DELL Technologies

Data Sheet

PowerEdge Server GPU Matrix

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	MI300X OAM	192 GB HBM3	Y	5.3 TB/sec	750W	AMD Infinity Fabric Links	896 GB/sec	N/A	N/A	N/A	AI / HPC
AMD	MI210	64 GB HBM2e	Y	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
Intel	Gaudi3 OAM	128GB HBM3	Y	3.6TB/s	850W	PCIe Gen 5x16	-	N/A	N/A	N/A	AI / HPC
Intel	Flex 140	12 GB GDDR6	Y	336 GB/Sec	75W	PCIe Gen4 x8	32 GB/sec (PCIe 4.0)	SW	HHHL/ FHHL	N/A	Inferencing/Edge
Nvidia	H100 NVL	94 GB HBM3	Y	3.9 TB/s	350W- 400W	PCIe Gen 5x16	600GB/s (PCIe 5.0)	DW	FHFL	PCIe 16 pin	AI / HPC
Nvidia	H200 SXM5 (x8)	141GB HBM3e	Y	4.8 TB/s	700W	NVIDIA NVLink	900 GB/sec	N/A	N/A	N/A	AI / HPC
Nvidia	H100 SXM5 (x8)	80 GB HBM3	Y	3 TB/sec	700W	NVIDIA NVLink	900 GB/sec	N/A	N/A	N/A	AI / HPC
Nvidia	H100 SXM5 (x4)	80 GB HBM3	Y	3 TB/sec	700W	NVIDIA NVLink	900 GB/sec	N/A	N/A	N/A	AI / HPC
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	AI/Performance graphics/VDI
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	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
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 Gen4 x16	64 GB/sec (PCIe 4.0)	DW	FHFL	CPU 8 pin	VDI
Nvidia	L4	24 GB GDDR6	Y	300 GB/s	72W	PCIe Gen4 x16	64 GB/sec (PCIe 4.0)	SW	HHHL	N/A	Inferencing/Edge/ VDI
Nvidia	L4	24 GB GDDR6	Y	300 GB/s	72W	PCIe Gen4 x16	64 GB/sec (PCIe 4.0)	SW	FHHL	N/A	Inferencing/Edge/ VDI
Nvidia	A2 (v2)	16 GB GDDR6	Y	200 GB/sec	60W	PCIe Gen4 x8	32 GB/sec (PCIe 4.0)	SW	HHHL	N/A	Inferencing/Edge/ VDI
Nvidia	A2 (v2)	16 GB GDDR6	Y	200 GB/sec	60W	PCIe Gen4 x8	32 GB/sec (PCIe 4.0)	SW	FHHL	N/A	Inferencing/Edge/ VDI
Nvidia	A10	24 GB GDDR6	Y	600 GB/sec	150W	PCIe Gen4 x16	64 GB/sec (PCIe 4.0)	SW	FHFL	PCIe 8 pin	mainstream graphics/VDI
Nvidia	T4	16 GB GDDR6	Y	300 GB/sec	70W	PCIe Gen3 x16	32 GB/sec (PCIe 3.0)	SW	HHHL	N/A	Inferencing/Edge/ VDI
Nvidia	T4	16 GB GDDR6	Y	300 GB/sec	70W	PCIe Gen3 x16	32 GB/sec (PCIe 3.0)	SW	FHHL	N/A	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 and H100 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 300GB/sec for MI210 when Infinity Fabric Link bridge 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. 8A100 w/Nvlink bridge is supported on R760XA, R750XA and DSS8440; A40 w/Nvlink bridge is supported on R760XA, R750XA, DSS8440 and T550; A30 w/NVLink bridge is supported on R760XA, R750XA, DSS8440 and T550; ; MI210 w/Infinity Fabric Link bridge is supported on R760XA and R750XA; H100 w/Nvlink bridge is supported on R760XA and R750XA; Max1100 w/XeLink bridge is supported on R760XA
9. DW - Double Wide, SW - Single Wide, FH- Full Height, FL - Full Length, HH - Half Height, HL - Half Length

