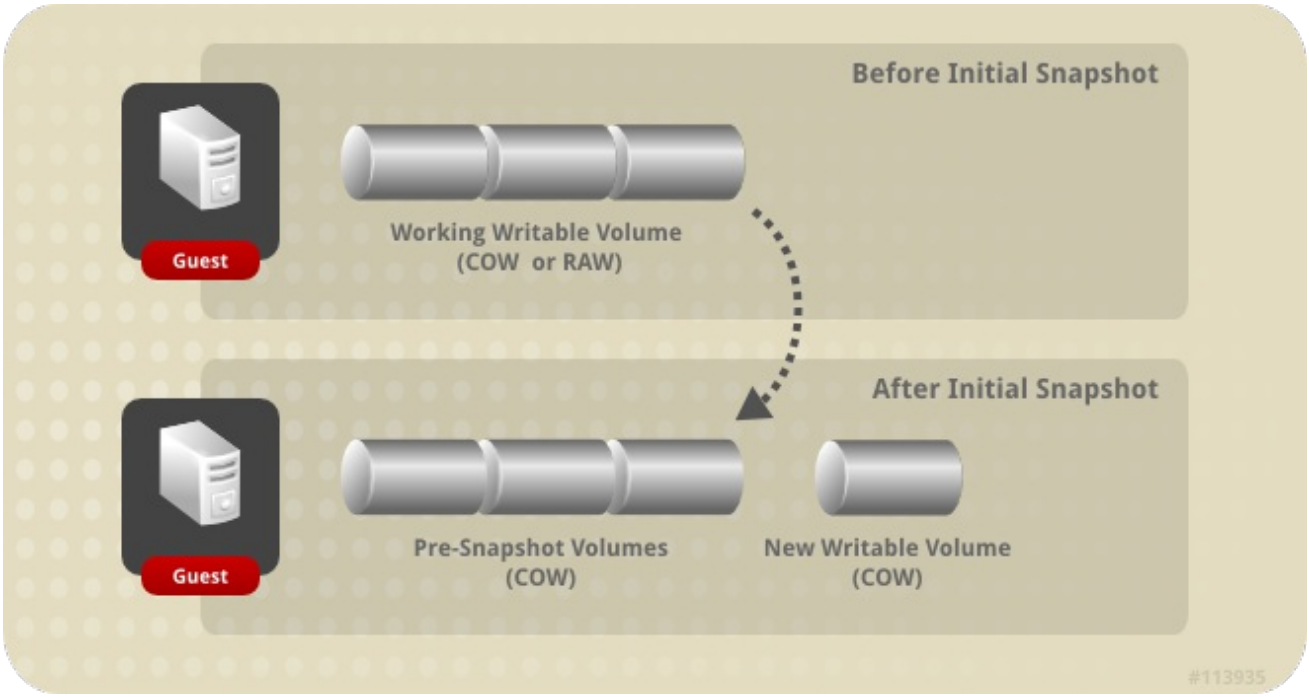
启用静止后尝试的所有实时快照。如果 snapshot 命令因为没有兼容的客户机代理失败，则会在不使用 use-quiescing 标志的情况下重新启动实时快照。当虚拟机恢复到带有静默文件系统的预快照状态时，它会完全引导，且不需要文件系统检查。使用未静止的文件系统恢复之前的快照需要在启动时进行文件系统检查。

8.3. 快照创建

在 Red Hat Virtualization 中，虚拟机的初始快照与初始快照保留其格式(QCOW2 或 RAW)不同。虚拟机的第一个快照将现有卷指定为基础镜像。其他快照是额外的 COW 层，跟踪自上一次快照以来对镜像中存储的数据所做的更改。

