了解使用 QEMU 和 Libvirt 实现 KVM 虚拟化

NetApp virtualization solutions

NetApp
August 18, 2025

This PDF was generated from https://docs.netapp.com/zh-cn/netapp-solutions-virtualization/kvm/kvm-overview.html on August 18, 2025. Always check docs.netapp.com for the latest.

目录

了解使用 QEMU 和 Libvirt 实现 KVM 虚拟化 1
组件概述 1
集群管理 2
计算 2
存储 3
网络 4
监控 5
数据保护 6

了解使用 QEMU 和 Libvirt 实现 KVM 虚拟化

了解 KVM 虚拟化以及它如何与 QEMU 和 Libvirt 协同工作以创建全面的 Linux 虚拟化平台。了解这些组件如何支持接近本机性能的虚拟机，同时为企业环境中的计算、存储和网络资源提供灵活的管理功能。

组件概述

1. KVM（基于内核的虚拟机）：
   - 作用：KVM是一个内核模块，通过利用CPU中的硬件辅助虚拟化扩展（例如Intel VT-x或AMD-V）来提供核心虚拟化功能。
   - 功能：KVM 允许 Linux 内核充当 1 型虚拟机管理程序，使其能够创建和管理具有接近本机 CPU 和内存操作性能的虚拟机。
   - 最低层：KVM 是最低级别的组件，直接与硬件交互，为完全虚拟化提供必要的基础设施。
2. QEMU（快速模拟器）：
   - 作用：QEMU是一个机器（硬件）模拟器，为客户系统提供其他虚拟化硬件组件，例如网络接口、磁盘控制器和显卡。
   - 功能：当 KVM 不可用时，QEMU 可以作为 2 型虚拟机管理程序运行，但由于它必须在软件中模拟客户 CPU，因此速度会慢得多。 QEMU 与 KVM 结合使用时，使用 KVM 加速 CPU 和内存操作，而 QEMU 处理外围设备和其他硬件设备的仿真。
   - 与 KVM 的交互：QEMU 通过设备文件（例如 /dev/kvm）使用 ioctl() 系统调用与 KVM 交互，以管理虚拟机进程并与 KVM 通信。
3. "Libvirt"：
   - 作用：Libvirt是一个虚拟化库和API，为各种虚拟化平台提供更高级别的管理接口，包括KVM/QEMU、Xen、VMware ESXi等。
   - 功能：Libvirt 通过提供统一的界面和工具集简化了虚拟机的管理。它充当抽象层，允许用户和应用程序管理虚拟机，而无需直接与底层虚拟机管理程序的特定命令或 API 交互。
   - 主要特点：
     - VM 生命周期管理：启动、停止、暂停、保存、恢复和迁移 VM。
     - 远程管理：通过 SSH 或其他协议控制远程主机上的虚拟机。
     - 存储管理：为虚拟机创建和管理存储池和卷。
     - 虚拟网络：配置NAT、桥接等模式的虚拟网络。
     - 安全性：与 SELinux 和 AppArmor 集成，以实现虚拟机的安全限制。
     - 热插拔：在虚拟机运行时添加或删除磁盘和网络接口等设备。
     - 工具：Libvirt 包括用于管理虚拟机的命令行工具（如 virsh）和图形工具（如 virt-manager）。
   - 它们如何协同工作：
     - KVM：提供内核级虚拟化基础架构。
     - QEMU：提供模拟硬件并管理VM进程。

        - Libvirt：充当管理层，提供 API 和工具来控制 KVM/QEMU 和其他虚拟机管理程序。
    - 本质上：KVM 为虚拟化提供硬件加速，QEMU 提供模拟硬件并运行 VM，Libvirt 提供用户友好的管理界面和 API 来控制整个设置。

