Red Hat Virtualization 4.4

SR-IOV を実装するためのハードウェアの考慮事項

Red Hat Virtualization で SR-IOV を実装するためのハードウェアの考慮事項

Last Updated: 2024-02-27

Red Hat Virtualization 4.4 SR-IOV を実装するためのハードウェアの考慮事項

Red Hat Virtualization で SR-IOV を実装するためのハードウェアの考慮事項

Red Hat Virtualization Documentation Team
Red Hat Customer Content Services
rhev-docs@redhat.com

法律上の通知

概要

このドキュメントでは、Red Hat Enterprise Linux で SR-IOV を実装する場合、および Red Hat Virtualization でデバイスを割り当てる場合のハードウェアの考慮事項について概説します。

目次

第1章 はじめに

これは、Red Hat Virtualization で SR-IOV を準備および設定する方法を示す一連のトピックの 1 つです。

- SR-IOV を実装するためのハードウェアの考慮事項 (本ガイド)
- PCI パススルー用のホストの設定 (ご使用の環境に適したインストールガイドを選択してください):
  - コマンドラインを使用したセルフホストエンジンの Red Hat Virtualization のインストール
  - ローカルデータベースが設定されたスタンドアロン Manager の Red Hat Virtualization のインストール
  - リモートデータベースが設定されたスタンドアロン Manager の Red Hat Virtualization のインストール
- NIC の仮想機能設定の編集
- vNIC プロファイルでパススルーの有効化
- SR-IOV を搭載した仮想マシンの移行を可能に

1.1. SR-IOV のハードウェアに関する考慮事項

Single Root I/O Virtualization (SR-IOV) は、単一の PCI Express (PCIe) エンドポイントを複数の個別のデバイスとして使用できるようにするハードウェアリファレンスです。これは、物理機能 (PF) と仮想機能 (VF) の 2 つの PCIe 機能を導入することで実現されます。

物理機能は、SR-IOV 機能を含む従来の PCIe 機能であり、データ移動を含む PCIe デバイスの完全な設定と制御を備えています。各 PCIe デバイスは、1〜8 個の独立した PF を利用できます。

仮想機能は、データ移動に必要なリソースと最小化された設定リソースのセットを含む軽量の PCIe 機能です。各 PF で複数の VF を作成でき、各 PF は異なる量の VF をサポートできます。許可される VF の総数は、PCIe デバイスのベンダーによって異なり、デバイスによって異なります。

PCIe 仕様は、8 つ以上の機能を可能にする PCIe ヘッダーのデバイス番号フィールドを再解釈する Alternative Routing ID Interpretation (ARI) の実装を通じて、より多くの VF をサポートします。この変換は、PCIe デバイスと、デバイスのすぐ上流にあるポート (ルートポートまたはスイッチ) の両方に依存し、ARI をサポートします。

システムファームウェア (BIOS または UEFI) は、メモリー、I/O ポートアパーチャ、PCIe バス番号範囲などのリソースを PCIe トポロジーに割り当てます。そのため、十分なリソースを割り当てるには、ファームウェアによって SR-IOV がサポートおよび有効化されている必要があります。

1.1.1. 概要

- ファームウェア (BIOS または UEFI) は SR-IOV に対応している必要があります。拡張機能がデフォルトで有効になっているかどうかを確認します。そうでない場合は、手動で有効にしてください。これは、仮想化拡張機能 (VT-d または AMD-Vi) を有効にすることに似ています。具体的な詳細については、ベンダーのマニュアルを参照してください。
- ルートポート、または PCIe デバイス (PCIe スイッチなど) のすぐ上流のポートは、ARI をサポートする必要があります。

- PCIe デバイスは SR-IOV に対応している必要があります。

ベンダーの仕様とデータシートをチェックして、お使いのハードウェアが要件を満たしていることを確認してください。

lspci -v コマンドを使用すると、システムにインストールされている PCI デバイスの情報を表示できます。

1.2. デバイス割り当てに関するハードウェアの考慮事項

デバイス割り当ては、仮想ゲストを PCIe デバイスに直接割り当てる機能を提供して、ゲストにフルアクセスを提供し、ネイティブに近いパフォーマンスを提供します。SR-IOV と組み合わせて実装されると、仮想ゲストには VF が直接割り当てられます。このようにして、複数の仮想ゲストを単一の PCIe デバイスの VF に直接割り当てることができます。

