

**NVIDIA vGPU types** ..... 3



# NVIDIA vGPU types

NVIDIA vGPU가 mdevctl types 가 vGPU

```
root@pve:~# mdevctl types
0000:01:00.0
  nvidia-156
    Available instances: 11
    Device API: vfio-pci
    Name: GRID P40-2B
    Description: num_heads=4, frl_config=45, framebuffer=2048M,
max_resolution=5120x2880, max_instance=12
  nvidia-215
    Available instances: 11
    Device API: vfio-pci
    Name: GRID P40-2B4
    Description: num_heads=4, frl_config=45, framebuffer=2048M,
max_resolution=5120x2880, max_instance=12
  nvidia-241
    Available instances: 0
    Device API: vfio-pci
    Name: GRID P40-1B4
    Description: num_heads=4, frl_config=45, framebuffer=1024M,
max_resolution=5120x2880, max_instance=24
  nvidia-46
    Available instances: 0
    Device API: vfio-pci
    Name: GRID P40-1Q
    Description: num_heads=4, frl_config=60, framebuffer=1024M,
max_resolution=5120x2880, max_instance=24
  nvidia-47
    Available instances: 11
    Device API: vfio-pci
    Name: GRID P40-2Q
    Description: num_heads=4, frl_config=60, framebuffer=2048M,
max_resolution=7680x4320, max_instance=12
  nvidia-48
    Available instances: 0
    Device API: vfio-pci
    Name: GRID P40-3Q
    Description: num_heads=4, frl_config=60, framebuffer=3072M,
max_resolution=7680x4320, max_instance=8
  ...
  nvidia-62
    Available instances: 0
    Device API: vfio-pci
    Name: GRID P40-1B
    Description: num_heads=4, frl_config=45, framebuffer=1024M,
```

max\_resolution=5120x2880, max\_instance=24

Name P40-1B, P40-1Q, P40-3A

vGPU		SM	CUDA ( )	
P40-1Q	1GB	1 SM	~128	가
P40-2Q	2GB	2 SM	~256	AI
P40-3Q	3GB	3 SM	~384	-
P40-8Q	8GB	8 SM	~1,024	
P40-24Q	24GB	30 SM ( )	~3,840	

Q (Fixed Share)	vGPU ( 가 )
B (Best Effort)	SM
A (Equal Share)	vGPU가 SM

Q . A 가 GPU  
 B . A 가 GPU  
 P40-2Q/B/A 8 가 .

	Q (Fixed)	B (Best Effort)	A (Equal Share)
1 VM (7 )	100% 가	100% 가	1/8(12.5%)
8 VM	VM 1/8	가 ( VM 가 )	VM 1/8
AI ( )			
VDI/가 ( )		(B4 가 )	
/			

From: <https://at1.kr/dokuwiki/> - AllThatLinux!

Permanent link: [https://at1.kr/dokuwiki/doku.php/nvidia\\_vgpu\\_types](https://at1.kr/dokuwiki/doku.php/nvidia_vgpu_types)

Last update: 2025/11/17 08:32

