Proxmox 6.2 GPU Passthrough Tutorial

- Proxmox 6.2 GPU Passthrough Tutorial
 - <u>Hardware</u>
 - BIOS Config
 - A. Enable VT-x
 - B. Enable VT-d
 - Host Config
 - A. Load Required Modules and Block GPU Drivers
 - B. Enable the IOMMU for systemd-boot (Proxmox on UEFI)
 - <u>C. IOMMU Interrupt Remapping</u>
 - D. Verification
 - <u>Guest VM Config</u>
 - A. Add PCI Device to VM
 - B. Other VM Hardware Configs
 - Guest OS Config
 - A. Prerequisities
 - B. Install NVIDIA Driver
 - C. GRUB Options
 - D. Verification
 - <u>References</u>

Hardware

MB: ASUS TUF GAMING B460-PLUS CPU: Intel i7-10700 GPU: NVIDIA GTX-1060

BIOS Config

Your hardware needs to support IOMMU (I/O Memory Management Unit) interrupt remapping, this includes the CPU and the mainboard.

A. Enable VT-x

In the Asus UEFI BIOS, this feature is in "Advanced -> CPU configuration" and is named "Intel Virtualization Technology".

10/26/23, 2:13 PM	Pro	oxmox 6.2 GPU Passthrough Tu	torial - HackMD	
VISUIS UEFI BIOS Utility - Ad	vanced Mode			
11/23/2015 22:28[‡] ⊕ English	🗐 MyFavorite(F3) 🛛 $ & \ensuremath{\mathcal{P}}$ Qfan C	Control(F6) 😳 EZ Tuning Wizard(F11)	🗐 Quick Note(F9)	Hot Keys
My Favorites Main Ai T	weaker <u>Advanced</u>	Monitor Boot Tool Ex	it 🔄 F	lardware Monitor
L3 Cache		6 MB	CPU	
L4 Cache		Not Supported	Frequ	ency Temperature
Intel Adaptive Thermal Monitor Proxmox 6.2 GPU Pass	Try 📑 HackMI	(https://hackmd.io?utm_s	ource=view-page&	^{IMHz} 30℃ <u>utm_medium=logo-nav)</u>
Active Processor Cores		All	BCLK	Vcore MHz 1.024 V
Limit CPUID Maximum		Disabled	Ratio	
Execute Disable Bit		Enabled	32x	
Intel Virtualization Technology		Enabled	Mer	nory
Hardware Prefetcher(L2 Cache)		Enabled	- Frequ 1866	iency Voltage MHz 1.517 V
Adjacent Cache Line Prefetcher		Enabled	Capa	ity
Boot performance mode		Max Non-Turbo Performa	3276	В МВ
Dynamic Storage Accelerator		Disabled	- Volt	age
➤ CPU Power Management Configuration	on		+12V 12.00	+5V 00 V 5.040 V
intel Virtualization Technology makes a independent operating systems to be re	single system appear as multiple in unning simultaneously on a single :	ndependent systems to software. This allow system.	ws for multiple, +3.3\ 3.328	v
			Last Modified	EzMode(F7) -
	Version 2.16.1240. Copyrigh	nt (C) 2014 American Megatrends, Inc.	Informa	ttiVleb.net 🖑

B. Enable VT-d

Then, if your motherboard supports it, you will find the "VT-d" option that matches IOMMU in "Advanced -> System Agent Configuration" or "Advanced -> North Bridge".



PS. SR-IOV option can also be found on this MB/BIOS, but will not be used in this tutorial.

