



Red Hat Virtualization 4.4

实施 SR-IOV 的硬件注意事项

使用 Red Hat Virtualization 实施 SR-IOV 的硬件注意事项

Red Hat Virtualization 4.4 实施 SR-IOV 的硬件注意事项

使用 Red Hat Virtualization 实施 SR-IOV 的硬件注意事项

Red Hat Virtualization Documentation Team

Red Hat Customer Content Services

rhev-docs@redhat.com

法律通告

Copyright © 2024 Red Hat, Inc.

The text of and illustrations in this document are licensed by Red Hat under a Creative Commons Attribution–Share Alike 3.0 Unported license ("CC-BY-SA"). An explanation of CC-BY-SA is available at

<http://creativecommons.org/licenses/by-sa/3.0/>

. In accordance with CC-BY-SA, if you distribute this document or an adaptation of it, you must provide the URL for the original version.

Red Hat, as the licensor of this document, waives the right to enforce, and agrees not to assert, Section 4d of CC-BY-SA to the fullest extent permitted by applicable law.

Red Hat, Red Hat Enterprise Linux, the Shadowman logo, the Red Hat logo, JBoss, OpenShift, Fedora, the Infinity logo, and RHCE are trademarks of Red Hat, Inc., registered in the United States and other countries.

Linux[®] is the registered trademark of Linus Torvalds in the United States and other countries.

Java[®] is a registered trademark of Oracle and/or its affiliates.

XFS[®] is a trademark of Silicon Graphics International Corp. or its subsidiaries in the United States and/or other countries.

MySQL[®] is a registered trademark of MySQL AB in the United States, the European Union and other countries.

Node.js[®] is an official trademark of Joyent. Red Hat is not formally related to or endorsed by the official Joyent Node.js open source or commercial project.

The OpenStack[®] Word Mark and OpenStack logo are either registered trademarks/service marks or trademarks/service marks of the OpenStack Foundation, in the United States and other countries and are used with the OpenStack Foundation's permission. We are not affiliated with, endorsed or sponsored by the OpenStack Foundation, or the OpenStack community.

All other trademarks are the property of their respective owners.

摘要

本文档概述了使用 Red Hat Enterprise Linux 实施 SR-IOV 的硬件注意事项，以及 Red Hat Virtualization 的设备分配。

目录

第1章 简介	3
1.1. SR-IOV 的硬件注意事项	3
1.2. 设备分配的硬件注意事项	3
附录 A. 法律通知	5

第 1 章 简介

这是显示如何在 Red Hat Virtualization 上设置和配置 SR-IOV 的一系列主题中的一个。

- 实施 SR-IOV 的硬件注意事项（本文档）
- 为 PCI 透传配置主机（为您的环境选择适当的安装指南）：
 - [使用命令行将 Red Hat Virtualization 安装为自托管引擎](#)
 - [使用本地数据库将 Red Hat Virtualization 安装为独立管理器](#)
 - [使用远程数据库将 Red Hat Virtualization 安装为独立管理器](#)
- [编辑 NIC 上的虚拟功能配置](#)
- [在 vNIC 配置文件中启用 passthrough](#)
- [使具有 SR-IOV 的虚拟机可以迁移](#)

1.1. SR-IOV 的硬件注意事项

单根 I/O 虚拟化(SR-IOV)是一个硬件引用，它允许将单个 PCI Express(PCIe)端点用作多个单独的设备。这通过引入两个 PCIe 功能来实现：物理功能(PF)和虚拟功能(VF)。

物理功能是包括 SR-IOV 功能的传统 PCIe 功能，具有对 PCIe 设备（包括数据移动）的完整配置和管理功能。每个 PCIe 设备可以有一个到 8 个独立的 PF。

虚拟功能是轻量级 PCIe 功能，包含数据移动和最少配置资源集所需的资源。可以在每个 PF 上创建多个 VF，每个 PF 可以支持不同的 VF 数量。允许的 VF 总数依赖于 PCIe 设备厂商，在设备之间是不同的。

PCIe 规格通过实现备选路由 ID 弃用(ARI)实现更大的 VF 数，它可以在 PCIe 标头中重新生成设备编号字段，允许超过 8 个功能。这个转换功能依赖于该设备的 PCIe 设备和端口（无论是根端口还是交换机），支持 ARI。

系统固件（BIOS 或 UEFI）为 PCIe 拓扑分配资源，包括内存、I/O 端口和 PCIe 总线号码范围。因此，SR-IOV 必须被固件支持并启用才能分配足够的资源。

1.1.1. 概述

- 固件（BIOS 或 UEFI）必须支持 SR-IOV。检查默认情况下是否启用扩展。如果没有，请手动启用。这与启用虚拟化扩展（VT-d 或 AMD-Vi）类似。具体详情请参考供应商手册。
- 根端口或立即由 PCIe 设备（如 PCIe 交换机）的上游端口必须支持 ARI。
- PCIe 设备必须支持 SR-IOV。