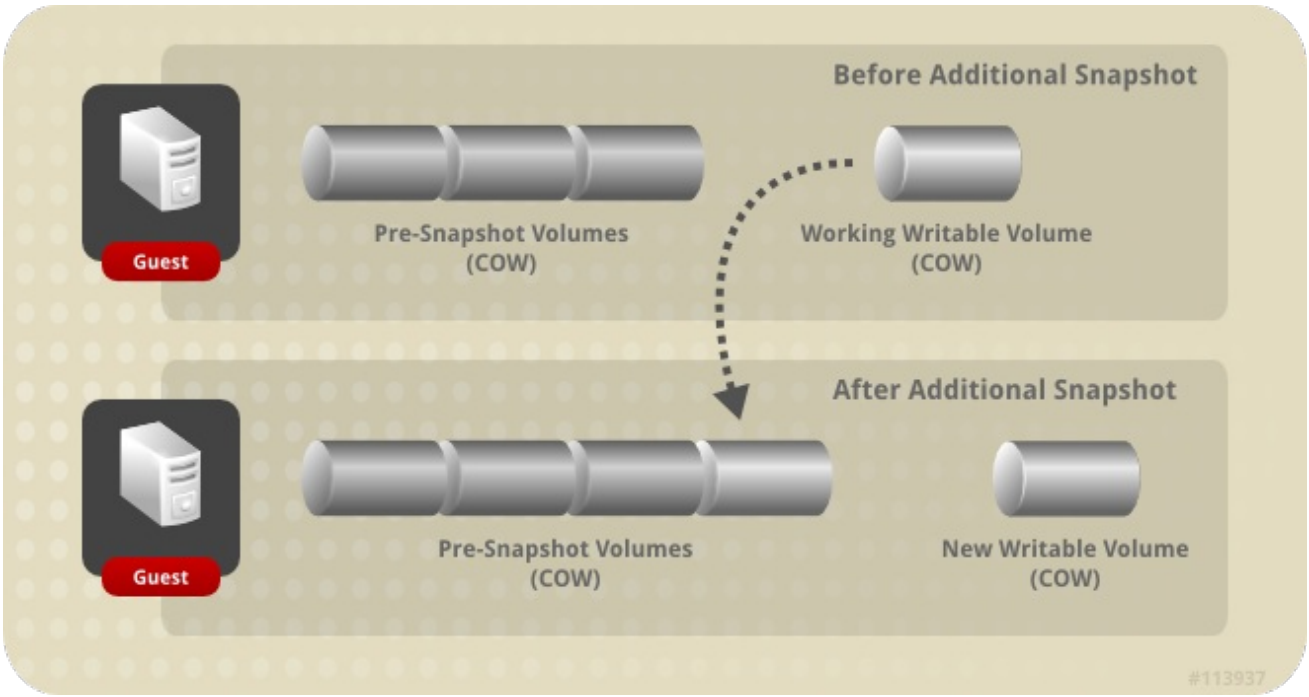
在 Red Hat Virtualization 中，客户虚拟机通常与 RAW 磁盘镜像交互，除非镜像创建为精简置备的镜像或用户特别要求是 QCOW2。如 图 8.1 "初始快照创建" 所述，快照的创建会导致组成虚拟磁盘镜像的卷充当所有后续快照的基础镜像。

图 8.1. 初始快照创建



在初始快照后执行快照会导致创建新的 COW 卷，其中创建快照后创建或更改的数据。每个新的 COW 层都只包含 COW 元数据。在快照写入新的 COW 层后，通过虚拟机使用和操作创建的数据。当使用虚拟机修改上一 COW 层中存在的数据时，数据会从上一层读取，并写入最新的层。虚拟机通过检查每个 COW 层从最新到最旧的虚拟机查找数据。

