

HPC & AI Update

Kim Tae Ryong | HPE Korea



Windows Server 2022: Be cloud ready with hybrid

Apollo의 새로운 이름, **HPE CRAY XD**

Bringing the legacy, power and innovation of high-performing computing technologies at any scale

Integrated HPE Cray HPC | AI Product Line **HPE Cray EX, HPE Cray XD 6500 and HPE Cray XD 2000**

“CRAY XD” – GEN11 and beyond

- Implementing leadership Cray Hardware, Software, Fabric & Tools in standard rack-mount server systems
- Model naming/numbering methodology consistent with current Apollo products

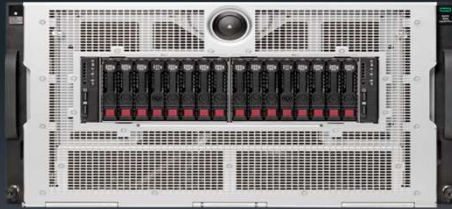
“APOLLO” – thru GEN10 PLUS

- No name change to existing HPE Apollo products

Apollo의 새로운 이름, HPE CRAY XD

CY 2022

Apollo 6500 Gen10 Plus



- AMD® EPYC 2P for up to 16 accelerators
- AMD® EPYC 1P for up to 8 accelerators
- GPU Accelerators from NVIDIA® and AMD®

CY 2023

HPE CRAY XD 6500

Models tailored for HPC, AI, Mixed Workloads

XD665



X4 GPU - Standard x86 Architecture

- 2x AMD Genoa CPU + 4x NVIDIA H100 GPU
- Mixed HPC & AI workloads

XD670



X8 GPU - Standard x86 Architecture

- 2P Intel Sapphire Rapids + 8x NVIDIA H100 GPU
- AI Training workloads

XD685



Leadership Cray Technology

- 4x AMD MI300A APU or 4x MI300C CPU
- Full-Spectrum HPC & AI with leadership performance
- Heterogeneous clusters to address any mix of GPU or CPU workloads

