

Scorpio™ Smart Fabric Switch

Family Brief

X-Series 320 Lane | Largest Open Memory-Semantic Fabric Switch | Hypercast™ | In-Network Compute

P-Series 32 to 320 Lanes | Broadest Family of PCIe® 6 Fabric Switches

Asteralabs (NASDAQ: ALAB) provides rack-scale AI infrastructure through purpose-built connectivity solutions. By collaborating with hyperscalers and ecosystem partners, Asteralabs enables organizations to unlock the full potential of modern AI. Asteralabs' Intelligent Connectivity Platform integrates CXL®, Ethernet, NVLink Fusion, PCIe®, and UALink™ semiconductor-based technologies with the company's COSMOS software suite to unify diverse components into cohesive, flexible systems that deliver end-to-end scale-up, and scale-out connectivity. The company's custom connectivity solutions business complements its standards-based portfolio, enabling customers to deploy tailored architectures to meet their unique infrastructure requirements. Discover more at www.asteralabs.com

Benefits and Features

- **X-Series** Largest open, memory-semantic fabric switch engineered to improve token economics
- **P-Series** 32 to 320 Lanes: Broadest family of PCIe 6 fabric switches, designed to rapidly scale compute capacity across diverse accelerators
- **Hypercast™ Data Replication and In-Network Compute engines** accelerate collective operations by up to 2x to maximize GPU utilization and tokens-per-watt performance
- Line-rate peer-to-peer traffic with non-blocking crossbar and ACS restriction capability
- Host and Downstream Port Containment (DPC) and Advanced Error Reporting (AER)
- **Connectivity System Management and Optimization Software (COSMOS™)** suite with deep telemetry and diagnostics for link and device health monitoring to provide unprecedented data center observability
- Host management/diagnostics through fast I2C/I3C and In-band
- Easy to deploy with next-gen accelerators and GPUs to improve signal integrity, reduce crosstalk, which improves system reliability
- **Datacenter-grade RAS diagnostics:** Signal, link, and packet diagnostics for rapid fault isolation and accelerated time-to-deployment
- Data center security features to provide end-to-end protection compliant with leading industry standards such as NIST, OCP, FIPS, etc.

Applications

- **X-Series:**
 - Large radix GPU scale-up clusters with low latency
 - GPU scale-up clusters with memory-semantic based open and platform-specific protocols
 - GPU scale-up clusters for frontier / Mixture of Experts models
- **P-Series:**
 - High performance AI servers using PCIe for CPU, GPU, NIC, or storage connectivity
 - KV Cache memory/storage disaggregation
 - AI compute trays and servers requiring additional PCIe lanes and performance
 - AI compute trays and servers requiring additional front-end NICs and SSDs

Description

The Scorpio Smart Fabric Switch family is purpose-built for AI labs, hyperscalers, neo-clouds, and enterprises that need to maximize tokens-per-watt performance, rapidly deploy accelerator infrastructure, and ensure operational resilience for production AI at scale.

The Scorpio X-Series 320L, largest open, memory-semantic fabric switch enables scale-up clusters with

up to 80 GPUs with single hop latency. Hardware-accelerated Hypercast and In-Network Compute engines boost collective operations by up to 2x. This maximizes GPU utilization by spending less time sending data across the fabric, improving time to first token and tokens per watt performance.

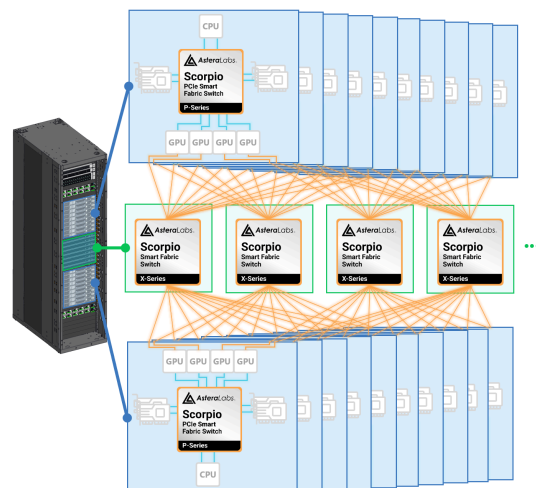
Hypercast and In-Network Compute engines are configured through COSMOS, extending its capabilities to now include platform performance optimizations. Scorpio X-series' memory-semantic connectivity enables accelerators to access fabric resources through native load/store operations and improving fabric efficiency at scale.

The Scorpio P-Series complements the X-Series in AI compute systems, delivering resiliency, dynamic configurability, and broad interoperability across GPUs and custom accelerators. The Scorpio P-series is now the broadest PCIe 6 Fabric Switch Family spanning 32 to 320 lanes.