图 8.2. 其他快照创建



8.4. 快照预览

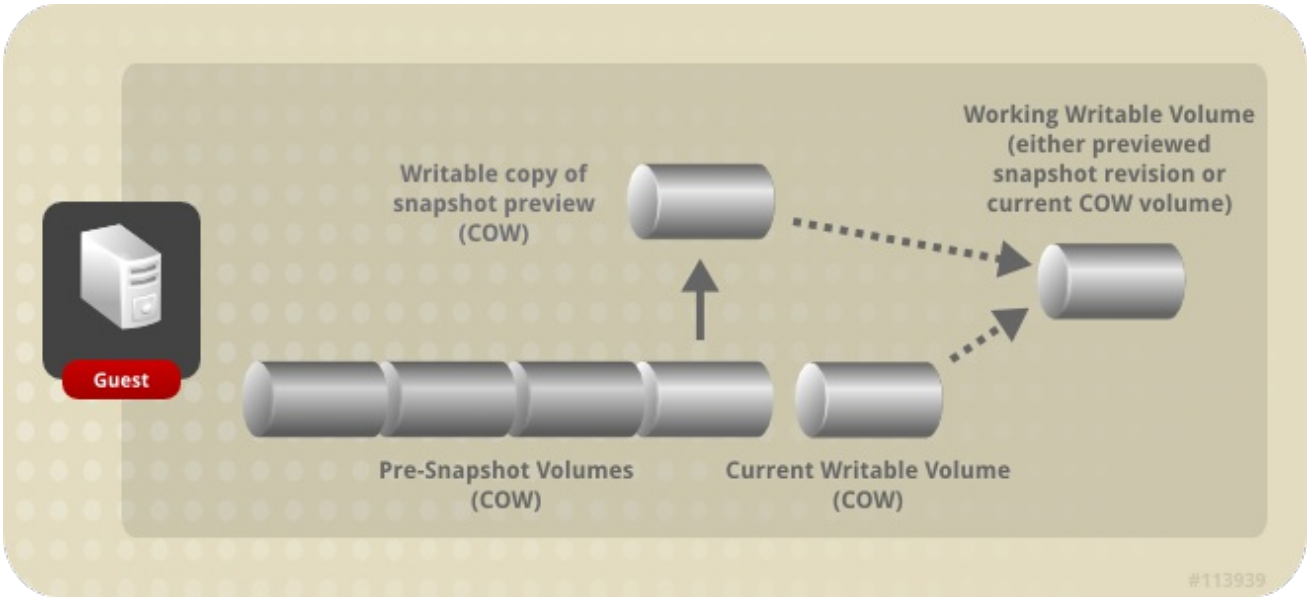
若要选择将虚拟磁盘映像恢复到哪个快照，管理员可以预览之前创建的所有快照。

从每个客户机的可用快照中，管理员可以选择快照卷以预览其内容。如 图 8.3 "预览快照" 所述，每个快照都保存为 COW 卷，当它被预览时，会从快照中复制一个新的预览层。客户机与预览进行交互，而不是实际的快照卷。

