Data Center GPU Interconnect

Overview

AI-driven data centers evolve from single-chip to heterogeneous multi-GPU architectures. High-speed optical interconnects enable scalability, while silicon photonics and co-packaged optics boost bandwidth and energy efficiency amid modular, ecosystem-based competition. NVIDIA's latest AI platforms—including B200, B300, GB200, and GB300—introduce cluster interconnect designs that combine NVLink fabrics, high-performance NICs, and large-scale switching networks. This article explores how these technologies work together, from node-level GPU communication to. Intra-rack interconnects primarily address communication requirements within a single server rack, connecting multiple compute nodes (servers) or accelerator resources inside the rack. Shift from single-node to. With low latency, massive networking bandwidth, and all-to-all connectivity, the sixth generation NVIDIA NVLink™ and NVLink Switch are designed to accelerate training and inference for faster reasoning and agentic AI workloads. The sixth-generation NVLink enables 3. This shift is pushing optical interconnect.

Article Content

Scale-up networks for AI clusters open new optical

Cloud providers building out services in data centers are raising the bar for optical connectivity needs.

QSFP-DD for AI Data Centers: 400G/800G GPU Interconnect Guide

Learn how QSFP-DD optical transceivers enable AI data centers with 400G/800G bandwidth. Compare modules, architectures, and deployment strategies for GPU clusters.

NVIDIA Data Center GPUs Explained: From A100 to

Understand NVIDIA data center GPUs for AI inference. Compare T4, L4, A100, H100, H200, and B200 on use cases, memory, and pricing to choose

Infrastructure for Scalable AI Reasoning | NVIDIA Vera

Real-time ray tracing and AI for advanced design and visualization Modular reference architecture for building accelerated data center servers High-speed

Nvidia outlines plans for using light for communication

You may like Nvidia updates data center roadmap with Rosa CPU and stacked Feynman GPUs AMD, Broadcom, and Nvidia join hyperscalers to

NVIDIA B200/B300/GB200/GB300 Cluster Interconnect Architecture

To connect GPU nodes to the cluster network, each system integrates multiple high-speed network interface cards (NICs). These NICs provide the external connectivity required for multi

Taking on Nvidia for data centre interconnect with UAlink

The data centre industry has created the UAlink promoter group to take on Nvidia''s NVlink GPU interconnect technology.

Confusion Grows With More Interconnect Options And Tradeoffs

References All AI Data Center Interconnects Will Be Optical Within 5 Years (Semiconductor Engineering) AI Optical Interconnect Boom Drives U.S. Firms to Expand

NVIDIA Data Center GPU Specs: A Complete

Updated 2026 comparison of NVIDIA data center GPUs: Blackwell Ultra B300, B200, GB200 NVL72, H100, H200, A100 & L40S — specs, FLOPS, NVLink,

NVIDIA HGX Platform

The NVIDIA HGX B300 integrates NVIDIA Blackwell Ultra GPUs with high-speed interconnects to propel the data center into a new era of accelerated computing

AWS and NVIDIA deepen strategic collaboration to accelerate AI from ...

This integration enables high-throughput, low-latency KV-cache data movement between GPU compute nodes performing token generation and distributed memory resources that store KV

Explaining NVIDIA Full-Scenario Compute

This article examines NVIDIA''s interconnect technologies across intra-rack, data center, and inter–data center environments, highlighting how

2025 Global AI Data Center Interconnect Trends

AI-driven data centers evolve from single-chip to heterogeneous multi-GPU architectures. High-speed optical interconnects enable scalability,

Next Gen Data Center CPU | NVIDIA Vera CPU

Real-time ray tracing and AI for advanced design and visualization Modular reference architecture for building accelerated data center servers High-speed

GPU Interconnects: NVLink vs PCIe

Compare GPU interconnect technologies like NVLink and PCIe and their effect on multi-GPU training performance.

AI Data Center Interconnect 2026: CPO, Optical Interconnect and ...

Explore AI data center interconnect trends in 2026, including CPO, optical interconnect, OCS, and the real challenges slowing large-scale deployment.

800G Interconnect Guide: DAC, ACC, AEC & AOC Comparison

Choose the right 800G interconnect for AI data centers. Compare DAC, ACC, AEC, AOC specs, distance frameworks, power costs & 1.6T upgrade paths for GPU clusters.

New Nvidia software gives data centers deeper visibility

Nvidia has released new open-source software that gives data center operators deeper visibility into the thermal and overall health of its AI GPUs,

NVLink & NVLink Switch: Fastest HPC Data Center

Real-time ray tracing and AI for advanced design and visualization Modular reference architecture for building accelerated data center servers High-speed

High-Speed Interconnect Technology Choices in the AI Data Center Era

InfiniBand is a high-speed, ultra-low-latency interconnect architecture specifically designed for HPC and AI supercomputing clusters.

H200 GPU | NVIDIA

Modular reference architecture for building accelerated data center servers High-speed interconnect for multi-GPU communication and large AI models Semi

7 Best Data Center Stocks, ETFs and REITs to Buy

Demand for AI is driving a data center boom. These stocks, ETFs and REITs are positioned to benefit from it.

NVIDIA Feynman GPU Gets 3D Die-Stacking, Custom

NVIDIA has revealed more details of its next-generation AI Data Center solution, Rosa Feynman, built using 3D Die-Stacking.

NVIDIA NVLink Explained: A Guide to the GPU

Modern supercomputers and data centers use thousands of NVIDIA GPUs to train deep neural networks, simulate scientific models, and process large data sets.

Build Semi-Custom AI Infrastructure | NVIDIA NVLink

Real-time ray tracing and AI for advanced design and visualization Modular reference architecture for building accelerated data center servers High-speed

CPUs are Back: The Datacenter CPU Landscape in 2026

CPUs are Back: The Datacenter CPU Landscape in 2026 RL and Agent Usage, Context Memory Storage, DRAM Pricing Impacts, CPU Interconnect Evolution, AMD Venice, Verano,

NVLink & NVLink Switch: Fastest HPC Data Center

With NVLink Switch, NVLink connections can be extended across nodes to create a seamless, high-bandwidth, multi-node GPU cluster—effectively forming a data

Data center semiconductor trends 2025: Artificial

Yole Group publishes its new report, Data Center Semiconductor Trends 2025, offering an in-depth analysis of how AI, HPC, and hyperscaler

Contact Us

For more information, pricing, or custom solutions, please contact us:

Website: https://truhope.co.za
Email: sales@truhope.co.za
Phone: +27 64 987 3021
Address: 22 Loop Street, Cape Town, 8001, South Africa

This document is for informational purposes only. Specifications subject to change without notice.