



# Red Hat Virtualization

## 4.1

### SR-IOV를 구현하기 위한 하드웨어 고려 사항

---

Red Hat Virtualization에서 SR-IOV를 구현하기 위한 하드웨어 고려 사항

Red Hat Virtualization 문서화 팀Red Hat



# Red Hat Virtualization 4.1 SR-IOV를 구현하기 위한 하드웨어 고려 사항

---

## Red Hat Virtualization에서 SR-IOV를 구현하기 위한 하드웨어 고려 사항

Red Hat Virtualization 문서화 팀  
Red Hat Customer Content Services  
rhev-docs@redhat.com

## 법적 공지

Copyright © 2017 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, 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 Software Collections 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.

## 초록

Red Hat Enterprise Linux에서 SR-IOV를 구현하고 Red Hat Virtualization에서 장치를 할당하는 데 필요한 하드웨어 고려 사항 가이드입니다.

---

## 차례

<b>1. 소개</b> .....	<b>2</b>
1.1. SR-IOV를 구현하기 위한 하드웨어 고려 사항 요약	2
<b>2. 장치 할당 시 필요한 추가적인 하드웨어 고려 사항</b> .....	<b>2</b>
2.1. 장치 할당 시 하드웨어 고려 사항 요약	3

## 1. 소개

SR-IOV(Single Root I/O Virtualization)는 단일 PCI Express (PCIe) 끝점이 여러 개별 장치로 사용될 수 있게 해주는 하드웨어 레퍼런스입니다. 이를 위해 두 개의 PCIe 기능인 물리적 기능(PF)과 가상 기능(VF)을 도입합니다.

PF는 일반적인 PCIe 기능이며 여기에는 SR-IOV 기능, 데이터 이동과 같은 PCIe 장치의 완전한 설정 및 제어 등의 기능이 포함됩니다. 각 PCIe 장치는 PF를 한 개에서 여덟 개까지 사용할 수 있습니다.

VF는 경량 PCIe 기능이며 데이터 이동에 필요한 리소스와 최소한의 설정 리소스 세트를 포함합니다. 각 PF에 여러 개의 VF를 생성할 수 있으며 각 PF는 다른 수의 VF를 지원할 수 있습니다. PCIe 장치 공급 업체에 따라 허용되는 총 VF 개수가 다르며 장치마다 다릅니다.

PCIe 사양으로 더 많은 수의 VF를 지원하려면 Alternative Routing ID Interpretation (ARI)을 구현하여 PCIe 헤더에 있는 장치 번호 필드를 다시 환산해서 8개가 넘는 VF를 사용할 수 있게 해줍니다. 이러한 환산은 PCIe 장치와 이 장치의 바로 위(upstream)에 있는 상위 포트(ARI를 지원하는 root 포트 또는 스위치) 모두에 의존합니다.

시스템 펌웨어(BIOS or UEFI)는 PCIe 토폴로지를 위해 메모리, I/O 포트 애퍼처(aperture), PCIe 버스 번호 범위 등과 같은 리소스를 할당합니다. 펌웨어가 SR-IOV를 지원 및 활성화해야 리소스를 충분히 할당할 수 있습니다.

### 1.1. SR-IOV를 구현하기 위한 하드웨어 고려 사항 요약

- ▶ 펌웨어(BIOS or UEFI)가 SR-IOV를 지원해야 합니다. 기본적으로 확장 기능이 활성화되어 있는지 확인합니다. 활성화되어 있지 않은 경우 수동으로 활성화합니다. 이는 가상화 확장 기능(VT-d or AMD-Vi)을 활성화하는 방법과 비슷합니다. 보다 자세한 내용은 공급 업체 사용 설명서를 참조하십시오.
- ▶ Root 포트 또는 PCIe 장치의 바로 업스트림에 있는 상위 포트(예: PCIe 스위치)가 ARI를 지원해야 합니다.
- ▶ PCIe 장치가 SR-IOV를 지원해야 합니다.

하드웨어가 이러한 요구 사항을 충족하는지 확인하기 위해 공급 업체 사양 및 데이터시트를 참조하시기 바랍니다.

**lspci -v** 명령을 사용하여 시스템에 이미 설치되어 있는 PCI 장치에 대한 정보를 표시합니다.

## 2. 장치 할당 시 필요한 추가적인 하드웨어 고려 사항

장치 할당을 통해 가상 게스트를 PCIe 장치에 직접 할당할 수 있으며, 그 결과 해당 게스트에게 완전한 액세스를 주고 네이티브에 가까운 (near-native) 성능을 제공합니다. 가상 게스트를 SR-IOV와 함께 구현하면 VF에 직접 연결됩니다. 이를 통해 다수의 가상 게스트를 단일 PCIe 장치에 있는 VF에 직접 할당할 수 있습니다.

가상 머신을 PCIe 장치에 직접 할당하기 위해 SR-IOV가 활성화되어 있지 않아도 되며 장치 할당이 VF를 생성하는 유일한 애플리케이션도 아니지만, 이 두 가지 기능은 상호 보완 관계에 있으며 이를 함께 사용하려면 추가적인 하드웨어 고려 사항이 있습니다.

