



P90G4 - 4U GPGPU Optimized Rack-mount Server

Artificial Intelligence (AI) nowadays not only being an academic subject but also moving fast toward the real world when talking about facial recognition, robotics, revolutionary analytics, to diseases prevention, smart city constructions, etc. Among all the groundbreaking scientific progresses that call for accelerations in machine-learning (ML) and deep-learning (DL), the adoption of great amounts of GPGPUs will satisfy the thirst of tremendous computing power.

P90G4 is a purpose-built 4U complex GPGPU server, composed of a 1U server node based on latest Intel® Xeon® Processor Scalable family, and a 3U PCIe Expansion Box with capacity supporting up to eight state-of-the-art GPGPU cards. To meet with different workloads from various demands, P90G4 can support multiple topologies and bandwidths between the GPGPUs and CPUs by simple cable routing adjustments. Moreover, the support of Infiniband network allows it to be easily scaled out to multiple GPGPU clusters. The series presents,

Datacenter-ready Platform with Low Latency and High Bandwidth

P90G4 is designed to accelerate High Performance Computing (HPC) and AI applications or frameworks. With the dual socket head node based on Intel® Xeon® Processor Scalable family and the PCIe Gen3 expansion switch that supports up to eight GPGPUs, it delivers a dynamic balance of GPGPU and CPU ratios. Furthermore, to boost data training jobs, the efficiency of parallel computing is critical, the architecture of P90G4 with high-bandwidth PCIe interconnections would drive efficient peer-to-peer communication and minimized latency.

Highlights -

- Support latest GPGPU technologies, up to 8 Nvidia® Tesla® or AMD® Radeon Instinct™ GPGPUs
- 10GbE onboard + up to 100GbE standard PCIe networking option
- Flexible GPGPU utilization
- Hot-swappable fans, drives and power supply units
- 3+1 Power Redundancy


Framework Flexibility for Variant AI Applications

P90G4 system is capable to support both single and dual root complex. Take ML cases for example, dual root complex can arrange large size of tasks in CPUs, and small or distributed data training jobs among GPGPUs. The flexibility of P90G4 makes it an elastic AI platform for all popular DL/ML frameworks like TensorFlow, Caffe, etc.

Elastic System Built for HPC

GPGPU acceleration is also one of the solutions to optimize High Performance Computing (HPC) applications and remote virtualization considering the physical limits of large and hyperscale datacenters. Enabled by various configuration methods, P90G4 not only supports single and multi-head nodes, but also allow random choices of GPGPU numbers based on alternative workloads generated by various applications. P90G4 is a HPC-ready platform to increase cloud-scale flexibility and fungibility.

Design for Reliability and Serviceability

Designed for reduced OPEX and system reliability, most hardware structure of P90G4 is hot swappable, including front-access 4+1 redundant fan modules and 3+1 power supplies in rear. Furthermore, the server node can be detached from the front of 4U chassis, and the GPGPU cards can be easily installed after removal of the top cover. P90G4 promotes efficient serviceability whilst optimal performance is delivered. 



Model Name	P90G4
Positioning	HPC/GPGPU
Form Factor	4U1N rack mount
Processor	Dual Socket; Intel® Xeon® Processor Scalable Family
Memory Slot	24x DDR4 DIMM slot
Chipset	Intel Lewisburg C620 Series (C622)
GPGPU	8x GPGPU slot
Disk Drive Bay	Front Primary Bay: Option1: 8x 2.5" SAS/SATA hot-plug drive Option2: 4x 2.5" SAS/SATA + 4x NVMe hot-plug drive (disabling OCP card support)
Expansion Slot	2x PCIe Gen3 x16 (Low Profile) 1x PCIe Gen3 x16 (FHHL) 1x PCIe Gen3 OCP 2.0 A+B NIC mezz 1x Inventec Storage mezz
Network Controller	Onboard: Dual 10G SFP+ Supporting 25GbE OCP 2.0 mezz/100GbE standard PCIe card Inventec network OCP mezz card options: Option1: NIC-I599-10GD (Dual port 10Gb SFP+) Option2: NIC-I540-10GDC (Dual port 10Gb RJ-45)
Storage Controller	Onboard: 8x SATA 6Gb/s port Inventec storage mezz card options Option1: SAS3-3008-8i (12Gb/s) Option2: SAS3-3316-16i (12Gb/s, RAID)
System Management	IPMI 2.0 compliant + KVM with Dedicated LAN
TPM	2.0
Power Supply	Option1: 1600W Platinum 3+1 redundancy Option2: 2000W Platinum 3+1 redundancy
Fan	GPGPU zone: 5x 8056 fan Server zone: 8x 4056 fan Support hot-swap; N+1 redundancy

Inventec® at core



ABOUT INVENTEC

Inventec Enterprise Business Group (EBG) was established in 1998 and has been focusing on the design and manufacturing of server systems. Inventec EBG is the key server system supplier of the global branding clients.

Inventec Corporation (TAO)

No.88, Dazhi Rd., Taoyuan Dist.,

Taoyuan City 33068, Taiwan

Tel: 886-3-390-0000

Fax: 886-3-376-2370

Email: TAOproductsupport@inventec.com

Website: EBG.Inventec.com