Host Config

A. Load Required Modules and Block GPU Drivers

Load VFIO Modules

Add to /etc/modules

```
vfio
vfio_iommu_type1
vfio_pci
vfio_virqfd
```

Block GPU Drivers

Block the original GPU drivers for attaching vfio-pci driver to the devices

echo	"blacklist	radeon" >> /etc/modprobe.d/blacklist.conf
echo	"blacklist	<pre>nouveau" >> /etc/modprobe.d/blacklist.conf</pre>
echo	"blacklist	<pre>nvidia" >> /etc/modprobe.d/blacklist.conf</pre>

B. Enable the IOMMU for systemd-boot (Proxmox on UEFI)

Get VendorID and DeviceID of the GPU

lspci -nn | grep -i nvidia

Sample output

01:00.0 VGA compatible controller [0300]: NVIDIA Corporation GP106 [GeForc 01:00.1 Audio device [0403]: NVIDIA Corporation GP106 High Definition Audi

.

=> For the first one: PCI ID: 01:00.0, VendorID: 10de, DeviceID: 1c03

Add kernel parameters

Add the following parameters into /etc/kernel/cmdline

intel_iommu=on vfio-pci.ids=<VendorID>:<DeviceID>,<VendorID>:<DeviceID> di

For example /etc/kernel/cmdline becomes

```
root=ZFS=rpool/ROOT/pve-1 boot=zfs intel_iommu=on vfio-pci.ids=10de:1c03,1
```

•

Update and reboot

```
pve-efiboot-tool
reboot
```

Check if kernel parameters are loaded.

C. IOMMU Interrupt Remapping

It will not be possible to use PCI passthrough without interrupt remapping.

To identify if your system has support for interrupt remapping

dmesg | grep 'remapping'

Sample output

[0.190148] DMAR-IR: Queued invalidation will be enabled to support x2a [0.191599] DMAR-IR: Enabled IRQ remapping in x2apic mode

If you see one of the following lines, then remapping is supported.

- AMD-Vi: Interrupt remapping enabled
- DMAR-IR: Enabled IRQ remapping in x2apic mode" ('x2apic' can be different on old CPUs, but should still work)

Allow Unsafe Interrupts (if interrupt remapping is not supported)

If your system **doesn't support** interrupt remapping, you can allow **unsafe interrupts** (not tested by me)

```
echo "options vfio_iommu_type1 allow_unsafe_interrupts=1" > /etc/modprobe.
update-initramfs -u
reboot
```

.

D. Verification

Kernel Parameters Loaded

cat /proc/cmdline

IOMMU Working

dmesg | grep -E "DMAR|IOMMU"

Sample output

Γ 0.009442] ACPI: DMAR 0x00000007936D000 0000A8 (v01 INTEL EDK2 Θ 0.110221] DMAR: IOMMU enabled Γ 0.190123] DMAR: Host address width 39 Γ 0.190125] DMAR: DRHD base: 0x000000fed90000 flags: 0x0 Γ Г 0.190131] DMAR: dmar0: reg_base_addr fed90000 ver 1:0 cap 1c0000c4066 0.190133] DMAR: DRHD base: 0x000000fed91000 flags: 0x1 Г Γ 0.190137] DMAR: dmar1: reg_base_addr fed91000 ver 1:0 cap d2008c40660 0.190140] DMAR: RMRR base: 0x00000079945000 end: 0x00000079b8efff Г 0.190142] DMAR: RMRR base: 0x0000007b000000 end: 0x0000007f7fffff Г 0.190144] DMAR-IR: IOAPIC id 2 under DRHD base 0xfed91000 IOMMU 1 Γ 0.190146] DMAR-IR: HPET id 0 under DRHD base 0xfed91000 Г 0.190148] DMAR-IR: Queued invalidation will be enabled to support x2a Γ Γ 0.191599] DMAR-IR: Enabled IRQ remapping in x2apic mode 0.942691] DMAR: No ATSR found Γ 0.942753] DMAR: dmar0: Using Queued invalidation Γ Γ 0.942757] DMAR: dmar1: Using Queued invalidation 0.951433] DMAR: Intel(R) Virtualization Technology for Directed I/O Г

VFIO Working

dmesg | grep -i vfio

Should see messages from vfio_pci driver.

