

SMC VDI + NVIDIA vGPU for Virtual Desktop



VDI solution for the Virtual Desktop

- Leveraging ServerDirect's expertise in server design with powerful NVIDIA GPUs and NVIDIA vGPU software
- Great solution for scaling out dense Virtual Desktop VDI that gives great User Experience while being cost effective and improving density of the VDI
- NVIDIA vGPU Software allows for virtualization of NVIDIA GPU to provide native, physical desktop experience in virtual environments
 - Many programs, including Win10 itself is graphically intensive and needs a GPU to offload the graphics workloads
 - Every computer has some sort of GPU to do GPU enabled tasks. VMs also need virtualized GPU resources to do the same
 - Easy to monitor remotely by IT and troubleshoot on the VM and not the devices

Good Solution

- Lowest Total Cost per user
- Better UX compared to VPN or basic VDI solution
- Basic office capable VDI system using NVIDIA GRID vPC 1B profile
- Uses 2 NVIDIA M10 GPUs in 2U form factor for up to 64 CCU
 - Upgradeable to 4 NVIDIA T4 GPUs for better GPU performance
- Good solution for organizations on a tight budget while offering a great user experience for knowledge workers using office productivity applications
- Limited upgradability using 2029U-E1CRT

Best Solution

- Much better UX compared to VPN or basic VDI solution
- Higher end VDI Graphics support with up NVIDIA GRID vPC 2B profile
- Uses 4 Highest End NVIDIA Quadro RTX Passive cards for up to 96 CCU
 - 96 CCU on NVIDIA GRID vPC 1B profile with 4 Quadro RTX 6000
 - 96 CCU on NVIDIA GRID vPC 2B profile with 4 Quadro RTX 8000
 - Good for running multiple monitors or 4K monitors
- Upgradable path to Quadro vDWS and remote workstation for higher graphics usage applications down the line.
- More upgradable server with room for addition GPUs if necessary in the 2029GP-TRT

Why Best is better

- More CPU, GPU, and RAM performance pool for more options to delegate resources
- 2GB framebuffer is recommended for multiple monitors or 4K monitors
- Hot-Swappable NVMe Drives
- More networking Options
 - 1 SIOM and 3 AOC; vs 1 AOC and 2-10G Ethernet
- Up to 50% more dense infrastructure
 - 96 CCU per 2U; vs 64 CCU per 2U

NVIDIA GPUs for Virtual Desktop



	NVIDIA QUADRO RTX 8000	NVIDIA QUADRO RTX 6000	NVIDIA T4	NVIDIA M10
GPU	1 NVIDIA Turing GPU	1 NVIDIA Turing GPU	1 NVIDIA Turing GPU	4 NVIDIA Maxwell GPUs
CUDA CORES	4,608	4,608	2,860	4 * 640 = Total 2,560
Tensor Cores	576	576	320	-
RT Cores	72	72	40	-
FP32 Peak Perf	14.9 TFlops		8.1 TFlops	6.6 TFlops
Max CCU	24 2B	24 1B	16 1B	32 1B
MEMORY SIZE	48 GB GDDR6	24 GB GDDR6	16 GB GDDR6	32GB GDDR6
FORM FACTOR	PCIe 3.0 Dual Slot		PCIe 3.0 Single Slot	PCIe 3.0 Dual Slot
POWER	295W / 250W Passive		70W	225W
THERMAL	Active / Passive		Passive	Passive
	PERFORMANCE optimized		BALANCED	DENSITY optimized

ServerDirect Windows 10 VDI Reference Comparison

	GOOD Lowest cost per user	BETTER Universal GPU high utilization	BEST (1B) Highest GPU performance GRID vPC @1B	BEST (vDWS) Highest Density GRID vPC @ 2B
Server	SYS-2029U-E1CRT (2U)		SYS-2029GP-TR (2U)	
Support User	64 CCU		96 CCU	
NVIDIA GRID SW	GRID vPC (1GB FB)		GRID vPC (1GB FB)	GRID vPC (2GB FB)
NVIDIA GPU	2 x NVIDIA M10 ~65\$/CCU	4 x NVIDIA T4 ~115\$/CCU	4 x NVIDIA Quadro RTX 6000-P ~145\$/CCU	4 x NVIDIA Quadro RTX 8000-P ~190\$/CCU
Performance/CCU*	~80 CUDA/CCU	~170 CUDA / CCU ~20 Tensor Core / CCU ~2.5 RT Core / CCU	~192 CUDA / CCU ~24 Tensor Core / CCU ~3 RT Core / CCU	
CPU	Slower and Lower Core Count CPU		Faster and Higher Core Count CPU	
System Memory	Recommend 384 GB ~6GB/CCU 24x DIMM slots	Recommend 512 GB ~8GB/CCU 24x DIMM slots	Recommend 768 GB ~8GB/CCU 16x DIMM slots	
Local Storage	More Storage Options Up to 4 NVMe Drives		Less Storage Options Up to 2 Hot Swappable NVMe Drives	
Network	Less Networking Options 2 x 10 GbE and 1 x AOC NIC		More Networking Options 1x SIOM and 3x AOC NIC	
Power Supply	Lower Wattage Power Supply. Limits upgradability		Higher Wattage Power Supply for Higher Watt GPU and CPU	

*Performance does not evenly split as it is based on workload and use case. This is just for estimation of performance availability.

ServerDirect Windows 10 VDI Reference Architecture

	GOOD Lowest cost per user	BETTER Universal GPU high utilization	BEST Highest Density vPC performance w/ GPU
Server	SYS-2029U-E1CRT (2U)		SYS-2029GP-TR (2U)
Support User	64 CCU		96 CCU
NVIDIA GRID SW	64 x GRID vPC (1GB FB) 1yr Subscription: SFT-NVD-G2P1S		96 x GRID vPC (2GB FB) 1yr Subscription: SFT-NVD-G2P1S
NVIDIA GPU	2 x NVIDIA M10 SMC PN : GPU-NVTM10	4 x NVIDIA T4 SMC PN: GPU-NVTT4	4 x NVIDIA Quadro RTX 8000-P SMC PN: GPU-NVQRTX8000-P
CPU	2 x Intel Xeon Gold 6240R (24c, 2.4 GHz)		2 x Intel Xeon Gold 6258R (28c, 2.7 GHz)
System Memory	<ul style="list-style-type: none"> Recommend 384 GB Support up to 24x DIMM slots Up to 3TB ECC RDIMM/LRDIMM, Up to DDR4-2666 	<ul style="list-style-type: none"> Recommend 512 GB Support up to 24x DIMM slots Up to 3TB ECC RDIMM/LRDIMM, Up to DDR4-2666 	<ul style="list-style-type: none"> Recommend 768 GB 16x DIMM slots, up to 2TB 3DS ECC LRDIMM, Up to DDR4-2666
Local Storage	<ul style="list-style-type: none"> Support up to 24 SATA/SAS ports via SAS Expander <ul style="list-style-type: none"> Optional 4 NVMe ports Optional 2 rear hot-swap 2.5" drive bay kit 		<ul style="list-style-type: none"> 8 hot-swap 2.5" SATA/SAS drive bays 2 hot-swap 2.5" SATA/SAS/NVMe drive bays
Network	<ul style="list-style-type: none"> 2 x 10 GbE Optional 1 x AOC NIC 		<ul style="list-style-type: none"> 1x SIOM and 3x AOC NIC
Power Supply	Redundant 1000W Titanium Level digital power supplies		Redundant 2000W Titanium Level digital power supplies