Brand	Model	Part Numbers	Tied SKU[2]					Customer Kit SKU[2]	
AMD	MI300X OAM	W9DCT	490-BJZD						
AMD	MI210	0RJ6Y/YJFV0/880VN	490-BHUU	490-BJYD				490-BHUT	490-BJYF
Intel	Gaudi3 OAM	3PNFV							
Intel	Flex 140	P7M0V/DHG15	490-BIIQ	490-BIIM				490-BIIK	490-BIIO
Nvidia	H100 NVL	MM3MP	330-BCFX						
Nvidia	H200 SXM5 (x8)	P1Y02	490-BKHF						
Nvidia	H100 SXM5 (x8)	W2YC4	489-BBGD						
Nvidia	H100 SXM5 (x4)	V0W6P	489-BBGJ						
Nvidia	L40S	7WK28	490-BJVT					490-BJVV	
Nvidia	A30	W3C1G/RV08M	490-BGUT	490-BGWJ	490-BGYE			490-BGUS	490-BHQK
Nvidia	L40	VHNWT/20FCY	490-BJHP	490-BJHQ				490-BJHN	
Nvidia	A40	M9WT5/W6GX6	490-BGKG	490-BGSJ				490-BGKE	
Nvidia	A16	8F04V/D1P1T	490-BHER	490-BHET				490-BHES	
Nvidia	L4	V9XT2	490-BJRN					490-BJRR	
Nvidia	L4	NG3PY	490-BJRP					490-BJRQ	
Nvidia	A2 (v2)	H98K0	490-BHVO					490-BHVQ	
Nvidia	A2 (v2)	MG5JX	490-BHVR					490-BHVP	
Nvidia	A10	CGNVH/6KNWW	490-BGKF	490-BGLB	490-BGYD			490-BGKH	
Nvidia	T4	7PH29	490-BFHN					490-BFLB	
Nvidia	T4	HCKJM	490-BEYL					490-BEYM	

		14G/15G Server Support																				16G Server Support																					
Brand	Model	End of Marketing Life	XE8545	R750XA	R750	R650	C6520	R640	R740	R740XD	R940XA	R6515	R6525	R75157	R75257	T550	C6525	DSS8440	XR4520C	XR12	XR11	XR2	XE9680	XE9640	XE8640	R760XA	R760	R760xs	R760XD2	R660	R7625	R7615	R6625	R6615	R960	T560	C6620	XR8620t	XR7620	XR5610	HS5620	GPU Perf numbers	Spec Sheets
AMD	MI300X OAM		X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	✓	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X		
AMD	MI210		X	✓	X	X	X	X	X	X	X	X	X	✓	✓	X	X	X	X	X	X	X	X	X	X	✓	X	X	X	X	✓	✓	X	X	X	X	X	X	X	X	X		
Intel	Gaudi3 OAM																						✓																					
Intel	Flex 140		X	X	✓	✓	X	X	X	X	X	X	X	X	X	X	X	X	X	X	✓	✓	X	X	X	X	✓	✓	X	X	✓	X	X	X	X	X	X	X	X	X	X	X		
Nvidia	H100 NVL		X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X				✓																	
Nvidia	H200 SXM5 (x8)		X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	✓																					
Nvidia	H100 SXM5 (x8)		X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	✓	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X		
Nvidia	H100 SXM5 (x4)		X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	✓	✓	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X		
Nvidia	L40S		X	✓	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	✓	✓	X	X	X	✓	✓	X	X	X	X	X	X	X	X	X	X		
Nvidia	A30		X	✓	✓	X	X	X	✓	✓	X	X	X	✓	✓	✓	X	✓	✓	✓	X	X	X	X	X	✓	✓	X	✓	X	✓	✓	X	X	X	✓	X	X	X	X	X		
Nvidia	L40		X	✓	✓	X	X	X	X	X	X	X	X	X	✓	X	X	X	X	X	X	X	X	X	X	✓	✓	X	X	X	✓	✓	X	X	X	✓	X	X	X	X	X		
Nvidia	A40		X	✓	✓	X	X	X	✓	✓	X	X	X	X	✓	✓	X	✓	X	✓	X	X	X	X	X	✓	✓	X	X	X	✓	✓	X	X	X	X	X	X	X	X	X		
Nvidia	A16		X	✓	✓	X	X	X	✓	✓	X	X	X	✓	✓	X	X	X	X	X	X	X	X	X	X	✓	✓	X	X	X	✓	✓	X	X	✓	X	X	X	X	X	X		
Nvidia	L4		X	✓	✓	✓	X	X	X	X	X	X	X	✓	✓	X	X	X	X	✓	✓	X	X	X	X	X	X	X	X	✓	X	X	X	X	X	X	X	X	✓	X	X		
Nvidia	L4		X	✓	✓	✓	X	X	X	X	X	X	X	✓	✓	X	X	X	✓	✓	✓	X	X	X	X	✓	✓	X	✓	✓	✓	✓	✓	✓	X	✓	X	✓	✓	✓	X		
Nvidia	A2 (v2)		X	✓	✓	✓	✓	✓	X	X	X	✓	✓	✓	✓	X	✓	X	X	X	✓	X	X	X	X	X	X	✓	X	✓	X	✓	X	X	X	X	X	✓	X	✓	X	✓		
Nvidia	A2 (v2)		X	✓	✓	✓	X	✓	✓	✓	X	X	✓	✓	✓	✓	X	X	✓	✓	✓	X	X	X	X	✓	✓	X	✓	✓	✓	✓	✓	✓	X	✓	X	X	✓	✓	X		
Nvidia	A10		X	✓	✓	X	X	X	✓	✓	X	X	X	X	✓	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X		
Nvidia	T4		X	✓	✓	✓	✓	✓	X	X	X	✓	✓	✓	✓	X	✓	X	X	X	X	✓	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X		
Nvidia	T4		X	✓	✓	✓	X	✓	✓	✓	X	X	✓	✓	✓	✓	X	✓	X	✓	✓	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X		