在管理员预览所选快照后，可以提交预览将客户机数据恢复到快照中捕获的状态。如果管理员提交预览，客户机将附加到预览层。

在快照被预览后，管理员可以选择 Undo 来丢弃所查看快照的预览层。尽管丢弃预览层，但保留包含快照本身的层。

图 8.3. 预览快照



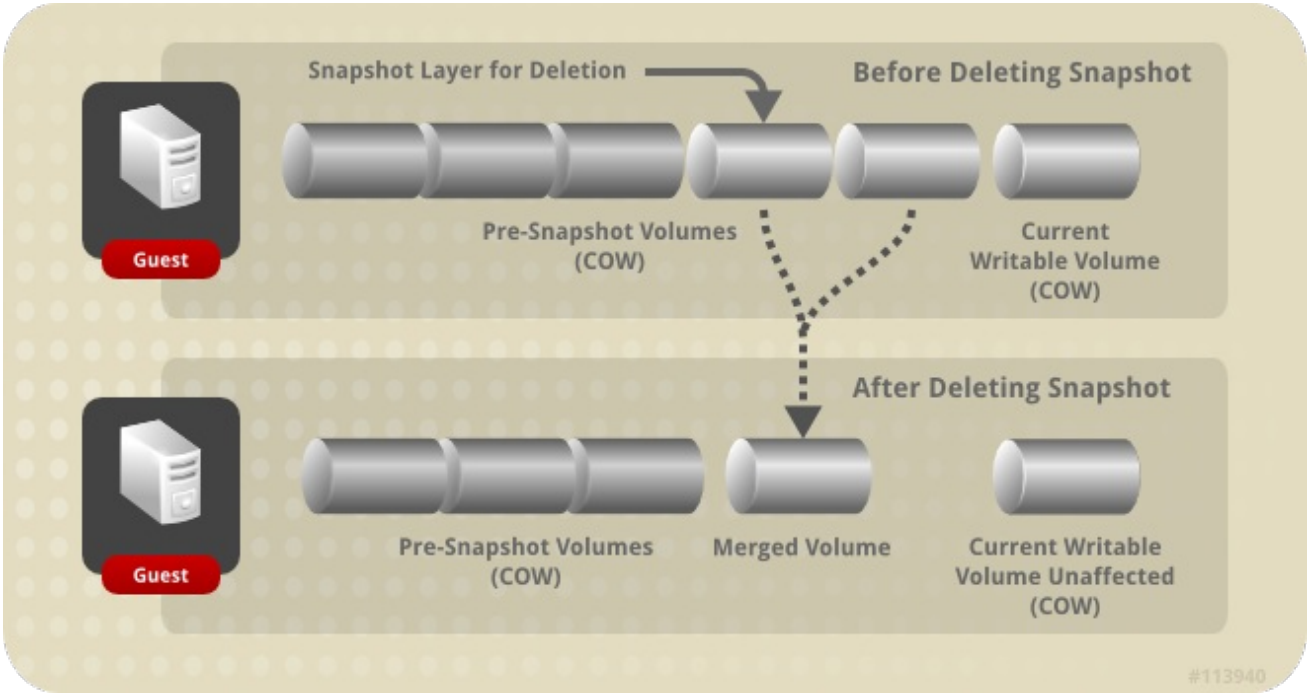
8.5. 快照删除

您可以删除个别快照或一系列不再需要的快照。删除快照会删除将虚拟磁盘镜像恢复到该特定恢复点的能力。它不一定会回收快照消耗的磁盘空间，也不会删除数据。只有在后续快照覆盖了已删除快照的数据时，才会回收磁盘空间。例如，如果第五个快照中的第三个快照被删除，则第三个快照中的更改数据必须保留在磁盘上供第四和第五个快照使用；但是，如果第四或第五个快照覆盖了第三个快照，那么第三个快照已冗余，并且可以回收磁盘空间。除了潜在的磁盘空间回收之外，删除快照也可以提高虚拟机的性能。

当选择快照删除时，QEMU 会创建一个相同大小的新逻辑卷，以将要删除的快照与后续快照合并。这个新逻辑卷会调整大小，以适应两个快照之间的所有区别。新逻辑卷可能是两个快照的总组合大小。合并了两个快照后，后续的快照会被重命名并标记为删除，并被新逻辑卷替换，它取其名称。快照最初标记为删除，其后续快照将被删除，其位置是单个合并的快照。

例如，快照 **Delete_snapshot** 是 200 GB，后续的快照 **Next_snapshot** 为 100 GB。**Delete_snapshot** 被删除，并创建了新逻辑卷，临时创建名为 **Snapshot_merge**，大小为 200 GB。**Snapshot_merge** 最终将调整为 300 GB，以适应 **Delete_snapshot** 和 **Next_snapshot** 的总合并内容。**Next_snapshot** 被重命名为 **Delete_me_too_snapshot**，以便可以重命名为 **Next_snapshot**。**最后，删除 Delete_snapshot 和 Delete_me_too_snapshot 被删除。**

图 8.4. 快照删除



用于从正在运行的虚拟机中删除快照的逻辑与已关闭的虚拟机略有不同。实时**快照删除作**为**异步块作业处理，其中 VDSM 在虚拟机的恢复文件中维护操作**记录，**以便可以跟踪作业，即使 VDSM 重新**启动，**或者虚拟机在操作过程中关闭。操作开始后，无法预览删除的快照或用作恢复点，即使操作失败或中断也是如此。在活跃层要与其父级合并的操作中，操作被分成两阶段，期间数据从活跃层复制到父层，磁盘写入会被镜像到活动的层和父层。最后，一旦删除快照中的数据与其父快照合并，VDSM 会被**视为已完成，VDSM 会在镜像链中同步更改。

