Brand	l 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	Part Numbers & SKL End of Marketing Life	14G/15G Server Sup	16G Server Support GPU Perf numbers	NOTE: Copy and paste links to browser. Clicking directly from XLS does not work consistently!
AMD	MI300X OAM	192 GB HBM3	Υ	5.3 TB/sec	750W	AMD Infinity Fabric Links	896 GB/sec	N/A	N/A	N/A	AI / HPC				https://www.amd.com/en/graphics/instinct-server-accelerators
AMD	MI210	64 GB HBM2e	Υ	1638 GB/sec	300W	PCIe Gen4x16/ Infinity Fabric Link bridge ⁸	64 GB/sec (PCIe 4.0)	DW	FHFL	CPU 8 pin	HPC/Machine learning training				https://www.amd.com/en/products/server-accelerators/amd-instinct-mi210
Intel	Max1550 OAM	128 GB HBM2e	Υ	3276.8 GB/sec	600W	Intel XeLink	-	N/A	N/A	N/A	AI / HPC				https://www.intel.com/content/www/us/en/products/sku/232873/intel-data-center-gpu-max-1550/support.html
Intel	Max1100	48 GB HBM2e	Y	1228.8 GB/sec	300W	PCIe Gen5x16/ XeLink bridge ⁸	128 GB/sec ⁵ (PCIe 5.0)	DW	FHFL	PCIe 16 pin	AI / HPC	_	41		https://www.intel.com/content/www/us/en/products/sku/232876/intel-data-center-gpu-max-1100/specifications.html
Intel Nvidia	Flex 140 H100 SXM5 (x8)	12 GB GDDR6 80 GB HBM3	Y	336 GB/Sec 3 TB/sec	75W 700W	PCIe Gen4 x8 NVIDIA NVLink	32 GB/sec (PCle 4.0) 900 GB/sec	SW N/A	HHHL/FHHL N/A	N/A N/A	Inferencing/Edge AI / HPC	_			https://www.intel.com/content/www/us/en/products/sku/230020/intel-data-center-gpu-flex-140/specifications.html
Nvidia	H100 SXM5 (x8)	80 GB HBM3 80 GB HBM3	Y	3 TB/sec 3 TB/sec	700W	NVIDIA NVLINK NVIDIA NVLINK	900 GB/sec 900 GB/sec	N/A N/A	N/A N/A	N/A N/A	AI / HPC	-	- 1		https://www.nvidia.com/en-us/data-center/h100/ https://www.nvidia.com/en-us/data-center/h100/
Nvidia	A100 SXM4 (x8)	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	—	1		https://www.nvidia.com/content/dam/en-zz/Solutions/Data-Center/a100/pdf/a100-80gb-datasheet-update-nvidia-us-1521051-
Nvidia	A100 SXM4 (x4)	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				https://www.nyidia.com/content/dam/en-zz/Solutions/Data-Center/4100/odf/a100-80gb-datasheet-update-nyidia-us-1521051-
Nvidia	A100 SXM4 (x4)	40 GB HBM2	Υ	1555 GB/sec	400W	NVIDIA NVLink	600 GB/sec (3rd Gen NVLink)	N/A	N/A	N/A	HPC/AI/Database Analytics				https://www.nvidia.com/content/dam/en-zz/Solutions/Data-Center/a100/pdf/a100-80gb-datasheet-update-nvidia-us-1521051-
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	PCIe 16 pin	HPC/AI/Database Analytics				https://www.nvidia.com/en-us/data-center/h100/