At Asteralabs' Cloud-Scale Interop Lab, Scorpio Smart Fabric Switches are rigorously tested across a broad range of CPUs, GPUs, NICs and storage devices to accelerate time to market and reduce deployment risk. On-die temperature sensors enable continuous environmental monitoring and proactive operational management, while real-time telemetry, and datacenter-grade diagnostics work together to deliver the resilience needed to maximize server uptime.

COSMOS software unifies the platform with purpose-built resiliency and serviceability, delivering non-disruptive firmware updates, OpenBMC management, and real-time telemetry across Asteralabs' complete rack-scale portfolio of fabric switches, PCIe copper connectivity, and optical solutions—enabling composable AI infrastructure from a single unified management software stack.

High Radix GPU Scale-up

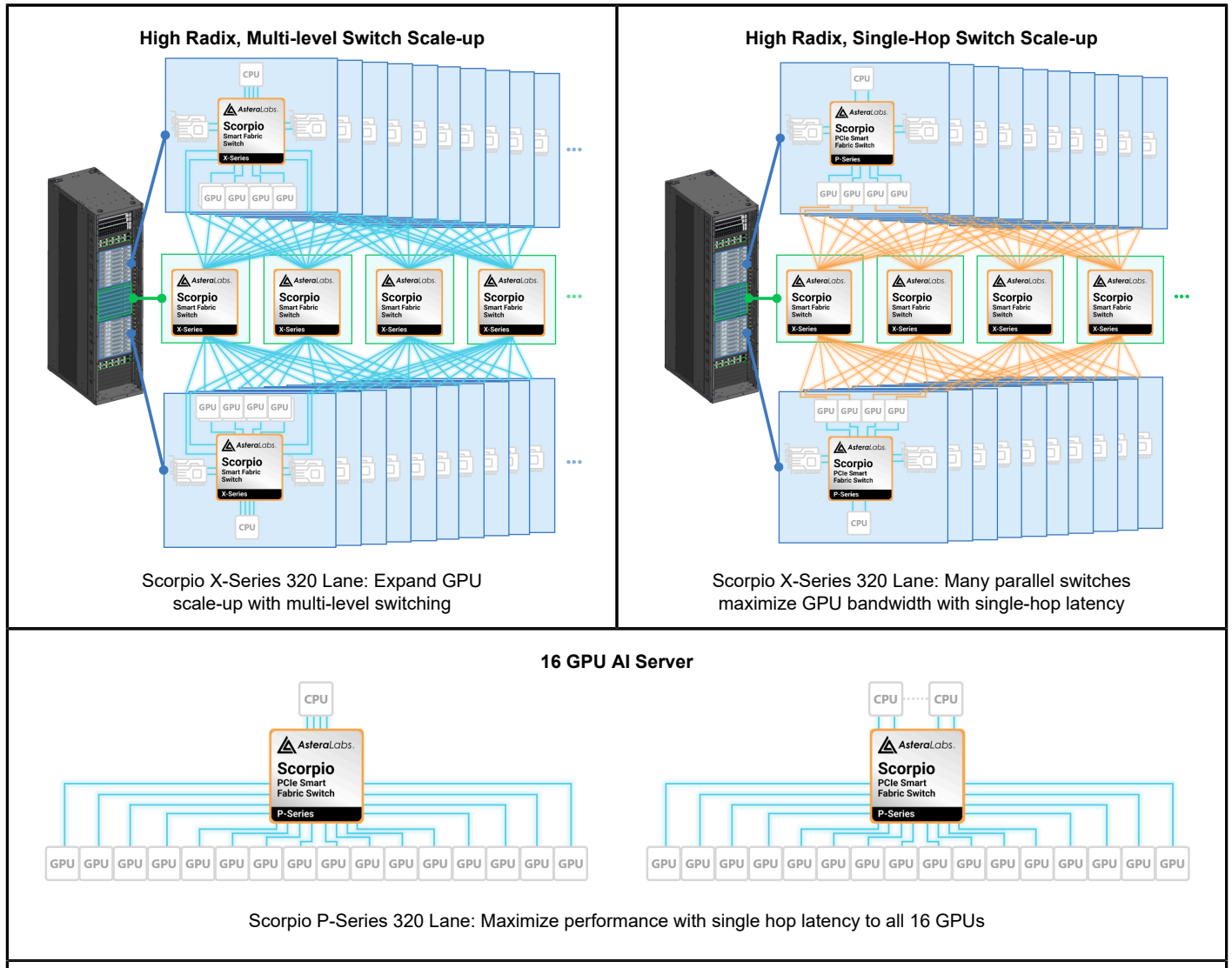


Product Family Information