第 9 章 硬件驱动程序和设备

9.1. 虚拟化硬件

Red Hat Virtualization 为虚拟客户机提供三种不同类型的系统设备。这些硬件设备都显示为物理附加的硬件设备到虚拟客户机，但设备驱动程序以不同的方式工作。

模拟设备

模拟设备（有时被称为 虚拟设备 ）完全存在于软件中。模拟设备驱动程序 是在主机上运行的操作系统（管理源设备）和客户机上运行的操作系统之间的翻译层。定向到模拟设备的设备级别指令会被拦截并被虚拟机监控程序转换。任何与 Linux 内核模拟和识别相同的设备都能够用作模拟驱动程序的后备源设备。

半虚拟化设备

半虚拟化设备需要在客户机操作系统上安装设备驱动程序，为其提供接口来与主机上的管理程序通信。此界面用于允许传统密集型任务（如磁盘 I/O）在虚拟化环境之外执行。以这种方式降低虚拟化固有的开销是，在直接在物理硬件上运行时，客户端操作系统性能可以接近预期。

物理共享设备

某些硬件平台允许虚拟客户机直接访问各种硬件设备和组件。虚拟化中的此过程称为 passthrough 或 设备分配。透传(passthrough)可让设备显示并的行为就像它们实际附加到客户端操作系统一样。

9.2. RED HAT VIRTUALIZATION 中的稳定设备地址

虚拟硬件 PCI 地址分配保留在 ovirt-engine 数据库中。

PCI 地址由 QEMU 在创建虚拟机时分配，并由 libvirt 报告给 VDSM。VDSM 将其报告回 Manager，它们存储在 ovirt-engine 数据库中。

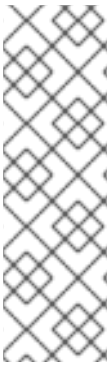
当虚拟机启动时，管理器会从数据库发送 VDSM 设备地址。VDSM 将它们传递给 libvirt，这使用虚拟机首次运行时分配的 PCI 设备地址启动虚拟机。

当从虚拟机中删除设备时，也会移除对该设备的所有引用（包括稳定的 PCI 地址）。如果添加设备来替换移除的设备，它将由 QEMU 分配，该地址不太可能与它替换的设备相同。

9.3. 中央处理单元(CPU)

集群中的每个主机都有多个 虚拟 CPU (vCPU)。虚拟 CPU 公开给主机上运行的客户端。当集群最初通过 Red Hat Virtualization Manager 创建时，由集群中的主机公开的所有虚拟 CPU 都是所选类型。无法在集群中混合虚拟 CPU 类型。

每种可用的虚拟 CPU 类型都具有相同名称的物理 CPU 的特征。虚拟 CPU 独立于物理 CPU 到客户端操作系统。



备注

支持 x2APIC ：

Red Hat Enterprise Linux 7 主机提供的所有虚拟 CPU 型号都包含对 x2APIC 的支持。这提供了 高级可编程中断控制器(APIC) 来更好地处理硬件中断。

9.4. 系统设备

系统设备对于要运行的客户机来说至关重要，且无法删除。附加到客户机的每个系统设备也会占用可用的 PCI 插槽。默认系统设备是：

- 主机网桥，
- ISA 网桥和 USB 网桥(USB 和 ISA 网桥是相同的设备)
- 图形卡（使用 Cirrus 或 qxl 驱动程序）以及
- 内存气球设备。

9.5. 网络设备

Red Hat Virtualization 能够向客户机公开三种不同类型的网络接口控制器。在创建客户机时，选择要公开给客户机的网络接口控制器类型，但可从 Red Hat Virtualization Manager 更改。

- e1000 网络接口控制器向客户机公开虚拟化 Intel PRO/1000 (e1000)。
- virtio 网络接口控制器向客户机公开半虚拟化网络设备。
- rtl8139 网络接口控制器向客户机公开一个虚拟化 Realtek Semiconductor Corp RTL8139。

