

ALPS-4800 Series

AI Training Platform

Features

- 8x PCIe x16 Gen3 GPGPU slots (300W/slot)
- Validated with NVIDIA® Tesla® P100/V100 accelerators
- Dual Intel® Xeon® Scalable processors
- 24x DDR4-2666 RDIMM
- 8x SATA 6Gb/s hot-swappable 2.5" drives
- 1x FHFL and 2x Low Profile PCIe x16 Gen3 add-on slots
- 1600W AC/DC Platinum PSU, 3+1 redundancy
- Separate airflow for CPU and GPU
- Flexible GPU/CPU ratio for DL/ML workloads



Introduction

AI Training Platform ALPS-4800 initiative enables operators and their customers to quickly build and enhance services with NVIDIA's GPU technologies. ADLINK's leadership in edge computing now includes NVIDIA's advanced GPU series of products in a stand-alone, scalable edge platform. ALPS-4800 provides endless possibilities for service innovation at the network edge.

Artificial Intelligence (AI) nowadays is not only an academic subject, but is moving fast towards the real world with applications in facial recognition, robotics, revolutionary analytics, disease prevention and smart city constructions. All the groundbreaking scientific progress calls for acceleration in machine-learning (ML) and deep-learning (DL) training, and the increasing adoption of GPUs will satisfy the thirst for tremendous computing power.

ALPS-4800 is a carrier-grade, multi-purpose platform designed for edge applications. Combining a server node, a PCIe expansion box with PCIe switching, the DEVKIT has a capacity to support a combination of up to eight NVIDIA GPUs depending on the needs of the application.

A configurable edge platform that supports different workloads and demands, the ALPS-4800 can support multiple topologies and bandwidths between GPUs and CPUs with simple cable routing adjustments. Moreover, Infiniband support allows it to be easily scaled up to multiple GPU clusters.

Framework Flexibility for Various AI Applications

The ALPS-4800 supports both single and dual root complexes for various AI applications. For deep learning applications, a single root complex can utilize all the GPU clusters to focus on large-size data training jobs and the use CPU to handle small tasks; For machine

learning applications, a dual root complex can allocate more tasks to the CPUs, and arrange fewer distributed data training jobs among GPUs. The flexible framework of the AI ALPS-4800 makes it an extremely flexible AI platform.

Enabling a switching option to configure specific PCIe lanes of GPU's to specific I/O and CPU cores enhances the ability to improve the overall flow of information to and from multiple virtualized applications. This provides the developer a broad range of options for configurability and manageability without the need to rack and stack systems, eating up valuable space, power and cooling.

Flexible System Built for Edge High Performance Computing (EHPC)

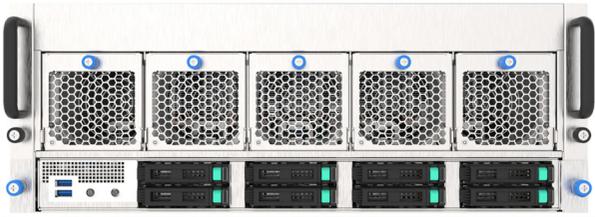
Moving high compute systems to the edge for GPU acceleration is critical for specific solutions needing to optimize high performance computing (HPC) applications and remote virtualization.

The ALPS-4800 is able to increase cloud-scale flexibility and agility at the edge.

It provides the flexibility to implement different head-nodes and the freedom to choose the numbers of GPUs per virtual machine (VM). It is an ideal hardware system that can support a wide variety of configurations via software implementation.

Rugged and Carrier-Grade for Reliability and Serviceability

Designed for reduced OPEX and system reliability, the hardware structure of the ALPS-4800 is all hot-swappable, front-accessible, has 4+1 redundant fan modules and redundant hot-swappable 3+1 power supplies. Easy access to the server node allows it to be removed from the front of the chassis, and GPU cards can be easily installed after removal of the top cover. The ALPS-4800 promotes efficient serviceability while delivering optimal performance.



Highlights

- Scalable**
 Supports the latest GPU technologies, up to 8 NVIDIA Tesla or 3rd party PCIe x16 acceleration cards
- Fast**
 10GbE onboard + OCP 2.0 to support 100Gb networking options
- Flexible GPU Configurations**
 Supports single & dual root complex design
- Hardened/Carrier Grade**
 Hot-swappable fans, drives and power supply units
- Packet Processing**
 Flow Processing – Packet Manager provides additional value for DPI, security and packet flow capabilities
- Broad Software Eco-System**
 Provides a platform supporting MEC, NFV/SDN and media applications using next generation GPUs, including tools for ease of development
- Compatibility**
 Wide range of OS, NFV/SDN/MEC middleware infrastructure, IPMI, and power management system

Model Name	ALPS-4800 Series
Form Factor	4U1N rack mount, 442mm(17.4") (W) x 174mm(6.8") (H) x 740mm(28.1") (D)
Processors	Dual Socket P0 (LGA-3647) Supports Intel® Xeon® Processor Scalable Family (up to 165W)
Memory	24x DDR4 DIMM Slot Up to 3TB 2666MHz DDR4 DIMM
Chipset	Intel® Lewisburg C622
GPGPU	8x PCIe x16 Gen3 GPU slots
Expansion Slots	2x PCIe Gen3 x16 (Low Profile) 1x PCIe Gen3 x16 (FHHL) 1x PCIe Gen3 OCP 2.0 A+B NIC mezz 1x Storage mezz
Storage	8x 2.5" SATA hot-swappable drives (Optional: SAS HBA/SAS RAID Storage Mezz Card)
System Management	AST2500 IPMI 2.0 and KVM with dedicated LAN
TPM	2.0
Power Supply	1600W AC/DC Platinum PSU, 3+1 redundancy Input: 1000W: 100-120 Vac / 12A / 50-60 Hz 1600W: 200-240 Vac / 10A / 50-60 Hz
Fans	GPGPU zone: 5x 8056 fan Server zone: 8x 4056 Fan Support hot-swap; N+1 redundancy
IO	1x VGA port 4x USB ports Front: 2x USB 2.0 ports Rear: 2x USB 3.0 ports 1x Dedicated BMC 1G port 2x 10G SFP+ NIC ports
OCP Slots	OCP 2.0 Supports 10GbE/25GbE/100GbE OCP 2.0 Mezz card