

GPUs on supported platforms

PLATFORM												AMD	
	A100 80GB PCIe	A100 40GB SXM4 (Nvlink)	A100 80GB SXM4 (Nvlink)	A40	A30	A16	A10	A2	M10	T4	MI100	MI210 AVAILABLE during 2022	
	XE8545		Shipping (4 ¹)	Shipping (4 ¹)									
R750xa	Shipping (4 ³)			Shipping (4 ³)	Shipping (4 ³)	Shipping (4 ³)	Shipping (4 ³)	Shipping (4)	Shipping (2 ³)	Shipping (6 ³)	Shipping (4 ³)		
R750	Shipping (2)			Shipping (2)	Shipping (2)	Shipping (2)	Shipping (3)	Shipping (4)	Shipping (2)	Shipping (6)			
R650										Shipping (3)			
C6520										Shipping (1)			
R7525 - Milan	Shipping (3)			Shipping (3)	Shipping (3)	Shipping (3)	Shipping (3)	Shipping (4)	Shipping (2)	Shipping (6)	Shipping (3)		
R7525 - Rome	Shipping (3)			Shipping (3)	Shipping (3)	Shipping (3)	Shipping (3)	Shipping (4)	Shipping (2)	Shipping (6)	Shipping (3)		
R7515 - Milan					Shipping (1)	Shipping (1)		Shipping (3)		Shipping (4)			
R7515 - Rome					Shipping (1)	Shipping (1)		Shipping (3)		Shipping (4)			
R6525 - Rome & Milan										Shipping (3)			
R6515 - Rome & Milan										Shipping (1)			
C6525 - Rome & Milan										Shipping (1)			
XR12	Shipping (2)			Shipping (2)	Shipping (2)			Shipping (2)		Shipping (2)			
XR11								Shipping (2)		Shipping (2)			
DSS8440	Shipping (4/8/10 ¹)			Shipping (4/8/10 ¹)	Shipping (4/8/10 ¹)					Shipping (8/12/16 ¹)			
R940XA	Shipping (4)												
R840									Shipping (2)				
R740/XD	Shipping (3)			Shipping (3)	Shipping (3)	Shipping (3)	Shipping (3)	Shipping (6)	Shipping (2)	Shipping (6 ^{**})			
R640										Shipping (3)			
T640									Shipping (2)				
T550				Shipping(2)	Shipping(2)			Shipping (5)		Shipping (5)			
XR2										Shipping (1)			
XE2420					Shipping (2)		Shipping (2)			Shipping (4)			

1 – XE8545, DSS8440, C4140 are set configs
 2 – subject to change
 3 - R750XA at a minimum requires 2GPUs to be installed at the factory
 (qty) - max number of GPUs allowed, maximum number of GPUs allowed might differ in different configurations on the same platform