Nvidia	A100	80 GB HBM2e	Υ	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				https://www.nvidia.com/content/dam/en-zz/Solutions/Data-Center/a100/pdf/nvidia-a100-datasheet-us-nvidia-1758950-r4-
Nvidia	L40S	48 GB GDDR6	Y	864 GB/sec	350W	PCIe Gen4 x16	64 GB/sec ⁵ (PCIe 4.0)	DW	FHFL	PCIe 16 pin	Al/Performance graphics/VDI				https://www.nvidia.com/en-us/data-center/l40s/
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 Al				https://www.nvidia.com/content/dam/en-zz/Solutions/data-center/products/a30-gpu/pdf/a30-datasheet.pdf
Nvidia	L40	48 GB GDDR6	Y	864 GB/sec	300W	PCIe Gen4 x16	64 GB/sec (PCIe 4.0)	DW	FHFL	PCIe 16 pin	Performance graphics/VDI				https://www.nvidia.com/en-us/data-center/l40/
Nvidia	A40	48 GB GDDR6	Υ	696 GB/sec	300W	PCIe Gen4x16/ NVLink bridge ⁸	64 GB/sec ⁵ (PCIe 4.0)	DW	FHFL	CPU 8 pin	Performance graphics/VDI				https://www.nvidia.com/content/dam/en-zz/Solutions/Data-Center/a40/proviz-print-nvidia-a40-datasheet-us-nvidia-1469711-r8-
Nvidia	A16	64 GB GDDR6	Y	800 GB/sec	250W	PCIe Gen4 x16	64 GB/sec (PCIe 4.0)	DW	FHFL	CPU 8 pin	VDI				https://www.nvidia.com/en-us/data-center/products/a16-gpu/
Nvidia	L4	24 GB GDDR6	Υ	300 GB/s	72W	PCIe Gen4 x16	64 GB/sec (PCIe 4.0)	sw	HHHL	N/A	Inferencing/Edge/VDI		-		https://www.nvidia.com/L4
Nvidia	L4	24 GB GDDR6	Υ	300 GB/s	72W	PCIe Gen4 x16	64 GB/sec (PCIe 4.0)	SW	FHHL	N/A	Inferencing/Edge/VDI				https://www.nyidia.com/t4
Nvidia	A2 (v2)	16 GB GDDR6	Υ	200 GB/sec	60W	PCIe Gen4 x8	32 GB/sec (PCIe 4.0)	sw	HHHL	N/A	Inferencing/Edge/VDI		TI		https://www.nvidia.com/en-us/data-center/products/a2/
Nvidia	A2 (v2)	16 GB GDDR6	Υ	200 GB/sec	60W	PCIe Gen4 x8	32 GB/sec (PCIe 4.0)	sw	FHHL	N/A	Inferencing/Edge/VDI		T		https://www.nvidia.com/en-us/data-center/products/a2/
Nvidia	A10	24 GB GDDR6	Υ	600 GB/sec	150W	PCIe Gen4 x16	64 GB/sec (PCIe 4.0)	sw	FHFL	PCIe 8 pin	mainstream graphics/VDI				https://www.nvidia.com/content/dam/en-zz/Solutions/Data-Center/a10/pdf/datasheet-new/nvidia-a10-datasheet.pdf
Nvidia	T4	16 GB GDDR6	Υ	300 GB/sec	70W	PCIe Gen3 x16	32 GB/sec (PCIe 3.0)	sw	HHHL	N/A	Inferencing/Edge/VDI				https://www.nvidia.com/content/dam/en-zz/Solutions/Data-Center/tesla-t4/t4-tensor-core-datasheet-951643.pdf
Nvidia	T4	16 GB GDDR6	Υ	300 GB/sec	70W	PCIe Gen3 x16	32 GB/sec (PCIe 3.0)	sw	FHHL	N/A	Inferencing/Edge/VDI				https://www.nvidia.com/content/dam/en-zz/Solutions/Data-Center/tesla-t4/t4-tensor-core-datasheet-951643.pdf
					1						35-1				mass of the state
1	and the demonstrate of														