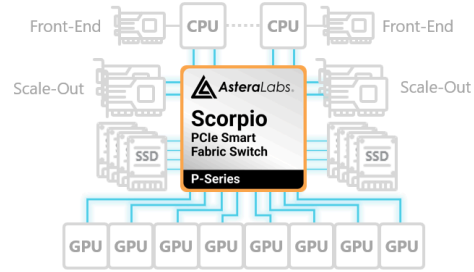
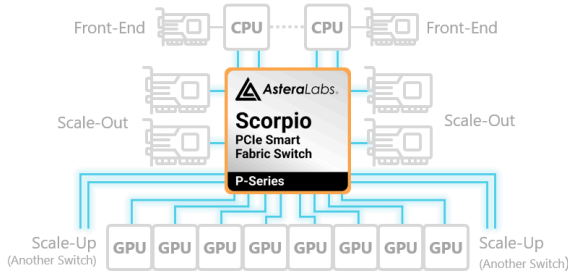
Series	Part #	Protocol	Lanes	Package Size [mm]
X-Series	Platform Specific up to 320 Lanes			
P-Series	PF63201L	PCIe 6	320	62.5x62.5
	PF61601L	PCIe 6	160	45x45
	PF60641L	PCIe 6	64	29x29
	PF60481L	PCIe 6	48	21x21
	PF60321L	PCIe 6	32	21x21
	PF50641L	PCIe 5	64	29x29

Example Topologies

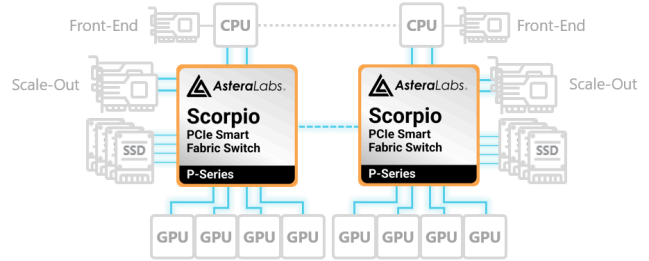
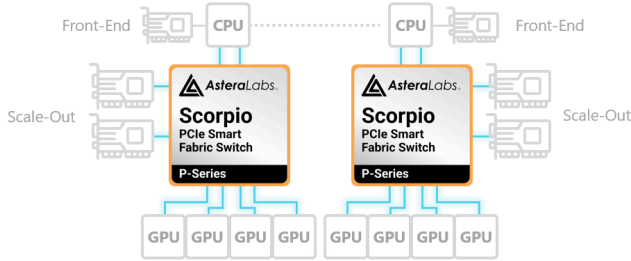
Fabric switches built from the ground up for AI clusters and servers, Scorpio Smart Fabric Switches support a diverse set of topologies. Below are some examples:



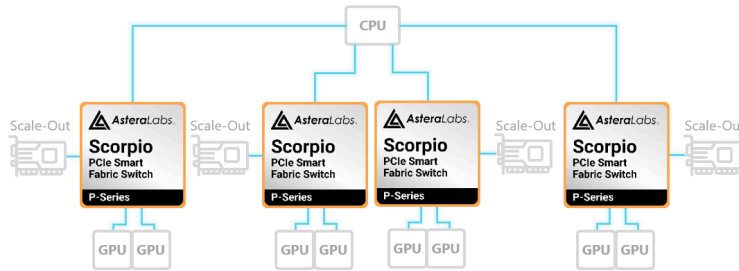
8 GPU AI Server



Scorio P-Series 320 Lane: All-to-all bandwidth across all 8 GPUs and devices with a single switch

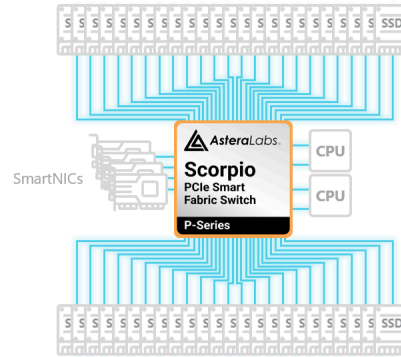
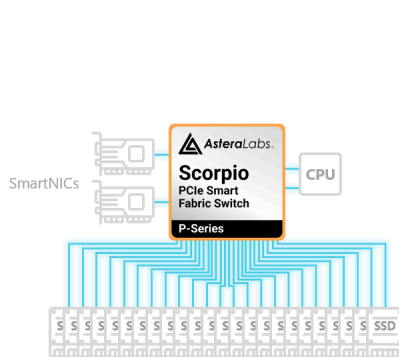


Scorio P-Series 160 Lane: High bandwidth GPU connectivity with optional switch-to-switch ports

