
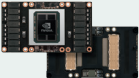




Choose the Right NVIDIA® Tesla® Solution for You

PRODUCT	DESIGNED FOR	BENEFITS	KEY FEATURES	RECOMMENDED SERVER CONFIGURATIONS
Tesla P100 PCIe 	HPC and Deep Learning	Replace 32 CPU servers with a single P100 server for HPC and deep learning	<ul style="list-style-type: none"> > 4.7 TeraFLOPS of double-precision performance > 9.3 TeraFLOPS of single-precision performance > 720 GB/s memory bandwidth (540 GB/s option available) > 16 GB of HBM2 memory (12 GB option available) 	2-4 GPUs per node
Tesla P100 with NVLink™ 	Deep Learning Training	10X faster deep learning training vs. last-gen GPUs	<ul style="list-style-type: none"> > 21 TeraFLOPS of half-precision performance > 11 TeraFLOPS of single-precision performance > 160 GB/s NVIDIA NVLink™ Interconnect > 720 GB/s memory bandwidth > 16 GB of HBM2 memory 	4-8 GPUs per node
Tesla P40 	Deep Learning Training and Inference	40X faster deep learning inference than a CPU server	<ul style="list-style-type: none"> > 47 TeraOPS of INT8 inference performance > 12 TeraFLOPS of single-precision performance > 24 GB of GDDR5 Memory > 1 decode and 2 encode video engines 	Up to 8 GPUs per node
Tesla P4 	Deep Learning Inference and Video Transcoding	40X higher energy efficiency than a CPU for inference	<ul style="list-style-type: none"> > 22 TeraOPS of INT8 inference performance > 5.5 TeraFLOPS of single-precision performance > 1 decode and 2 encode video engines > 50 W/75 W Power > Low profile form factor 	1-2 GPUs per node