^{*}suggested ideal workloads, but can be used for other workloads

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

³upto 100GB/sec when RTX NVLink bridge is used, RTX NVLink bridge is only supported on T640 ⁴Structural Sparsity enabled

⁶upto 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, upto 400GB/sec for A800 when NVLink bridge is used

Option Dockstease. In a 70 of all Trib Winter In Virtal Ringles is seed, upon 2004cested to 14 of 2005cested on An Owner In Virtal Ringles is used.

*Refer to MaarGPUs on supported platforms safe for feetal supported on R750 x Max 110 of 2005cested on R750 x Max 120 of 2005cested on R7

	Partners																					
																	AMD				Intel	_
PLATFORM	H100 80GB PCle	A100 80GB PCIe	H100 SXM5 (X8)	A100 SXM4 (X8)	H100 SXM5 (X4)) A100 40GB SXM4 (X4)	A100 80GB SXM4 (X4)	L40S	L40	L4	A40	A30	A16	A10	A2	T4	MI100 (EOML)	MI210	MI300X OAM (X8)	Flex 140	Max 1100	Max 1550 OAM (X4)
XE9680			Shipping	Shipping															Ready to Quote			
XE9640					Shipping																	Shipping
XE8640					Shipping																	
R760XA	Shipping (4 ³)	Shipping (4 ³)						Shipping (4 ³)	Shipping (4 ³)	Shipping (8 ³)	Shipping (4 ³)	Shipping (4 ³)	Shipping (4 ³)		Shipping (12 ³)			Shipping (4 ³)		Shipping (10 ³)*	Shipping (4 ³)	l
R760	Shipping (2)	Shipping (2)						Shipping (2)	Shipping (2)	Shipping (4)	Shipping (2)	Shipping (2)	Shipping (2)		Shipping (6)					Shipping (6)*	Shipping (2)	
R760xs															Shipping (2)							
R760xd2										Shipping (2)		Shipping (1)			Shipping (2)							
R660										Shipping (3)					Shipping (2)					Shipping (2)*		
R7625	Shipping (2)	Shipping (2)						Shipping (2)	Shipping (2)	Shipping (4)	Shipping (2)	Shipping (2)	Shipping (2)		Shipping (6)			Shipping (2)				
R7615	Shipping (2)	Shipping (3)						Shipping (3)	Shipping (3)	Shipping (4)	Shipping (3)	Shipping (3)	Shipping (3)		Shipping (6)		İ	Shipping (3)				
R6625										Shipping (2)					Shipping (2)		İ					
R6615										Shipping (2)					Shipping (2)		İ					
R960		Shipping (4)											Shipping (4)									
T560									Shipping (2)	Shipping (5)		Shipping (2)			Shipping (6)							
C6620															Shipping (2)							
XR7620		Shipping (2)								Shipping (5)		Shipping (2)			Shipping (5)						Shipping (2)	
XR5610										Shipping (2)					Shipping (2)							
XR8000										Shipping (2)												
HS5620															Shipping (2)							
XE8545						Shipping (4 ¹)	Shipping (4 ¹)								- 11 317							1
R750xa	Shipping (4 ³)	Shipping (4 ³)						Shipping (4 ³)	Shipping (4 ³)	Shipping (6 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)						11 0 7	Shipping (2)	Shipping (6)	Shipping (2)	Shipping (2)	Shipping (2)	Shipping (3)	Shipping (6)	Shipping (6)	- 11 5 7	- 11 0 7		Shipping (6)		
R650										Shipping (3)					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	11 300	111 345							- 11 507	Shipping (4)	- 11 507	Shipping (1)	Shipping (1)	- 11 500	Shipping (4)	Shipping (4)	- 11 517	Shipping (1)				1
R6525 - Rome & Milan										11 317		- 11 5(7	- 11 5(7		Shipping (3)	Shipping (3)		- 11 3 17				
R6515 - Rome & Milan															Shipping (2)	Shipping (1)						1
C6525 - Rome & Milan															Shipping (1)	Shipping (1)						
XR4520C										Shipping (2)		Shipping (1)			Shipping (2)							
XR12		Shipping (2)								Shipping (3)	Shipping (2)	Shipping (2)			Shipping (2)	Shipping (2)				Shipping (2)		
XR11		Onipping (E)								Shipping (3)	Onipping (E)	Chipping (E)			Shipping (2)	Shipping (2)				Shipping (3)		
DSS8440		non-the-								Snipping (0)	Shipping	Shipping			Simpping (2)	Shipping				Snipping (0)		
		Shipping (4/8/10 ³)		1	ļ						(4/8/10 ¹)	(4/8/10 ¹)			1	(8/12/16 ¹)			-			├
R940XA		Shipping (4)		-		-								ļ					-			↓
R740/XD		Shipping (3)									Shipping (3)	Shipping (3)	Shipping (3)	Shipping (3)	Shipping (6)	Shipping (6**)			ļ			↓
R640					ļ										Shipping (3)	Shipping (3)			ļ			ļ
T550					ļ						Shipping(2)	Shipping(2)			Shipping (5)	Shipping(5)			ļ			
XR2										e: Internal Use On						Shipping (1)						<u> </u>

1 – XE8545, DSS8440 are set configs
2 – subject to change
3 - R760XA, R750XA at a minimum require 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
Currently available as Customer install only
** 3 FI, 3 additional T4s can be added through Customer Install