Accelerators FOR Gen 10

COMPUTE



DL Training



Scientific Research



Data Analytics

**Fastest Compute,
FP64 up to 7 MIG
instances**

A100

300W | 80GB
2-Slot FHFL
NVLink



Language Processing



Conversational AI



Recommender Systems

**Versatile
Mainstream
Compute FP64, Up
to 4 MIG instances**

A30

165W | 24GB
2-Slot FHFL
NVLink

COMPUTE & GRAPHICS



Edge AI & Small Inference



Edge Video



Mobile Cloud Gaming

**Entry-level inference
Video & Graphics
Compact & Versatile**

A2

40-60W |
16GB
1-Slot Low
Profile

T4

70W
16GB
1-Slot Low
Profile

GRAPHICS



4K Cloud Gaming

Virtual Workstation



AI acceleration

**Mainstream
Graphics and Video
with AI**

A10

150W | 24GB
1 Slot FHFL



Cloud Rendering



Cloud XR and vWS



3D Design
Collaboration

**Fastest RT Graphics
Largest render
models**

A40

300W | 48GB
2-Slot FHFL
NVLink | 3x DP



Virtual Desktop



Virtual Workstation



Transcoding

**4K Resolution
Max # of encode/
decode streams**

A16

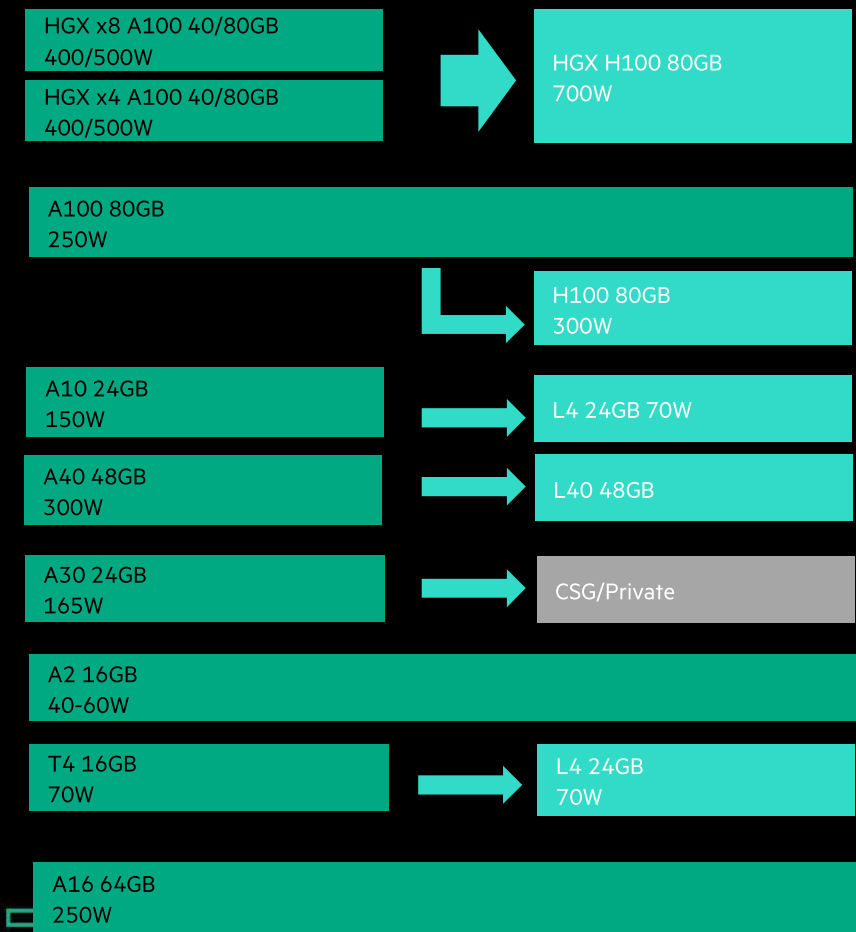
250W | 4x 16GB
2-Slot FHFL

Hewlett Packard
Enterprise



Windows Server 2022: Be cloud ready with hybrid

Accelerators FOR Gen 11





Hewlett Packard
Enterprise



Windows Server 2022: Be cloud ready with hybrid

GPU 지원 현황

Accelerator Model	용도	HPE PN:	ProLiant Server	Card Count	Date
A16 non-CEC	VDI	R8T26C	DL380 Gen10	verifying	5-Jul-22
			DL380 Gen10 Plus	3	
			SY480 Gen10 Plus E2	2	
		R8T26C	DL385 Gen10 Plus v2	verifying	1-Aug-22
A10 non-CEC	Mainstream Graphic	R9W59C	DL380 Gen10 Plus	5	15-Aug-22
			DL345 Gen10 Plus	3	3-Oct-22
			SY480 Gen10 Plus	4	3-Oct-22
		R9W59A	XL645d Gen10 Plus	8	15-Aug-22
			XL675d Gen10 Plus	16	
A30	Light Deep Learning Mainstream Graphic	R7G39C	DL380 Gen10	2	단종
DL385 Gen10 Plus v2			3		
A30 non-CEC		R9S38C	DL380 Gen10	2	15-Aug-22
			DL380 Gen10 Plus	3	
			DL385 Gen10 Plus v2	3	
A40	High end Graphic	R7E31C	DL380 Gen10	2	단종
R9S37C		DL380 Gen10	2	16-May-22	
		DL380 Gen10 Plus	3		
		DL385 Gen10 Plus v2	3		
		DL580 Gen10	4		
		SY480 Gen10 E2	2		
		SY480 Gen10 Plus E2	2		
R9S37A		XL645d Gen10 Plus	4	16-May-22	
		XL675d Gen10 Plus	10		
					



Accelerator Model	용도	HPE PN:	ProLiant Server	Card Count	Date
A100-80GB PCIe non-CEC	DL Training / Inferencing Data Analysis	R9P49C	DL380 Gen10 Plus	2	16-May-22
			DL385 Gen10 Plus v2	3	
			DL580 Gen10	4	
		R9P49A	XL290n Gen10 Plus	2	
			XL645d Gen10 Plus	4	
			XL675d Gen10 Plus	10	
A2 non-CEC	Entry Inferencing 저전력, Edge	R9H23C	DL360 Gen10	1	20-Jun-22
			DL360 Gen10 Plus	2	
			DL380 Gen10	8	
			DL380 Gen10 Plus	8	
			DL385 Gen10 Plus v2	8	
			SY480 Gen10 E4	4	
			SY480 Gen10 Plus H	8	
			SY480 Gen10 Plus P	1	
			XL290n Gen10 Plus	4	1-Aug-22
			ML350 Gen10	4	
			DL325 Gen10 Plus v2	2	3-Oct-22
H100 PCIe	DL Training / Inferencing Data Analysis	R9S41A	XL645d Gen10 Plus	4	3-Dec-22
			XL675d Gen10 Plus	10	
		R9S41C	DL380 Gen10 Plus	2	6-Feb-23
MI210 PCIe	HPC & AI	R6V51A	DL380 Gen10 Plus	3	1-Aug-22
			XL645d Gen10 Plus	4	4-Apr-22
			XL675d Gen10 Plus	10	

Windows Server 2022: Be cloud ready with hybrid

A100 40GB PCIe (R6B53C/R6B53A) 단종 | A100 40GB 8 GPU HGX (R3V64A/R4W71A) 6월 30일 단종

THANK YOU

Contact e-mail: ryong@hpe.com



Windows Server 2022: Be cloud ready with hybrid