Sample output

[0.000000]	Command line: initrd=\EFI\proxmox\5.4.34-1-pve\initrd.img-5
[0.110162]	<pre>Kernel command line: initrd=\EFI\proxmox\5.4.34-1-pve\initr</pre>
[0.987220]	VFIO - User Level meta-driver version: 0.3
[0.987271]	vfio-pci 0000:01:00.0: vgaarb: changed VGA decodes: olddeco
[1.006157]	<pre>vfio_pci: add [10de:1c03[fffffffffffffffff]] class 0x000000</pre>
[1.026154]	<pre>vfio_pci: add [10de:10f1[fffffffffffffff]] class 0x000000</pre>
[5.320737]	vfio-pci 0000:01:00.0: vgaarb: changed VGA decodes: olddeco
[36.517767]	vfio-pci 0000:01:00.0: vfio_ecap_init: hiding ecap 0x19@0x9
[36.520535]	vfio-pci 0000:01:00.0: Invalid PCI ROM header signature: ex
[36.538147]	vfio-pci 0000:01:00.1: enabling device (0000 -> 0002)
[39.276682]	vfio-pci 0000:01:00.0: Invalid PCI ROM header signature: ex

•

VF-PCI Driver Loaded

lspci -nnk

Kernel driver in use should be: vfio-pci

Sample output

. . .

	01:00.0 VGA compatible controller [0300]: NVIDIA Corporation GP106 [GeForc
	Subsystem: Micro-Star International Co., Ltd. [MSI] GP106 [GeForce
	Kernel driver in use: vfio-pci
	Kernel modules: nvidiafb, nouveau
	01:00.1 Audio device [0403]: NVIDIA Corporation GP106 High Definition Audi
	Subsystem: Micro-Star International Co., Ltd. [MSI] GP106 High Def
	Kernel driver in use: vfio-pci
	Kernel modules: snd_hda_intel
-	

IOMMU Groups Isolation

For working PCI passthrough, you need a dedicated IOMMU group for all PCI devices you want to assign to a VM.

```
find /sys/kernel/iommu_groups/ -type 1
```

Sample output

/sys/kernel/iommu_groups/7/devices/0000:00:1c.0
/sys/kernel/iommu_groups/7/devices/0000:04:00.3
/sys/kernel/iommu_groups/7/devices/0000:04:00.1
/sys/kernel/iommu_groups/7/devices/0000:04:00.2
/sys/kernel/iommu_groups/7/devices/0000:00:1c.4
/sys/kernel/iommu_groups/7/devices/0000:04:00.0
/sys/kernel/iommu_groups/5/devices/0000:00:17.0
/sys/kernel/iommu_groups/3/devices/0000:00:14.0
/sys/kernel/iommu_groups/1/devices/0000:00:01.0
/sys/kernel/iommu_groups/1/devices/0000:01:00.0
/sys/kernel/iommu_groups/1/devices/0000:01:00.1
/sys/kernel/iommu_groups/8/devices/0000:05:00.0
/sys/kernel/iommu_groups/8/devices/0000:00:1d.0
/sys/kernel/iommu_groups/6/devices/0000:00:1b.0
/sys/kernel/iommu_groups/4/devices/0000:00:16.0
/sys/kernel/iommu_groups/2/devices/0000:00:02.0
/sys/kernel/iommu_groups/0/devices/0000:00:00.0
/sys/kernel/iommu_groups/9/devices/0000:00:1f.2
/sys/kernel/iommu_groups/9/devices/0000:00:1f.0
/sys/kernel/iommu_groups/9/devices/0000:00:1f.3
/sys/kernel/iommu_groups/9/devices/0000:00:1f.6
/sys/kernel/iommu_groups/9/devices/0000:00:1f.4