请参阅供应商规格和数据表，确认硬件是否符合这些要求。

`lspci -v` 命令可用于打印系统上已安装的 PCI 设备的信息。

1.2. 设备分配的硬件注意事项

设备分配提供了将虚拟客户端直接分配给 PCIe 设备的能力，为客户机提供接近原生的性能。在与 SR-IOV 结合使用时，虚拟客户端会直接分配 VF。这样，可以将多个虚拟客户机直接分配给一个 PCIe 设备的 VF。

SR-IOV 不需要启用将虚拟机直接分配给 PCIe 设备，也没有设备分配唯一应用程序以创建 VF，但这两种功能是补充的，并且有额外的硬件注意事项。

设备分配需要在 CPU 和固件中支持 I/O 内存管理单元(IOMMU)。IOMMU 在 I/O 虚拟地址(IOVA)和物理内存地址之间转换。这允许虚拟客户机使用客户机物理地址对设备进行编程，然后由 IOMMU 转换为主机物理地址。

IOMMU 组是可与系统中所有其他设备隔离的一组设备。IOMMU 组代表 IOMMU granularity 的最小设备集合，以及系统中所有其他 IOMMU 组之间的隔离。这允许 IOMMU 区分事务和 IOMMU 组外的设备之间的直接内存访问(DMA)和 IOMMU 控件。

在虚拟客户端和 PCIe 设备的虚拟功能之间隔离事务是设备分配的基础。在 PCIe 和服务器规格中定义的访问控制服务(ACS)功能是维护 IOMMU 组内隔离的硬件标准。如果没有原生 ACS，否则来自硬件厂商的确认，IOMMU 组中的任何多功能都会在 IOMMU 之外的功能之间公开对等的 DMA，将 IOMMU 组扩展为包括缺乏适当隔离的功能。

对于服务器的根端口，还建议对原生 ACS 的支持，否则这些端口上安装的设备将被分组在一起。root 端口、基于处理器(northbridge)根端口和基于控制器 hub 的 root 端口(southbridge)根端口有两个变量关系。如上所述，如果设备分配与 SR-IOV 结合使用，而虚拟客户机被分配给 VF，则这些端口必须同时支持 ACS 和 ARI。

Intel 的 Xeon 处理器 E5 系列、Xeon 处理器 E7 系列和 High End Desktop Processors 包括基于处理器的根端口上原生 ACS 支持。

Intel Platform Controller Hub(PCH)PCI Express 根端口当前不支持 ACS 或使用非标准 ACS 实施，从而对通过这些根端口连接的设备进行精细隔离。其中许多根端口确实支持 ACS 等效功能。Red Hat Enterprise Linux 7.3 内核包括对 X79、X99、X99、5 系列到 9 系列以及 100-series PCI Express 芯片组启用此 ACS 等效功能的支持。

在安装 PCIe 设备时，请参阅厂商规格以确定基于处理器的和基于控制器的 root 端口，以确保 root 端口支持 ACS。

另外，I/O 拓扑中的任何 PCIe 交换机或桥接也需要 ACS 支持，否则它可能会扩展 IOMMU 组。

1.2.1. 概述

- CPU 必须支持 IOMMU（如 VT-d 或 AMD-Vi）。IBM POWER8 默认支持 IOMMU。
- 固件必须支持 IOMMU。
- 使用的 CPU 根端口必须支持 ACS 或 ACS 等效功能。
- PCIe 设备必须支持 ACS 或 ACS 等效功能。
- 建议 PCIe 设备和根端口之间的所有 PCIe 交换机和网桥均应支持 ACS。例如，如果交换机不支持 ACS，该交换机后面的所有设备共享同一个 IOMMU 组，并且只能分配到同一虚拟机。

请参阅供应商规格和数据表，确认硬件是否符合这些要求。

lspci -v 命令可用于打印系统上已安装的 PCI 设备的信息。

附录 A. 法律通知

Copyright © 2022 Red Hat, Inc.

Licensed under the ([Creative Commons Attribution–ShareAlike 4.0 International License](#)).从(oVirt Project)的文档衍生而来。如果您发布本文档或对其进行改编，您必须提供原始版本的 URL。

修改后的版本必须删除所有红帽商标。

Red Hat、Red Hat Enterprise Linux、Red Hat 商标、Shadowman 商标、JBoss、OpenShift、Fedora、Infinity 商标以及 RHCE 都是在美国及其他国家的注册商标。

Linux® 是 Linus Torvalds 在美国和其他国家/地区的注册商标。

Java® 是 Oracle 和/或其附属公司的注册商标。

XFS® 是 Silicon Graphics International Corp. 或其子公司在美国和/或其他国家的商标。

MySQL® 是 MySQL AB 在美国、欧盟和其他国家/地区的注册商标。

Node.js® 是 Joyent 的官方商标。Red Hat Software Collections 与官方 Joyent Node.js 开源或商业项目没有正式关联或被正式认可。

The OpenStack® Word Mark 和 OpenStack 标识是 OpenStack Foundation 在美国及其他国家的注册商标/服务标记或商标/服务标记，可根据 OpenStack Foundation 授权使用。我们不附属于 OpenStack Foundation 或 OpenStack 社区。

所有其他商标均由其各自所有者所有。