SR-IOV を有効にして仮想マシンを PCIe デバイスに直接割り当てる必要はなく、デバイス割り当てが VF を作成するための唯一のアプリケーションでもありませんが、2 つの機能は補完的であり、一緒に使用する場合はハードウェアに関する追加の考慮事項があります。

デバイスの割り当てには、CPU とファームウェアでの I/O メモリー管理ユニット (IOMMU) のサポートが必要です。IOMMU は、I/O 仮想アドレス (IOVA) と物理メモリーアドレスの間で変換を行います。これにより、仮想ゲストはゲストの物理アドレスを使用してデバイスをプログラムでき、IOMMU によってホストの物理アドレスに変換されます。

IOMMU グループは、システム内の他のすべてのデバイスから分離できるデバイスのセットです。IOMMU グループは、IOMMU の粒度と、システム内の他のすべての IOMMU グループからの分離の両方を備えたデバイスの最小セットを表します。これにより、IOMMU は、IOMMU グループ外のデバイスと IOMMU の制御の間のダイレクトメモリーアクセス (DMA) を制限しながら、IOMMU グループとの間のトランザクションを区別できます。

仮想ゲストと PCIe デバイスの仮想機能間のトランザクションの分離は、デバイス割り当ての基本です。PCIe およびサーバー仕様で定義されているアクセス制御サービス (ACS) 機能は、IOMMU グループ内の分離を維持するためのハードウェア標準です。ネイティブ ACS がない場合、またはハードウェアベンダーからの確認がない場合、IOMMU グループ内の多機能デバイスは、IOMMU の保護外で発生する機能間でピアツーピア DMA を公開し、適切な分離がない機能を含むように IOMMU グループを拡張するリスクがあります。

サーバーのルートポートにはネイティブ ACS サポートも推奨されます。そうしないと、これらのポートにインストールされているデバイスがグループ化されます。ルートポートには、プロセッサーベース (ノースブリッジ) のルートポートとコントローラーハブベース (サウスブリッジ) のルートポートの 2 種類があります。上記のように、デバイス割り当てが SR-IOV と組み合わせて使用され、仮想ゲストが VF に割り当てられている場合、これらのポートは ACS と ARI の両方をサポートする必要があります。

Intel の Xeon Processor E5 ファミリー、Xeon Processor E7 ファミリー、および High End Desktop Processor には、プロセッサーベースのルートポートでのネイティブ ACS サポートが含まれています。

Intel Platform Controller Hub ベース (PCH) の PCI Express Root Port は現在、ACS をサポートしていないか、非標準の ACS 実装を使用しているため、これらのルートポートを介して接続されたデバイスをきめ細かく分離することは困難です。これらのルートポートの多くは、ACS と同等の機能をサポートしています。Red Hat Enterprise Linux 7.3 カーネルには、X79、X99、5 シリーズから 9 シリーズ、および 100 シリーズ PCIExpress チップセットでこの ACS と同等の機能を有効にするためのサポートが含まれています。

ルートポートが ACS をサポートしていることを確認するために、PCIe デバイスをインストールするときにプロセッサーベースおよびコントローラーハブベースのルートポートを決定する方法については、ベンダーの仕様を参照してください。

さらに、I/O トポロジー内の PCIe スイッチまたはブリッジにも ACS サポートが必要です。そうでない場合、IOMMU グループが拡張される可能性があります。

1.2.1. 概要

- CPU が IOMMU (例: VT-d または AMD-Vi) をサポートしていること。IBM POWER8 はデフォルトで IOMMU をサポートしています。
- ファームウェアが IOMMU をサポートしていること。
- 使用する CPU ルートポートが ACS または ACS と同等の機能をサポートしていること。
- PCIe デバイスが ACS または ACS と同等の機能をサポートしていること。
- PCIe デバイスとルートポート間のすべての PCIe スイッチおよびブリッジが ACS をサポートすることをお勧めします。たとえば、スイッチが ACS をサポートしていない場合には、そのスイッチの背後にあるデバイスはすべて同じ IOMMU グループを共有し、同じ仮想マシンにしか割り当てることができません。

ベンダーの仕様とデータシートをチェックして、お使いのハードウェアが要件を満たしていることを確認してください。

lspci -v コマンドを使用すると、システムにインストールされている PCI デバイスの情報を表示できます。

付録A 法的通知