每个客户机都允许多个网络接口控制器。添加的每个控制器都需要客户机上有一个可用的 PCI 插槽。

9.6. 图形设备

提供了两个模拟图形设备。这些设备可以与 SPICE 协议或 VNC 连接到。

- ac97 模拟 Cirrus CLGD 5446 PCI VGA 卡。
- vga 使用 BochsVESA 扩展（硬件级别，包括所有非标准模式）模拟 dummy VGA 卡。

9.7. 存储设备

存储设备和存储池可以使用块设备驱动程序将存储设备附加到虚拟客户机。请注意，存储驱动程序不是存储设备。驱动程序用于将后备存储设备、文件或存储池卷附加到虚拟客户机中。后备存储设备可以是任何受支持的存储设备、文件或存储池卷的类型。

- IDE 驱动程序向客户机公开仿真块设备。模拟的 IDE 驱动程序可用于将最多四个虚拟化 IDE 硬盘或虚拟化 IDE CD-ROM 驱动器的任意组合附加到每个虚拟客户机。模拟的 IDE 驱动程序也用于提供虚拟化 DVD-ROM 驱动器。
- VirtIO 驱动程序向客户机公开半虚拟化块设备。半虚拟化块驱动程序是附加到虚拟化客户机的管理程序支持的所有存储设备的驱动程序（但软盘磁盘驱动器除外，必须模拟）。

9.8. 声音设备

有两个模拟声音设备可用：

- ac97 模拟 Intel 82801AA AC97 Audio 兼容声卡。
- es1370 模拟 ENSONIQ AudioPCI ES1370 声卡。

9.9. 串行驱动程序

半虚拟化串行驱动程序(virtio-serial)是一个字节型字符流驱动程序。半虚拟化串行驱动程序提供主机用户空间和客户机用户空间（网络不可用或不可用）之间的简单通信接口。

9.10. BALLOON DRIVER

balloon 驱动程序允许客户机表达其需要的 hypervisor 量。balloon 驱动程序允许主机高效地为客户机分配和内存，并允许将可用内存分配给其他客户机和进程。

使用 balloon 驱动程序的客户机可以将客户机 RAM 的部分标记为不使用（膨胀）。hypervisor 可以释放内存，并将内存用于其他主机进程或该主机上的其他虚拟机。当客户机再次需要可用内存时，虚拟机监控程序可以重新分配 RAM 到客户机（膨胀）。

第 10 章 最低要求和技术限制

10.1. 最低要求和支持的限制

Red Hat Virtualization 环境有多个物理和逻辑限制。目前不支持具有这些限制之外的配置环境。

10.2. 资源限值

某些限制适用于存储域和主机等资源。

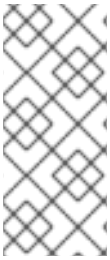
表 10.1. 资源限值

项	限制
存储域	建议每个数据中心至少 2 个存储域： • 需要数据存储域。 • 建议使用 ISO 存储域。
主机	红帽支持每个 Red Hat Virtualization Manager 最多 200 个主机。

10.3. 集群限制

集群是一组物理主机，这些主机被视为一组虚拟机的资源池。集群中的主机共享相同的网络基础架构和相同的存储。集群是一个迁移域，其中的虚拟机可以从主机移到主机。为确保每个集群都有多个限制的稳定性。

- 所有受管虚拟机监控程序都必须位于集群中。
- 集群内的所有受管虚拟机监控程序都必须具有相同的 CPU 类型。Intel 和 AMD CPU 无法在同一集群中共存。



备注

有关集群的更多信息，请参阅管理指南中的 集群。https://access.redhat.com/documentation/en/red-hat-virtualization/4.0/single/administration-guide/#chap-Clusters