Version: May 2022

GPUs on supported platforms

Brand	Model	GPU Memory	Memory ECC	Memory Bandwidth	Max Power Consumption	Graphic Bus/ System Interface	Interconnect Bandwidth	Slot Width	GPU Height/Length	Auxiliary Cable	Workload ¹
AMD	MI210	64 GB HBM2e	Y	1638 GB/sec	300W	PCIe Gen4x16/ Infinity Fabric Link b	64 GB/sec (PCIe 4.0)	DW	FHFL	CPU 8 pin	HPC/Machine learning training
AMD	MI100	32 GB HBM2	Y	1228 GB/sec	300W	PCIe Gen4x16	64 GB/sec (PCIe 4.0)	DW	FHFL	PCIe 8 pin	HPC/Machine learning training
Nvidia	A100	80 GB HBM2	Y	2039 GB/sec	500W	NVIDIA NVLink	600 GB/sec (3rd Gen NVLink)	N/A	N/A	N/A	HPC/AI/Database Analytics
Nvidia	A100	40 GB HBM2	Y	1555 GB/sec	400W	NVIDIA NVLink	600 GB/sec (3rd Gen NVLink)	N/A	N/A	N/A	HPC/AI/Database Analytics
Nvidia	A100	80 GB HBM2e	Y	1935 GB/sec	300W	PCIe Gen4x16/ NVLink bridge ⁸	64 GB/sec ⁵ (PCIe 4.0)	DW	FHFL	CPU 8 pin	HPC/AI/Database Analytics
Nvidia	A30	24 GB HBM2	Y	933 GB/sec	165W	PCIe Gen4x16/ NVLink bridge ⁸	64 GB/sec ⁵ (PCIe 4.0)	DW	FHFL	CPU 8 pin	mainstream AI
Nvidia	A40	48 GB GDDR6	Y	696 GB/sec	300W	PCIe Gen4x16/ NVLink bridge ⁸	64 GB/sec ⁵ (PCIe 4.0)	DW	FHFL	CPU 8 pin	Performance graphics/VDI
Nvidia	A16	64 GB GDDR6	Y	800 GB/sec	250W	PCIe Gen4x16	64 GB/sec (PCIe 4.0)	DW	FHFL	CPU 8 pin	VDI
Nvidia	A2	16 GB GDDR6	Y	200 GB/sec	60W	PCIe Gen 4x8	32 GB/sec (PCIe 4.0)	DW	HHHL	N/A	Inferencing/Edge/VDI
Nvidia	A2	16 GB GDDR6	Y	200 GB/sec	60W	PCIe Gen 4x8	32 GB/sec (PCIe 4.0)	DW	HHHL	N/A	Inferencing/Edge/VDI
Nvidia	A10	24 GB GDDR6	Y	600 GB/sec	150W	PCIe Gen4x16	64 GB/sec (PCIe 4.0)	SW	FHFL	PCIe 8 pin	mainstream graphics/VDI
Nvidia	M10	32 GB GDDR5	N	332 GB/sec	225W	PCIe Gen3x16	32 GB/sec (PCIe 3.0)	DW	FHFL	PCIe 8 pin	VDI
Nvidia	T4	16 GB GDDR6	Y	300 GB/sec	70W	PCIe Gen3x16	32 GB/sec (PCIe 3.0)	SW	HHHL	N/A	Inferencing/Edge/VDI
Nvidia	T4	16 GB GDDR6	Y	300 GB/sec	70W	PCIe Gen3x16	32 GB/sec (PCIe 3.0)	SW	FHHL	N/A	Inferencing/Edge/VDI
Nvidia	A100	40 GB HBM2	Y	1555 GB/sec	250W	PCIe Gen4x16/ NVLink bridge ⁸	64 GB/sec ⁵ (PCIe 4.0)	DW	FHFL	CPU 8 pin	HPC/AI/Database Analytics
Nvidia	V100S	32 GB HBM2	Y	1134 GB/sec	250W	PCIe Gen3x16	32 GB/sec (PCIe 3.0)	DW	FHFL	CPU 8 pin	HPC/AI/Database Analytics
Nvidia	V100	32 GB HBM2	Y	900 GB/sec	250W	PCIe Gen3x16	32 GB/sec (PCIe 3.0)	DW	FHFL	CPU 8 pin	HPC/AI/Database Analytics
Nvidia	V100	16 GB HBM2	Y	900 GB/sec	900 GB/sec	PCIe Gen3x16	32 GB/sec (PCIe 3.0)	DW	FHFL	CPU 8 pin	HPC/AI/Database Analytics
Nvidia	V100	32 GB HBM2	Y	900 GB/sec	300W	NVIDIA NVLink	300 GB/sec (2nd Gen NVLink)	N/A	N/A	N/A	HPC/AI/Database Analytics
Nvidia	V100	16 GB HBM2	Y	900 GB/sec	300W	NVIDIA NVLink	300 GB/sec (2nd Gen NVLink)	N/A	N/A	N/A	HPC/AI/Database Analytics
Nvidia	RTX6000	24 GB GDDR6	Y	624 GB/sec	250W	PCIe Gen3x16/ NVLink bridge ³	32 GB/sec ³ (PCIe 3.0)	DW	FHFL	CPU 8 pin	VDI/ Performance Graphics
Nvidia	RTX8000	48 GB GDDR6	Y	624 GB/sec	250W	PCIe Gen3x16/ NVLink bridge ³	32 GB/sec ³ (PCIe 3.0)	DW	FHFL	CPU 8 pin	VDI/ Performance Graphics
Nvidia	P100	16 GB HBM2	Y	732 GB/sec	300W	NVIDIA NVLink	160 GB/sec (1st Gen NVLink)	N/A	N/A	N/A	HPC/AI/Database Analytics
Nvidia	P100	16 GB HBM	Y	732 GB/sec	250W	PCIe Gen3x16	32 GB/sec (PCIe 3.0)	DW	FHFL	CPU 8 pin	HPC/AI/Database Analytics
Nvidia	P100	12 GB HBM2	Y	549 GB/sec	250W	PCIe Gen3x16	32 GB/sec (PCIe 3.0)	DW	FHFL	CPU 8 pin	HPC/AI/Database Analytics
Nvidia	P40	24 GB DDR5	N	346 GB/sec	250W	PCIe Gen3x16	32 GB/sec (PCIe 3.0)	DW	FHFL	CPU 8 pin	HPC/AI/Database Analytics

¹ suggested ideal workloads, but can be used for other workloads

² Different SKUs are mentioned because different platforms might support different SKUs. This sheet doesn't specifically call out platform-SKU associations

³ upto 100GB/sec when RTX NVLink bridge is used, RTX NVLink bridge is only supported on T640

⁴ Structural Sparsity enabled

⁵ upto 600GB/sec for A100 when NVLink bridge is used, upto 200GB/sec for A30 when NVLink bridge is used, upto 112.5GB/sec for A40 when NVLink bridge is used

⁶ Peak performance numbers shared by Nvidia or AMD for MI100

⁷ Refer to Max#GPUs on supported platforms tab for detail support on Rome vs Milan processors

⁸ A100 w/Nvlink bridge is supported on R750XA and DSS8440, A40 w/Nvlink bridge is supported on R750XA, DSS8440 and T550, A30 w/NvLink bridge is supported on R750XA and T550

DW - Double Wide, SW - Single Wide, FH- Full Height, FL - Full Length, HH - Half Height, HL - Half Length