Guest VM Config

Guest OS: Ubuntu 18.04.4 Server

A. Add PCI Device to VM

Only user root can add PCI device to guest VM

Hardware => Add => PCI Device

	Summary	1	Add	d 🗸 Remove	Edit	Resize disk Move disk Revert
>_	Console	Ģ	₿	Hard Disk		2.00 GiB [balloon=0]
Ţ	Hardware	ŧ	👔 💿 CD/DVD Drive	CD/DVD Drive	8 (1 sockets, 8 cores)	
۵	Cloud-Init	-	=	Network Device		SeaBIOS
ø	Options	C	۲ ۵	EFI Disk		Default
		4	•	USB Device		q35
=	Task History			PCI Device		VirtIO SCSI
۲	Monitor	0		Serial Port CloudInit Drive Audio Device VirtIO RNG	2)	none media=cdrom
B	Backup		•		_,	local-zfervm-104-disk-0 size=20G
•7	Poplication		4 0)		-+0)	virtia EE:41:54:80:60:80 bridge verbat0 figure II 1
L.+	Replication	4	•		ett)	virtio=5E:41:5A:B9:60:D0,bridge=vmbr10,firewall=1
୭	Snapshots	Ŀ	₽⊦	CI Device (hostp	ci0)	01:00,pcie=1,x-vga=1
D	Firewall >					

If the steps above are successfully done, should see the GPU here

Add: PCI Devic	се			\otimes	
Device:			✓ MDev Type:	~	
All Functions:	ID 个	IO	Vendor	Device	Media
	0000:00:17.0	5	Intel Corporation		No
ROM-Bar:	0000:00:1f.3	9	Intel Corporation		No
	0000:00:1f.4	9	Intel Corporation		No
Help	0000:00:1f.6	9	Intel Corporation		No
	0000:01:00.0	1	NVIDIA Corporation	GP106 [GeForce GTX 1060 6GB]	No
	0000:01:00.1	1	NVIDIA Corporation		No
	0000:04:00.0	7	Intel Corporation	I350 Gigabit Network Connection	No
	0000:04:00.1	7	Intel Corporation	1350 Gigabit Network Connection	No
	0000:04:00.2	7	Intel Corporation	1350 Gigabit Network Connection	No
	0000:04:00.3	7	Intel Corporation	I350 Gigabit Network Connection	No
	0000:05:00.0	8	Micron/Crucial Tec		No

Check all options

- [v] All Functions
- [v] Primary GPU
- [v] ROM-Bar
- [v] PCI-Express

10/26/23, 2:13 PM

Proxmox 6.2 GPU Passthrough Tutorial - HackMD

Add: PCI Devi	ce		\otimes
Device:	0000:01:00.0 ~	MDev Type:	~ ~
All Functions:		Primary GPU:	
ROM-Bar:		PCI-Express:	
Help			Advanced 🗹 🛛 Add

B. Other VM Hardware Configs

Choose either option 1 or option 2 to configure.

Warning: lost of the web GUI terminal access

Once the *Primary GPU* (x-vga=1) option is set, the VNC consle on Web GUI will **NOT** able to connect to the VM, so make sure you are able to access the VM by **SSH**, and the network setting in the VM will not be changed after reboot.

Option 1. Edit the VM Config

Edit /etc/pve/qemu-server/<VMID>.conf

```
bios: seabios
machine: q35
hostpci0: 01:00,pcie=1,x-vga=1
```

Option 2. By the Web GUI

Set BIOS, Machine, PCI Device (hostpci0) by the Web GUI

Summary	Add ~ Remove Edit	Resize disk Move disk Revert
>_ Console	🚥 Memory	2.00 GiB [balloon=0]
- Hardware	Processors	8 (1 sockets, 8 cores)
Cloud-Init	BIOS	SeaBIOS
Options	Display	Default
Task History	🌣 Machine	q35
	SCSI Controller	VirtIO SCSI
Monitor	 CD/DVD Drive (ide2) 	none,media=cdrom
🖺 Backup	🖨 Hard Disk (scsi0)	local-zfs:vm-104-disk-0,size=20G
13 Replication		virtio=5E:41:5A:B9:60:D0,bridge=vmbr10,firewall=1
Snapshots	📼 PCI Device (hostpci0)	01:00,pcie=1,x-vga=1

Guest OS Config

A. Prerequisities

```
sudo add-apt-repository ppa:graphics-drivers/ppa
sudo apt update
sudo apt install ubuntu-drivers-common
```