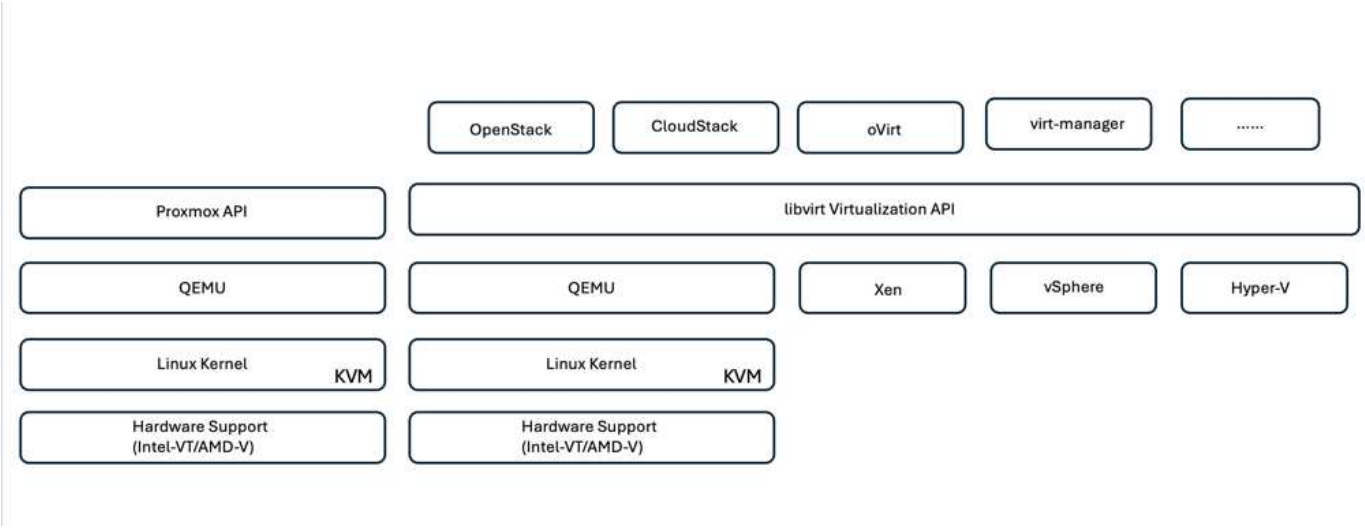
Libvirt 客户端工具可用于通过驾驶舱机器从 CLI、GUI 或基于 Web 管理虚拟机或操作。要查看使用 libvirt 的应用程序列表，请查看 "此处"。

当应用程序进入 Kubernetes 环境时，请查看 Kubevirt 以在这些环境中将虚拟机作为 pod 运行。

"Ansible 模块可用于 libvirt"用于自动化目的。

集群管理

通常，虚拟化管理器或 virsh cli 工具一次管理单个主机。为了管理集群中的多个主机，通常使用更高级别的应用程序，如 oVirt、CloudStack 或 OpenStack。这些工具将协助放置虚拟机并分配负载。如果在小型集群环境中寻找某些虚拟机的高可用性，则可以使用 Pacemaker 和 Corosync 或检查您的管理堆栈选项。



计算

Libvirt 提供了一个用于管理虚拟机资源和功能的综合框架。这包括与以下相关的任务：

1. 虚拟机（域）管理：
    - 生命周期操作：Libvirt 提供了一整套用于管理虚拟机状态的操作（在 libvirt 术语中称为“域”）。这包括启动、停止、暂停、恢复、保存、还原和迁移虚拟机。
    - XML 配置：虚拟机配置使用 XML 文件定义。您可以使用 virsh 或 virt-manager 等工具来创建、修改和删除这些 XML 配置。
    - 远程管理：您可以使用 libvirt 的远程协议管理远程主机上的虚拟机，该协议支持 SSH 等各种网络传输。
2. 资源分配和管理：
    - CPU 管理：Libvirt 允许您配置客户机 CPU，包括指定虚拟 CPU 的数量、控制 CPU 固定（将 vCPU 与主机上的特定物理 CPU 关联）以及管理 CPU 模式（如主机直通以向客户机公开主机的 CPU 功能）。
    - 内存管理：您可以为虚拟机分配内存并配置内存过量使用（允许分配给虚拟机的内存总和超过主机上可用的物理内存）。

- 存储管理：Libvirt 可以管理虚拟机的各种类型的存储，包括磁盘映像（格式包括 qcow2、vmdk 和 raw ）、NFS 共享、LVM 卷组、iSCSI 共享和原始磁盘设备。
- 主机设备管理：您可以管理物理和虚拟主机设备，如 USB、PCI、SCSI 和网络设备，包括 SR-IOV 和 NPIV 等虚拟化功能。

3. 虚拟网络：
   - 虚拟网络交换机：Libvirt 创建虚拟网络交换机（网桥）以将虚拟机相互连接并连接到主机网络。
   - 网络模式：它支持各种网络模式，例如 NAT、桥接、隔离和路由，以配置虚拟机与网络的交互方式。
   - 防火墙规则：Libvirt 自动管理防火墙规则（使用 iptables）来控制虚拟网络的网络流量。
4. 性能优化：
   - CPU 固定：将 vCPU 固定到特定的物理 CPU 可以提高缓存效率和性能，尤其是在 NUMA 环境中。
   - NUMA 调整：您可以通过将客户机大小限制为单个 NUMA 节点上的资源量并将 vCPU 和内存固定到连接到 I/O 适配器的同一物理插槽来优化 NUMA 系统的性能。
   - 大页面：使用大页面可以减少与管理小内存页面相关的开销，从而提高性能。
5. 与其他工具集成：
   - virsh：与 libvirt 交互的命令行界面。
   - virt-manager：用于管理虚拟机和 libvirt 资源的图形工具。
   - OpenStack：Libvirt 是 OpenStack 中常用的虚拟化驱动程序。
   - 第三方工具：许多其他工具和应用程序利用 libvirt 的 API 来管理虚拟机，包括云管理平台和备份解决方案。