10.4. 存储域限制

存储域为虚拟磁盘镜像和 ISO 镜像存储提供空间，以及虚拟机的导入和导出。虽然可以在给定数据中心中创建许多存储域，但每个存储域都有一些限制和建议。

表 10.2. 存储域限制

项	限制
存储类型	支持的存储类型有： • 光纤通道协议 (FCP) • Internet Small Computer System Interface (iSCSI) • 网络文件系统 (NFS) • POSIX Compliant 文件系统(POSIX) • Red Hat Gluster Storage (GlusterFS) Red Hat Virtualization 4.0 中的新 ISO 和导出存储域可由任何基于文件的存储(NFS、Posix 或 GlusterFS)提供。
逻辑单元号(LUN)	每个由 iSCSI 或 FCP 提供的存储域都不允许超过 300 个 LUN。
逻辑卷(LV)	在 Red Hat Virtualization 中，逻辑卷代表虚拟机、模板和虚拟机快照的虚拟磁盘。 对于由 iSCSI 或 FCP 提供的每个存储域，我们不推荐使用超过 350 个逻辑卷。如果给定存储域中的逻辑卷数量超过这个数字，则建议将可用存储拆分为单独的存储域，且每个逻辑卷都不超过 350 个逻辑卷。 这个限制的根本原因是 LVM 元数据的大小。随着逻辑卷的数量增加，与这些逻辑卷关联的 LVM 元数据也会增加。当此元数据大小超过 1 MB 时，置备操作（如创建新磁盘或快照）的性能会减少，在运行 QCOW 磁盘时，用于精简配置逻辑卷的 lvextend 操作需要更长的时间才能运行。 有关逻辑卷的详情请参考 https://access.redhat.com/solutions/441203。

备注

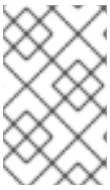
有关存储域的更多信息，请参阅管理指南中的 存储。

10.5. RED HAT VIRTUALIZATION MANAGER LIMITATIONS

Red Hat Virtualization Manager servers must run Red Hat Enterprise Linux 7.另外，还需要满足很多额外的硬件要求。

表 10.3. Red Hat Virtualization Manager Limitations

项	限制
RAM	• 至少 4 GB RAM。
PCI 设备	• 至少有一个网络控制器，最小带宽为 1 Gbps。
存储	• 建议至少具有 25 GB 的可用磁盘空间。



备注

有关 Red Hat Virtualization Manager 的更多信息，请参阅 安装指南。

10.6. HYPERVISOR 要求

Red Hat Virtualization Host (RHVH)有多个硬件要求和支持的限制。Red Hat Enterprise Linux 主机的存储要求根据其现有配置所使用的磁盘空间量而有所不同，但应该大于 RHVH。

表 10.4. Red Hat Virtualization 主机要求和支持的限制

项	支持限制

项	支持限制
CPU	至少 1 个物理 CPU。Red Hat Virtualization 支持在主机中使用这些 CPU 型号： • AMD Opteron G1 • AMD Opteron G2 • AMD Opteron G3 • AMD Opteron G4 • AMD Opteron G5 • Intel Conroe • Intel Penryn • Intel Nehalem • Intel Westmere • Intel Haswell • Intel SandyBridge Family • IBM POWER 8 所有 CPU 都必须支持 Intel® 64 或 AMD64 CPU 扩展，并且启用了 AMD-V™ 或 Intel VT® 硬件虚拟化扩展。还需要支持 **No eXecute** 标志(NX)。
RAM	每个虚拟机所需的 RAM 量因以下不同： • 客户机操作系统要求， • 客户机应用程序要求，以及 • 虚拟机的内存活动和使用. 此外，KVM 能够为虚拟机过量使用物理 RAM。它仅根据需要为虚拟机分配 RAM，并将利用率不足的虚拟机转换为交换。 有关最大和最低支持 RAM，请参见 https://access.redhat.com/articles/rhel-limits。