B. Install NVIDIA Driver

Check the latest available driver for the GPU

```
ubuntu-drivers devices
```

Sample Output

```
== /sys/devices/pci0000:00/0000:00:1c.0/0000:01:00.0 ==
modalias : pci:v000010DEd00001C03sv00001462sd00003283bc03sc00i00
vendor : NVIDIA Corporation
model : GP106 [GeForce GTX 1060 6GB]
driver : nvidia-driver-415 - third-party free
driver : nvidia-driver-435 - distro non-free
driver : nvidia-driver-440 - distro non-free
driver : nvidia-driver-440-server - distro non-free
driver : nvidia-driver-440-server - distro non-free
driver : nvidia-driver-440 - distro non-free
driver : nvidia-driver-440 - third-party free
driver : nvidia-driver-410 - third-party free
driver : nvidia-driver-450 - third-party free recommended
driver : nvidia-driver-418-server - distro non-free
driver : xserver-xorg-video-nouveau - distro free builtin
```

=> The output shows that we can install nvidia-driver-440-server

sudo apt install nvidia-driver-440-server nvidia-utils-440-server
sudo reboot

Verification

lsmod | grep nvidia

C. GRUB Options

Warning: lost of access to the VNC terminal

There will be no output to the VNC terminal after these options are set, make the ssh is ready.

Add the following options to GRUB_CMDLINE_LINUX_DEFAULT in /etc/default/grub

video=vesafb:off,efifb:off

Run

sudo update-grub sudo reboot

D. Verification

sudo nvidia-smi

Sample output

```
Sat Aug 1 19:45:48 2020
Driver Version: 440.95.01
| NVIDIA-SMI 440.95.01
                        CUDA Version: 10.2
| GPU Name Persistence-M| Bus-Id Disp.A | Volatile Uncorr.
| Fan Temp Perf Pwr:Usage/Cap|
                   Memory-Usage | GPU-Util Compute
0 GeForce GTX 106... Off | 00000000:01:00.0 Off |
      P5
        14W / 140W |
  0%
   45C
                 0MiB / 6078MiB |
                           0%
                               Defa
 | Processes:
                              GPU Mem
  GPU PID Type Process name
                              Usage
No running processes found
               -----
```

=> The GPU information is successfully retrieved without any error.

References

BIOS

Proxmox 6.2 GPU Passthrough Tutorial - HackMD

<u>https://us.informatiweb.net/tutorials/it/bios/enable-iommu-or-vt-d-in-your-bios.html</u>
 <u>(https://us.informatiweb.net/tutorials/it/bios/enable-iommu-or-vt-d-in-your-bios.html)</u>

PVE

- <u>https://pve.proxmox.com/wiki/Pci_passthrough</u>
 (<u>https://pve.proxmox.com/wiki/Pci_passthrough</u>)
- <u>https://mathiashueber.com/pci-passthrough-ubuntu-2004-virtual-machine/</u> (<u>https://mathiashueber.com/pci-passthrough-ubuntu-2004-virtual-machine/</u>) <u>https://stackoverflow.com/questions/48199261/proc-cmdline-does-not-updated-</u> <u>with-update-grub (https://stackoverflow.com/questions/48199261/proc-cmdline-does-not-updated-with-update-grub)</u>
- <u>https://pve.proxmox.com/wiki/PCI(e)_Passthrough</u>
 (<u>https://pve.proxmox.com/wiki/PCI(e)_Passthrough</u>) (outdated)

Guest OS

- <u>https://gitpress.io/@chchang/install-nvidia-driver-cuda-pgstrom-in-ubuntu-1804</u>
 (<u>https://gitpress.io/@chchang/install-nvidia-driver-cuda-pgstrom-in-ubuntu-1804</u>)
- <u>https://www.facebook.com/groups/pve.tw/permalink/1556562391178983/</u>
 <u>(https://www.facebook.com/groups/pve.tw/permalink/1556562391178983/)</u>