Specification	NV A10	NV A40	NV A6000	NV L4	NV L20	NV L40	NV L40S	NV 6000 Ada	ILV MR100	MTX C500 PCIe
Scenario	AI Reasoning AI Fine-tuning	AI Reasoning AI Fine-tuning	AI Reasoning AI Fine-tuning	AI Reasoning AI Fine-tuning	AI Reasoning AI Fine-tuning	AI Reasoning AI Fine-tuning	AI Reasoning AI Fine-tuning	AI Reasoning AI Fine-tuning	AI Reasoning AI Fine-tuning ***0.8x A40; 0.4x L40***	HPC - Science... **AI Training** ***0.85~0.9x A100***
Vendor	Nvidia, USA	Nvidia, USA	Nvidia, USA	Nvidia, USA	Nvidia, USA	Nvidia, USA	Nvidia, USA	Nvidia, USA	Iluvatar, China	MetaX, China
Architecture	Ampere	Ampere	Ampere	Ada Lovelace	Ada Lovelace	Ada Lovelace	Ada Lovelace	Ada Lovelace	Data N/A	Data N/A
CUDA Cores	9,216	10,752	10,752	7,776	11,776	18,176	18,176	18,176		
Tensor Cores	288	336	336	72	0	568	568	568	Data N/A CUDA-Compatible	Data N/A CUDA-Compatible
Ray Trace Cores	72	84	84	0	0	142	142	142		
GPU Memory	24 GB GDDR6	48 GB GDDR6	48 GB GDDR6	24 GB GDDR6	48 GB GDDR6	48 GB GDDR6	48 GB GDDR6	48 GB GDDR6	32 GB HBM2e	64 GB HBM2e
Memory Bandwidth	600 GB/s	696 GB/s	768 GB/s	300 GB/s	864 GB/s	864 GB/s	864 GB/s	960 GB/s	800 GB/s	1.8 TB/s
Peak FP16	62.3 TFLOPs	74.1 TFLOPs	78.2 TFLOPs	60.6 TFLOPs	119.5 TFLOPs	180.9 TFLOPs	182.4 TFLOPs	182.6 TFLOPs	96 TFLOPs	240 TFLOPs
Peak INT8	124.6 TOPs	148.1 TOPs	148.6 TOPs	122.4 TOPs	239 TOPs	360.0 TOPs	362.0 TOPs	360.0 TOPs	192 TOPs	480 TOPs
Power Consumption	150 W	300 W	300 W	72 W	275 W	300 W	350 W	300 W	150 W	350 W
PCIe Interface	PCIe 4.0 x16	PCIe 4.0 x16	PCIe 4.0 x16	PCIe 4.0 x16	PCIe 4.0 x16	PCIe 4.0 x16	PCIe 4.0 x16	PCIe 4.0 x16	PCIe 4.0 x16	PCIe 5.0 x16
Form Factor	Full-height Full-length Single Slot	Full-height Full-length Single Slot	Full-height Full-length Dual Slot	Half-height Half-length Single Slot	Full-height Full-length Dual Slot	Full-height Full-length Dual Slot	Full-height Full-length Dual Slot	Full-height Full-length Dual Slot	Full-height Full-length Single Slot	Full-height Full-length Dual Slot
Released Date	22-Nov	21-Mar	20-Dec	23-Nov	23-Nov	23-Jan	23-Nov	22-Dec	23	23
List Price US$	US$8,956	US$14,289	US$8,975	US$3,633	US$5,999	US$8,999	US$9,999	US$6,800	**Morris** morris@ici-cn.com +86 13901209254 (WA)	
Ref Price 24-Sep	US$3,500	US$8,000	US$6,500	US$2,800	US$5,500	US$7,000	US$9,500	US$7,900		

ICI can help you in below scenarios:

- Supply of popular AI hardware, software products or systems are unstable.
- You need to integrate with the CUDA architecture at lowest cost and in shortest time.
- AI content generation processes must be deployed within a private environment.
- Legal regulations prohibit the upload of data to public cloud services.
- Advanced applications such as computer graphics, genomics, molecular biology, and climate simulation, etc., require advanced mathematical and physical libraries, cannot rely on specialized resources like Google TPU cloud.
- Cost of cloud services, such as AWS or Nvidia DGX Cloud, becomes prohibitively high.

A100/H100/H200/DGX..., and more vendors/models are also available.
Please inquire ICI for more technical or commercial details.

产品代号	羲云®C500 PCIe5.0x16
算力	TF32: 120 TFLOPS FP16:240TFLOPS BF16:240 TFLOPS INT8:480 TOPS
内容规格	64GBHBM2e,带宽1.8TB/s
视频/JPEG解码	160路1080p@30FPS
视频/JPEG编码	12路1080p@30FPS
互联	MetaXLink 2卡4卡全互联
虚拟化示例	1/2/4/8
功耗	350W