Server Compability Legend: ✓ = Supported | X = Not supported

PLATFORM	NVIDIA												
	H100 NVL	H200 SXM5 (x8)	H100 SXM5 (X8)	H100 SXM5 (X4)	L40S	L40	L4	A40	A30	A16	A10	A2	T4
XE9680		RTQ RTS 7/17/24[2]	Shipping										
XE9640				Shipping									
XE8640				Shipping									
R760XA	RTQ (4[3]) RTS 7/10/24[2]				Shipping (4[3])	Shipping (4[3])	Shipping (8[3])	Shipping (4[3])	Shipping (4[3])	Shipping (4[3])		Shipping (12[3])	
R760	Sep FY25[2] (2)				Shipping (2)	Shipping (2)	Shipping (4)	Shipping (2)	Shipping (2)	Shipping (2)		Shipping (6)	
R760xs												Shipping (2)	
R760xd2							Shipping (2)		Shipping (1)			Shipping (2)	
R660							Shipping (3)					Shipping (2)	
R7625	Sep FY25[2] (2)				Shipping (2)	Shipping (2)	Shipping (4)	Shipping (2)	Shipping (2)	Shipping (2)		Shipping (6)	
R7615	Sep FY25[2] (3)				Shipping (3)	Shipping (3)	Shipping (4)	Shipping (3)	Shipping (3)	Shipping (3)		Shipping (6)	
R6625							Shipping (3)					Shipping (2)	
R6615							Shipping (2)					Shipping (2)	
R960										Shipping (4)			
T560						Shipping (2)	Shipping (5)		Shipping (2)			Shipping (6)	
R360												Shipping (1)	
T360												Shipping (1)	
C6620												Shipping (2)	
XR7620							Shipping (5)		Shipping (2)			Shipping (5)	
XR5610							Shipping (2)					Shipping (2)	
XR8620t							Shipping (3)						
HS5620												Shipping (2)	
R750						Shipping (2)	Shipping (6)	Shipping (2)	Shipping (2)	Shipping (2)	Shipping (3)	Shipping (6)	Shipping (6)
R650							Shipping (3)					Shipping (3)	Shipping (3)
R7525 - Rome & Milan						Shipping (3)	Shipping (6)	Shipping (3)	Shipping (3)	Shipping (3)	Shipping (3)	Shipping (6)	Shipping (6)
R6525 - Rome & Milan												Shipping (3)	Shipping (3)
XR4520C							Shipping (2)		Shipping (1)			Shipping (2)	
XR12							Shipping (3)	Shipping (2)	Shipping (2)			Shipping (2)	Shipping (2)
XR11							Shipping (3)					Shipping (2)	Shipping (2)

PLATFORM	AMD		INTEL	
	MI210	MI300X OAM (X8)	Flex 140	Gaudi3 OAM
XE9680		Shipping		Q4
XE9640				
XE8640				
R760XA	Shipping (4[3])		Shipping (10[3])*	
R760			Shipping (6)*	
R760xs				
R760xd2				
R660			Shipping (3)*	
R7625	Shipping (2)			
R7615	Shipping (3)			
R6625				
R6615				
R960				
T560				
R360				
T360				
C6620				
XR7620				
XR5610				
XR8620t				
HS5620				
R750			Shipping (6)	
R650			Shipping (2)	
R7525 - Rome & Milan	Shipping (3)			
R6525 - Rome & Milan				
XR4520C				
XR12			Shipping (2)	
XR11			Shipping (3)	

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