PLATFORM	Partners															
														AMD		INTEL
	H100 80GB PCIe	A100 80GB PCIe	H100 SXM5	A100 40GB SXM4 (Nvlink)	A100 80GB SXM4 (Nvlink)	L40	L4	A40	A30	A16	A10	A2	T4	MI100	MI210	Max 1550 OAM (X4)
XE9680			Shipping		1H2023											
XE9640																1H2023
XE8640			1H2023													
R760	Shipping (2)	Shipping (2)						Shipping (2)	Shipping (2)	Shipping (2)		Shipping (6)				
R660												Shipping (2)				
R7625		Shipping (2)						Shipping (2)	Shipping (2)	Shipping (2)		Shipping (6)			Shipping (2)	
R7615		Shipping (3)						Shipping (3)	Shipping (3)	Shipping (3)		Shipping (6)			Shipping (3)	
R6625												Shipping (2)				
R6615												Shipping (2)				
C6620												Shipping (2)				
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[3])	Shipping (4[3])	Shipping (6[3])	Shipping (6[3])	Shipping (4[3])	Shipping (4[3])	
R750	Shipping (2)	Shipping (2)				Shipping (2)	Shipping (6)	Shipping (2)	Shipping (2)	Shipping (2)	Shipping (3)	Shipping (6)	Shipping (6)			
R650							Shipping (3)					Shipping (3)	Shipping (3)			
C6520												Shipping (1)	Shipping (1)			
R7525 - Rome & Milan	Shipping (3)	Shipping (3)				Shipping (3)	Shipping (6)	Shipping (3)	Shipping (3)	Shipping (3)	Shipping (3)	Shipping (6)	Shipping (6)	Shipping (3)	Shipping (3)	
R7515 - Rome & Milan									Shipping (1)	Shipping (1)		Shipping (4)	Shipping (4)		Shipping (1)	
R6525 - Rome & Milan												Shipping (3)	Shipping (3)			
R6515 - Rome & Milan												Shipping (2)	Shipping (1)			
C6525 - Rome & Milan												Shipping (1)	Shipping (1)			
XR4000									Shipping (1)			Shipping (2)				
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)														
R740/XD		Shipping (3)						Shipping (3)	Shipping (3)	Shipping (3)	Shipping (3)	Shipping (6)	Shipping (6**)			
R640												Shipping (3)	Shipping (3)			
T550								Shipping(2)	Shipping(2)			Shipping (5)	Shipping (5)			
XR2													Shipping (1)			

1 – XE8545, DSS8440 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

Brand	Model	GPU Memory	Memory ECC	Memory Bandwidth	Max Power Consumption	Graphic Bus/ System Interface	Interconnect Bandwidth	Slot Width	GPU Height/Length	Workload[1]
AMD	MI210	64 GB HBM2e	Y	1638 GB/sec	300W	PCIe Gen4x16/ Infinity Fabric Link bridge	64 GB/sec (PCIe 4.0)	DW	FHFL	HPC/Machine learning training
AMD	MI100	32 GB HBM2	Y	1228 GB/sec	300W	PCIe Gen4x16	64 GB/sec (PCIe 4.0)	DW	FHFL	HPC/Machine learning training
Intel	Max 1550	128 GB HBM		600W	X^eLink bridge		N/A	N/A	N/A	
Nvidia	A100	80 GB HBM2	Y	2039 GB/sec	500W	NVIDIA NVLink	600 GB/sec (3rd Gen NVLink)	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	HPC/AI/Database Analytics
Nvidia	H100	80 GB HBM2e	Y	2000 GB/sec	300-350W	PCIe Gen5x16/ NVLink bridge[8]	128 GB/sec[5] (PCIe 5.0)	DW	FHFL	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	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	mainstream AI
Nvidia	L40	48 GB GDDR6	Y	864 GB/sec	300W	PCIe Gen4x16	64 GB/sec[5] (PCIe 4.0)	DW	FHFL	Performance graphics/VDI
Nvidia	L4	24 GB GDDR6	Y	300 GB/s	72W	PCIe Gen4 x16	64 GB/sec (PCIe 4.0)	SW	HHHL	Inferencing/Edge/VDI
Nvidia	L4	24 GB GDDR6	Y	300 GB/s	72W	PCIe Gen4 x16	64 GB/sec (PCIe 4.0)	SW	FHHL	Inferencing/Edge/VDI
Nvidia	A40	48 GB GDDR6	Y	696 GB/sec	300W	PCIe Gen4x16/ NVLink bridge[8]	64 GB/sec[5] (PCIe 4.0)	DW	FHFL	Performance graphics/VDI
Nvidia	A16	64 GB GDDR6	Y	800 GB/sec	250W	PCIe Gen4x16	64 GB/sec (PCIe 4.0)	DW	FHFL	VDI
Nvidia	A2 (v2)	16 GB GDDR6	Y	200 GB/sec	60W	PCIe Gen 4x8	32 GB/sec (PCIe 4.0)	SW	HHHL	Inferencing/Edge/VDI
Nvidia	A2 (v2)	16 GB GDDR6	Y	200 GB/sec	60W	PCIe Gen 4x8	32 GB/sec (PCIe 4.0)	SW	FHHL	Inferencing/Edge/VDI
Nvidia	A10	24 GB GDDR6	Y	600 GB/sec	150W	PCIe Gen4x16	64 GB/sec (PCIe 4.0)	SW	FHFL	mainstream graphics/VDI
Nvidia	T4	16 GB GDDR6	Y	300 GB/sec	70W	PCIe Gen3x16	32 GB/sec (PCIe 3.0)	SW	HHHL	Inferencing/Edge/VDI
Nvidia	T4	16 GB GDDR6	Y	300 GB/sec	70W	PCIe Gen3x16	32 GB/sec (PCIe 3.0)	SW	FHHL	Inferencing/Edge/VDI

[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