CPU 및 펌웨어가 I/O Memory Management Unit(IOMMU)을 지원해야 장치를 할당할 수 있습니다. IOMMU는 I/O Virtual Addresses(IOVA)와 물리적인 메모리 주소 사이에서 변환을 합니다. 이를 통해 가상 게스트가 물리적 주소로 장치를 프로그래밍하면 IOMMU가 이 주소를 호스트의 물리적인 주소로 변환합니다.

IOMMU 그룹이란 시스템 내의 다른 모든 장치에서 분리할 수 있는 장치의 모음입니다. IOMMU 그룹은 시스템 내의 다른 모든 IOMMU 그룹으로부터 분리되며 IOMMU 세분성도 있는 가장 작은 장치의 모음을 말합니다. 이를 통해 IOMMU는 IOMMU 그룹 외에 있으며 IOMMU의 제어를 벗어난 장치 간의 직접 메모리 액세스(DMA)를 제한하면서 IOMMU 그룹과 주고받는 통신을 구별할 수 있습니다.

장치 할당을 하려면 반드시 가상 게스트와 PCIe 장치의 VF 사이의 통신을 분리해야 합니다. PCIe 및 서버 사양에 정의되어 있는 액세스 제어 서비스 (ACS) 기능은 IOMMU 그룹 내에서 분리를 유지하기 위한 하드웨어 표준입니다. 네이티브 ACS가 없는 경우, 또는 이 기능이 탑재되어 있지 않다는 것을 하드웨어 공급 업체가 명시하지 않은 경우, IOMMU 그룹 내의 다기능 (multifunction) 장치 중 하나로 인해 IOMMU의 보호를 벗어난 상태로 발생하는 기능 간의 peer-to-peer DMA가 공개되어서 그 결과 IOMMU 그룹이 확장되어 제대로 분리되지 않은 기능이 이 그룹에 포함되게 되는 리스크가 발생합니다.

서버의 root 포트에 설치된 장치가 그룹화되지 않도록 하려면 이 root 포트에도 네이티브 ACS 지원이 되는 것이 좋습니다. root 포트에는 프로세서 기반 (northbridge) root 포트와 컨트롤러 허브 기반 (southbridge) root 포트의 두 종류가 있습니다. 앞서 말한 것과 같이 장치 할당 기능을 SR-IOV와 함께 구현하며 가상 게스트가 VF에 연결되는 경우 해당 포트는 ACS 및 ARI를 모두 지원해야 합니다.

Intel의 Xeon 프로세서 E5 제품군, Xeon 프로세서 E7 제품군, 그리고 고급형 데스크탑 프로세서에는 프로세서 기반 root 포트에 네이티브 ACS 지원이 포함되어 있습니다.

Intel 장치에는 일반적으로 컨트롤러 허브 기반 root 포트에 네이티브 ACS 지원이 포함되어 있지 않지만 Red Hat Enterprise Linux 7.2 커널에는 어느 정도 유연성이 있어서 X99, X79, 그리고 5 시리즈부터 9 시리즈까지의 칩셋(chipset) root 포트에서 ACS와 동급의 분리를 활성화할 수 있습니다.

PCIe 장치 설치 시 공급 업체 사양을 참조하여 root 포트가 ACS를 지원하는지 확인한 후에 프로세서 기반 및 컨트롤러 허브 기반 root 포트를 결정하십시오.

또한 I/O 토폴로지 내의 모든 PCIe 스위치 및 브리지도 ACS를 지원해야 합니다. 그렇지 않은 경우 IOMMU 그룹이 확장될 수 있습니다.

## 2.1. 장치 할당 시 하드웨어 고려 사항 요약

- ※ CPU가 IOMMU(예: VT-d 또는 AMD-Vi)를 지원해야 합니다. 기본적으로 IBM POWER8은 IOMMU를 지원합니다.
- ※ 펌웨어가 IOMMU를 지원해야 합니다.
- ※ CPU root 포트가 ACS 또는 ACS와 동급의 기능을 지원해야 합니다.
- ※ PCIe 장치가 ACS 또는 ACS와 동급의 기능을 지원해야 합니다.
- ※ 모든 PCIe 스위치, 그리고 PCIe 장치와 root 포트 간의 모든 브리지가 ACS를 지원하는 것이 권장됩니다. 예를 들어, 어떤 스위치가 ACS를 지원하지 않는 경우 해당 스위치의 모든 장치는 같은 IOMMU 그룹을 공유하기 때문에 동일한 가상 머신에만 할당될 수 있습니다.

하드웨어가 이러한 요구 사항을 충족하는지 확인하기 위해 공급 업체 사양 및 데이터시트를 참조하시기 바랍니다.

**lspci -v** 명령을 사용하여 시스템에 이미 설치되어 있는 PCI 장치에 대한 정보를 표시합니다.