项	支持限制
存储	主机支持的最小内部存储是以下列表总数： • root (/)分区至少需要 6 GB 存储。 • /boot 分区至少需要 1 GB 存储。 • /var 分区至少需要 15 GB 存储。对于自托管引擎部署，这必须至少为 60 GB。 • 交换分区需要至少 8 MB 的存储。建议的 swap 分区的大小因主机被安装的系统以及环境预期的过量使用级别而有所不同。如需更多信息 https://access.redhat.com/solutions/15244 ，请参阅。 请注意，它们是主机安装的 *最低存储要求*。建议您使用使用更多存储空间的默认分配。
PCI 设备	至少需要一个网络控制器，建议的最小带宽为 1 Gbps。

重要

当 Red Hat Virtualization Host 引导一个信息时，可能会出现信息：

```
Virtualization hardware is unavailable.
(No virtualization hardware was detected on this system)
```

这个警告表示虚拟化扩展被禁用，或者您的处理器中没有虚拟化扩展。确保 CPU 支持列出的扩展，并在系统 BIOS 中启用。

检查处理器是否有虚拟化扩展，并启用了它们：

- **在主机引导屏幕中，按任意键，然后从列表中选择 Boot 或 Boot with serial console 条目。按 Tab 编辑所选选项的内核参数。列出的最后一个内核参数后，确保 有一个空格 并附加 rescue 参数。**
- **按 Enter 键引导进入救援模式。**
- **在出现提示时，确定您的处理器具有虚拟化扩展，并通过运行以下命令启用它们：**

  ```
  # grep -E 'svm|vmx' /proc/cpuinfo
  ```

  如果显示任何输出，处理器将支持硬件虚拟化。如果没有显示输出，您的处理器仍然可以支持硬件虚拟化。在某些情况下，制造商在 BIOS 中禁用虚拟化扩展。你认为这是这种情况，请咨询系统的 BIOS 以及制造商提供的主板手册。
- **作为额外的检查，验证 kvm 模块是否已在内核中载入：**

  ```
  # lsmod | grep kvm
  ```

  如果输出包含 kvm_intel 或 kvm_amd，则 kvm 硬件虚拟化模块已加载，您的系统满足要求。

10.7. 客户机要求和支持限制

以下要求和支持限制适用于在 Red Hat Virtualization 主机(RHVH)上运行的客户机：

表 10.5. 虚拟化硬件

项	限制
CPU	Red Hat Enterprise Linux 7 每个客户机最多支持 240 个虚拟化 CPU。
RAM	不同的客户机有不同的 RAM 要求。每个客户机所需的 RAM 量因客户机操作系统的要求以及客户机正在操作的负载而有所不同。 有关 https://access.redhat.com/articles/rhel-kvm-limits 客户机机器的最大和最低支持 RAM，请参阅。
PCI 设备	每个客户机最多支持 31 个虚拟化 PCI 设备。许多系统设备对这个限制进行计数，其中一些是强制的。针对 PCI 设备限制的强制设备包括 PCI 主机网桥、ISA 网桥、USB 网桥、板板网桥、图形卡以及 IDE 或 VirtIO 块设备。
存储	每个客户机最多支持 28 个虚拟化存储设备，其中包括可能的 3 IDE 和 25 Virtio。

10.8. SPICE 限制

SPICE 目前支持最大分辨率 2560x1600 像素。

10.9. 其他参考资源

这些附加文档资源不作为 Red Hat Virtualization 文档套件的一部分。但是，它们确实包含管理 Red Hat Virtualization 环境的系统管理员的有用信息，并可通过 获得 https://access.redhat.com/documentation/en/red-hat-enterprise-linux/。

Red Hat Enterprise Linux - System Administrator's Guide

Red Hat Enterprise Linux 的部署、配置和管理指南。

Red Hat Enterprise Linux - DM-Multipath Guide

在 Red Hat Enterprise Linux 中使用 Device-Mapper 多路径指南。

Red Hat Enterprise Linux - Installation Guide

安装 Red Hat Enterprise Linux 的指南。

Red Hat Enterprise Linux - Storage Administration Guide

在 Red Hat Enterprise Linux 中管理存储设备和文件系统的指南。

Red Hat Enterprise Linux - 虚拟化部署和管理指南

Red Hat Enterprise Linux 中虚拟化技术的安装、配置、管理和故障排除指南。