KVM 虚拟机管理程序允许过度使用 CPU 和内存，因为通常 VM 客户端的利用率较低。但需要进行监控和平衡以获得更好的性能。

VM 元数据以 XML 形式存储在 /etc/libvirt/qemu 中。可以使用 virt-install 或 virsh cli 创建 VM。如果优先使用 UI 或使用上层管理堆栈，则可以使用 Virt-Manager。

总之，libvirt 为虚拟化的计算方面提供了一个全面的管理层，允许您控制虚拟机生命周期、分配资源、配置网络、优化性能以及与其他工具和平台集成。

存储

VM 磁盘可以在存储池上动态配置，也可以由存储管理员为 VM 预先配置。 libvirt 支持多种池类型。以下是适用的池类型以及支持的存储协议的列表。常用的选择是 dir。然后是 netfs 和 logical。iscsi 和 iscsi-direct 使用单目标，不提供容错功能。mpath 提供多路径，但不支持动态分配。它的使用方式更像是 vSphere 中的原始设备映射。对于文件协议（NFS/SMB/CIFS），可以在 "自动安装程序"或使用 fstab 和目录池类型。对于块协议（iSCSI、FC、NVMe-oF），使用 ocfs2 或 gfs2 等共享文件系统。

存储协议	目录	文件系统	净流表	逻辑	磁盘	**iscsi**	**iscsi直接**	**mpath**
SMB/CIFS	是	否	是	否	否	否	否	否
NFS	是	否	是	否	否	否	否	否
iSCSI	是	是	否	是	是	是	是	是
FC	是	是	否	是	是	否	否	是

存储协议	目录	文件系统	净流表	逻辑	磁盘	**iscsi**	**iscsi直接**	**mpath**
NVMe-oF	是	是	否	是	是	否	否	否[1]

注意： 1 - 可能需要额外的配置。

根据所使用的存储协议，主机上需要提供额外的包。这是示例列表。

存储协议	**Fedora**	**Debian**	吃豆人
SMB/CIFS	samba 客户端/cifs-utils	smbclient/cifs实用程序	smbclient/cifs实用程序
NFS	nfs实用程序	nfs-通用	nfs实用程序
iSCSI	iscsi 启动器实用程序、设备映射器多路径、ocfs2 工具/gfs2 实用程序	open-iscsi、多路径工具、ocfs2 工具/gfs2 实用程序	open-iscsi、多路径工具、ocfs2 工具/gfs2 实用程序
FC	sysfsutils、设备映射器多路径、ocfs2 工具/gfs2 实用程序	sysfsutils、多路径工具、ocfs2 工具/gfs2 实用程序	sysfsutils、多路径工具、ocfs2 工具/gfs2 实用程序
NVMe-oF	nvme-cli、ocfs2-工具/gfs2-utils	nvme-cli、ocfs2-工具/gfs2-utils	nvme-cli、ocfs2-工具/gfs2-utils

存储池详细信息存储在 /etc/libvirt/storage 的 XML 文件中。

要从 vSphere 环境导入 VM 数据，请查看"Shift 工具包"。

网络

Libvirt 为管理虚拟机和容器提供了强大的虚拟网络功能。它通过虚拟网络交换机或网桥的概念实现这一点。

核心概念：* 虚拟网络交换机（网桥）：这就像主机服务器上基于软件的网络交换机。虚拟机连接到此交换机，并且流量通过它流动。 * TAP 设备：这些是特殊的网络设备，充当连接虚拟机网络接口和 libvirt 桥的“虚拟电缆”。