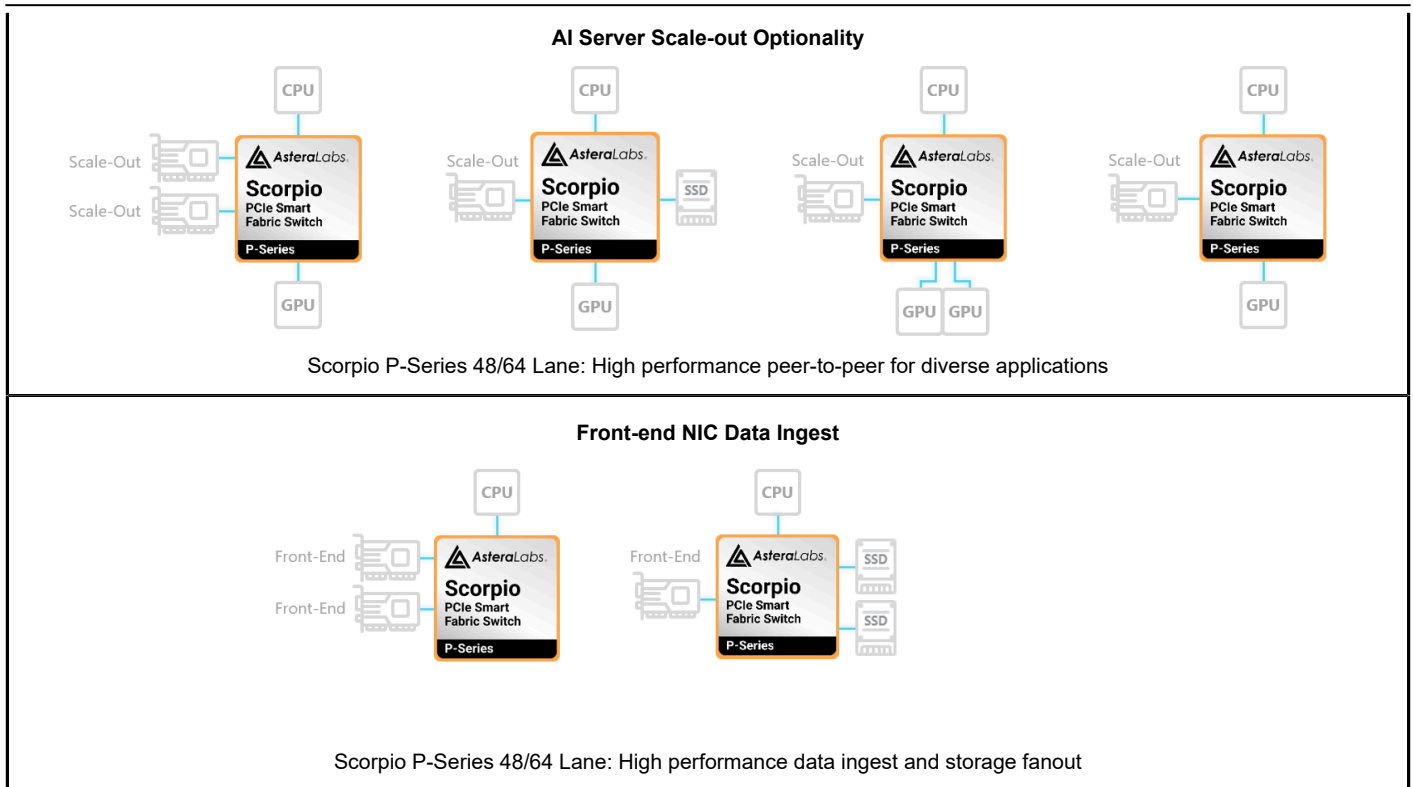


Scorio P-Series 64 Lane: High performance GPU to NIC scale-out with modular server design

KV Cache Memory/Storage Disaggregation



Scorio P-Series 160/320 Lane: Increase inference performance with disaggregated KV Cache appliance



Operational Resilience and Deep Diagnostics

COSMOS

Scorio is architected from the ground up for operational resilience, keeping production workloads running continuously while giving infrastructure teams the deep diagnostics to stay ahead of failures. Powering it all is COSMOS, delivering deep telemetry, link and device health monitoring, and unprecedented data center observability.

- **Resilient architecture:** COSMOS enables non-disruptive, customer-controlled firmware updates that isolate the switching plane from the management processor, so infrastructure stays live while your teams maintain it. No rack downtime. No cluster restarts.
- **Signal diagnostics:** The capability to pinpoint physical issues including damaged caps, loose cables, and disconnected pins at every link section through COSMOS, accelerating repair and recovery.
- **Link diagnostics:** When link training issues arise, COSMOS isolates the exact phase and source, whether a bad preset or a bad request, eliminating guesswork and dramatically compressing time-to-resolution.
- **Packet diagnostics:** For complex failures that evade standard diagnostics, COSMOS captures and surfaces link layer packet analysis, giving teams the precision needed to quickly resolve issues and accelerate platform deployment.