SMC VDI + NVIDIA vGPU

for Virtual Workstation

Supermicro's VDI solution for the Virtual Workstation

- Leveraging Supermicro's expertise in server design with powerful NVIDIA GPUs and NVIDIA vGPU software
- Great solution to build high performance VMs for remote workstation for high powered graphics users
- NVIDIA vGPU Software virtualizes an NVIDIA GPU to provide physical workstation user experience in a virtual environment
  - Able to allocate multiple VMs per GPU or multiple GPUs for a single VM
  - Enables users to use any device, anywhere and get physical workstation like performance
  - Centralized data on the server, secured in the data center, rather than client devices

Supermicro Servers with NVIDIA Quadro vDWS

- Offers sizes of 2U GPU server and 4U GPU server
  - 2029GP-TR and 4029GP-TRT
- Flexibility to reconfigure VMs to the profile size the users need
  - Each GPU can be allocated differently. And up to 4 GPUs can be combined for 1VM
- Offers variety of sizing and scaling to be able to match each user's power and performance needs

Supermicro's Good Solution

- Comparable pricing to WKSTN with NVIDIA Quadro RTX 4000
- BYOD allows for any device to have the workstation like performance while using the VM with vGPU
- 2 Different reference sizing for 2U or 4U systems
- Leverages NVIDIA Quadro RTX 6000 with 8Q Profile (8GB framebuffer)
  - 3 CCU per GPU for 15 CCU in 2U or 24 CCU in 4U
  - Can be reallocated to however is necessary

Supermicro's Best Solution

- Comparable pricing to WKSTN with NVIDIA Quadro RTX 5000
- BYOD allows for any device to have the workstation like performance while using the VM with vGPU
- 2 Different reference sizing for 2U or 4U systems
- Leverages NVIDIA Quadro RTX 8000 with 16Q profile (16 GB framebuffer)
  - 3 CCU per GPU for 15 CCU in 2U or 24 CCU in 4U
  - Can be reallocated to however is necessary
- RTX 8000 has double vRAM than RTX 6000
  - Allows for same density with higher framebuffer (for bigger models)

NVIDIA GPUs for Quadro vDWS Solution



	NVIDIA QUADRO RTX 8000	NVIDIA QUADRO RTX 6000	NVIDIA T4
GPU	1 NVIDIA Turing GPU	1 NVIDIA Turing GPU	1 NVIDIA Turing GPU
CUDA CORES	4,608	4,608	2,860
Tensor Cores	576	576	320
RT Cores	72	72	40
FP32 Peak Perf	**14.9 TFlops**		**8.1 TFlops**
Max CCU	**24 2B**	**24 1B**	**16 1B**
MEMORY SIZE	48 GB GDDR6	24 GB GDDR6	16 GB GDDR6
FORM FACTOR	PCIe 3.0 Dual Slot		PCIe 3.0 Single Slot
POWER	295W / 250W Passive		70W
THERMAL	Active / Passive		Passive
	PEFORMANCE Optimized		**BALANCED**

Supermicro Quadro vDWS Reference Comparison

	2U Solution		4U Solution	
Server	**SYS-2029GP-TR (2U)**		**SYS-4029GP-TRT (4U)**	
Support User	**15 CCU**		**24 CCU**	
NVIDIA GRID SW	**Quadro vDWS (8GB FB)**	**Quadro vDWS (16GB FB)**	**Quadro vDWS (8GB FB)**	**Quadro vDWS (16GB FB)**
NVIDIA GPU	**5 x NVIDIA RTX 6000-P** GPU-NVQRTX6000-P	**5 x NVIDIA RTX 8000-P** GPU-NVQRTX8000-P	**8 x NVIDIA RTX 6000-P** GPU-NVQRTX6000-P	**8 x NVIDIA RTX 8000-P** GPU-NVQRTX8000-P
CPU	Can use lower core CPU because less VMs 6226R 16C @ 2.9 GHZ		Need higher core CPU because more VMs 6248R 24C @ 3.0 GHZ	
System Memory	Recommend 960 GB ~64GB/CCU		Recommend 1536 GB ~64GB/CCU	