- 网络模式：Libvirt 支持多种网络配置以满足不同的需求：
  - NAT（网络地址转换）：这是默认模式。连接到NAT网络的虚拟机可以通过主机的IP地址访问外部网络，但外部主机无法直接向虚拟机发起连接。
  - Bridged：此模式下虚拟网络直接与主机连接在同一个网段。这使得虚拟机看起来就像直接连接到物理网络一样。
  - 隔离：隔离网络上的虚拟机可以相互通信并与主机通信，但无法访问主机外部的任何内容。这对于测试或安全环境很有用。
  - 路由：来自虚拟网络的流量无需 NAT 即可路由到物理网络。这需要在主机网络上进行正确的路由配置。
  - 开放：类似于路由模式，但没有 libvirt 自动应用任何防火墙规则。这假设网络流量将由其他系统管理。
- DHCP 和 DNS：Libvirt 可以使用 dnsmasq 管理其虚拟网络的 DHCP 服务，从而允许它为虚拟机分配 IP 地址并处理虚拟网络内的 DNS 解析。
- 防火墙规则：Libvirt 自动设置 iptables 规则来控制虚拟网络的流量，尤其是在 NAT 模式下。

管理 Libvirt 网络：

- virsh：virsh 命令行工具提供了一套全面的用于管理虚拟网络的命令，包括列出、启动、停止、定义和取消定义网络。
- 虚拟机管理器（virt-manager）：此图形工具通过直观的用户界面简化了虚拟网络的创建和管理。
- XML 配置：Libvirt 使用 XML 文件来定义虚拟网络的配置。您可以直接编辑这些 XML 文件或使用 virsh net-edit 等工具来修改网络配置。

常见用例：

- NAT：为具有单个网络接口的主机上的虚拟机提供简单的基本连接。
- 桥接：将虚拟机无缝集成到现有网络中。
- 隔离：创建安全或测试环境，限制虚拟机的外部访问。
- 路由：需要特定路由的更高级场景。
- Open vSwitch (OVS)：适用于需要高级网络管理和自动化的复杂、大规模部署。

通过利用这些功能，libvirt 提供了一个灵活而强大的框架来管理 Linux 环境中的虚拟机网络。

监控

NetApp Data Infrastructure Insights （以前称为Cloud Insights）是一个基于云的基础架构监控和分析平台，可为您的 IT 基础架构（包括虚拟机）提供全面的可视性。

虽然Data Infrastructure Insights以高度关注监控NetApp存储和 VMware 环境而闻名，但它也具有监控其他类型的基础设施和工作负载的能力。

您可以使用NetApp Data Infrastructure Insights监控基于 Libvirt 的虚拟机：

1. 数据收集器：
    - Data Infrastructure Insights通过采集单元软件运行，该软件使用各种数据收集器从您的基础设施收集数据。
    - Data Infrastructure Insights具有异构基础设施和工作负载的收集器，包括 Kubernetes。还有一个开放的 Telegraf 收集器和开放的 API，可轻松与其他系统集成。
2. 与 Libvirt 的潜在集成：
    - 自定义数据收集：您可以使用开放的 Telegraf 收集器或Data Infrastructure InsightsAPI 从基于 Libvirt 的系统收集数据。您需要编写或配置收集器以使用其 API 从 Libvirt 收集指标（例如，通过 virsh 命令或访问 Libvirt 的内部指标）。
3. 使用Data Infrastructure Insights监控 Libvirt 的好处：
    - 统一可视性：获得虚拟化环境的单一视图，包括NetApp存储和基于 Libvirt 的虚拟机。
    - 性能监控：识别性能瓶颈和资源限制，无论它们是虚拟机内部的还是与支持它们的底层基础设施相关的。
    - 资源优化：分析工作负载概况以调整虚拟机大小、回收未使用的资源并优化整个环境中的资源利用率。
    - 故障排除：通过将虚拟机性能指标与后端存储指标关联起来，实现端到端可视性，快速识别和解决问题。

- 预测分析：使用机器学习获得智能洞察，并在潜在问题影响性能之前主动识别它们。

总而言之，虽然Data Infrastructure Insights对 VMware 有强大的支持，但可以通过使用自定义数据收集器或利用其开放 API 将其与基于 Libvirt 的虚拟化集成。这将为Data Infrastructure Insights平台内的 Libvirt 环境提供统一的可见性、增强的性能监控和资源优化功能。

数据保护

可以通过多种方法使用NetApp ONTAP保护基于 Libvirt 的虚拟机的数据，通常利用 ONTAP 的内置数据保护功能。以下是常见策略的细分：

1. 使用 ONTAP 的原生数据保护功能：
   - 快照：ONTAP 的核心数据保护技术是快照。这些是数据卷的快速、时间点副本，需要的磁盘空间最少，并且性能开销可以忽略不计。您可以使用快照来创建 Libvirt VM 磁盘的频繁备份（假设它们存储在ONTAP卷上）。
   - SnapMirror： SnapMirror用于将 Snapshot 副本从一个ONTAP存储系统异步复制到另一个 ONTAP 存储系统。这使您可以在远程站点或云中创建 Libvirt VM 的灾难恢复 (DR) 副本。
   - SnapVault： SnapVault用于将数据从多个存储系统备份到中央ONTAP系统。这是将来自不同主机的许多 Libvirt VM 的备份整合到中央备份存储库的一个不错的选择。
   - SnapRestore： SnapRestore允许您从 Snapshot 副本快速恢复数据。这对于在数据丢失或损坏时恢复 Libvirt VM 至关重要。
   - FlexClone： FlexClone根据 Snapshot 副本创建卷的可写副本。这对于根据生产虚拟机数据快速创建测试/开发环境很有用。
   - MetroCluster/ SnapMirror主动同步：为了实现自动零 RPO（恢复点目标）和站点到站点可用性，您可以使用ONTAP MetroCluster或 SMas，这使得能够在站点之间建立延伸集群。
2. 与第三方备份解决方案集成：许多第三方备份解决方案与NetApp ONTAP集成并支持备份虚拟机。您可以使用这些解决方案将 Libvirt VM 备份到ONTAP存储，利用 ONTAP 的数据保护功能。例如，一些备份解决方案使用 ONTAP 的 Snapshot 技术实现快速、无代理备份。
3. 脚本和自动化：您可以创建脚本来自动化创建 Libvirt VM 卷的ONTAP快照的过程。这些脚本可以利用 ONTAP 的命令行界面或 API 与存储系统进行交互。

主要考虑因素：

- 存储位置：您的 Libvirt VM 磁盘映像应存储在ONTAP卷上，以利用 ONTAP 的数据保护功能。
- 网络连接：确保 Libvirt 主机和ONTAP存储系统之间的网络连接。
- HBA 管理：如果使用光纤通道 (FC) 进行存储连接，请确保在 Libvirt 主机上安装了必要的 HBA 管理包。
- 监控和报告：监控您的数据保护操作并确保它们成功完成。通过将 Libvirt 的功能与 ONTAP 强大的数据保护功能相结合，您可以为虚拟化环境实施全面的数据保护策略。

版权信息

版权所有 © 2025 NetApp, Inc.。保留所有权利。中国印刷。未经版权所有者事先书面许可，本文档中受版权保护的任何部分不得以任何形式或通过任何手段（图片、电子或机械方式，包括影印、录音、录像或存储在电子检索系统中）进行复制。

从受版权保护的 NetApp 资料派生的软件受以下许可和免责声明的约束：

本软件由 NetApp 按“原样”提供，不含任何明示或暗示担保，包括但不限于适销性以及针对特定用途的适用性的隐含担保，特此声明不承担任何责任。在任何情况下，对于因使用本软件而以任何方式造成的任何直接性、间接性、偶然性、特殊性、惩罚性或后果性损失（包括但不限于购买替代商品或服务；使用、数据或利润方面的损失；或者业务中断），无论原因如何以及基于何种责任理论，无论出于合同、严格责任或侵权行为（包括疏忽或其他行为），NetApp 均不承担责任，即使已被告知存在上述损失的可能性。

NetApp 保留在不另行通知的情况下随时对本文档所述的任何产品进行更改的权利。除非 NetApp 以书面形式明确同意，否则 NetApp 不承担因使用本文档所述产品而产生的任何责任或义务。使用或购买本产品不表示获得 NetApp 的任何专利权、商标权或任何其他知识产权许可。

本手册中描述的产品可能受一项或多项美国专利、外国专利或正在申请的专利的保护。

有限权利说明：政府使用、复制或公开本文档受 DFARS 252.227-7013（2014 年 2 月）和 FAR 52.227-19（2007 年 12 月）中“技术数据权利 — 非商用”条款第 (b)(3) 条规定的限制条件的约束。

本文档中所含数据与商业产品和/或商业服务（定义见 FAR 2.101）相关，属于 NetApp, Inc. 的专有信息。根据本协议提供的所有 NetApp 技术数据和计算机软件具有商业性质，并完全由私人出资开发。 美国政府对这些数据的使用权具有非排他性、全球性、受限且不可撤销的许可，该许可既不可转让，也不可再许可，但仅限在与交付数据所依据的美国政府合同有关且受合同支持的情况下使用。除本文档规定的情形外，未经 NetApp, Inc. 事先书面批准，不得使用、披露、复制、修改、操作或显示这些数据。美国政府对国防部的授权仅限于 DFARS 的第 252.227-7015(b)（2014 年 2 月）条款中明确的权利。

商标信息

NetApp、NetApp 标识和 http://www.netapp.com/TM 上所列的商标是 NetApp, Inc. 的商标。其他公司和产品名称可能是其各自所有者的商标。