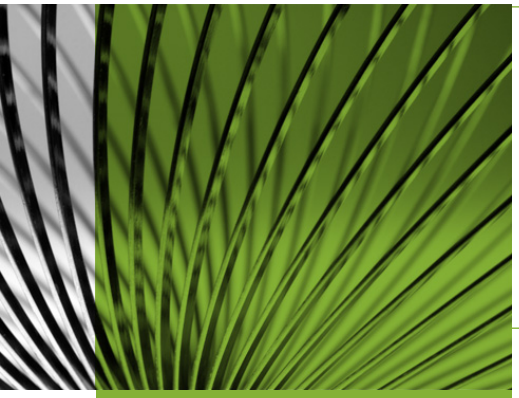

OPEN, ETHERNET-BASED AI DATA CENTER NETWORKING SOLUTION BRIEF

End-To-End, Multivendor, Intent-Based AI Networking For Speed, Efficiency, And Economics

Challenge
Deploying AI clusters is complex, time consuming, and expensive. Enterprises investing in AI applications have limited resources to deploy AI data centers and optimize expensive GPU resources without vendor lock-in.

Solution
Juniper's AI Data Center solution is the fastest and most flexible way to deploy high performing AI training, inference, and storage clusters, and the simplest to operate with limited IT resources.

Benefits
- Simplified Day 0/1/2+ operations use fewer resources and save time and money.
- Open flexibility to design GPU-agnostic networks using proven technologies and products that avoid vendor lock-in with performance that is on par with InfiniBand.
- Turnkey end-to-end validated solutions ensure confidence in choice of products with expedited deployment times.

Traditional data center technologies and designs fall short of demanding performance, capacity, and latency requirements AI workloads place on infrastructure. AI data centers, characterized by specialized back-end training and front-end inference fabric designs, require cutting-edge solutions in compute, storage, and networking. Amid these developments, there has been a surge in demand for powerful Graphics Processing Units (GPUs) and a reliance on single-sourced, proprietary InfiniBand networking. Unfortunately, this has led to escalated pricing and supply chain bottlenecks. Lowering costs while fueling innovation and flexibility is imperative for enterprises. Recognizing the need to optimize GPU performance while minimizing costs, enterprises are increasingly turning to Ethernet as the preferred open networking alternative for AI data centers.

The Challenge

Across multiple industries, AI is transitioning from engineering lab to boardroom deliverable, and AI architects and their IT counterparts are suddenly navigating new territories. To realize AI's full value, they must first manage the complexities of AI deployment and deal with the realities of AI costs and lead times.

With few practical GPU options and proprietary InfiniBand networking, supply chain delays and low inventory have increased the costs of expensive and rare AI infrastructure. A single GPU, for example, costs more than $30,000 and the AI servers that use these chips can cost upwards of $400,000. With such a high price, even small GPU clusters can cost in the millions while large GPU clusters can cost hundreds of millions of dollars.

As the engine behind AI, data center networks play a critical role in interconnecting and maximizing GPUs. Reducing job completion time (JCT)—the time it takes to complete AI training—is the key to faster speed and cost savings.

Successful AI adoption also depends on fast market response to demand. To ease the reliance on single vendor solutions, the industry has made significant moves to foster an open, competitive market with GPU diversity and the most widely deployed L2 technology in the world, Ethernet.

The Juniper Networks AI Data Center Networking Solution

Juniper's AI Data Center solution is the fastest and most flexible way to deploy high-performing AI training, inference, and storage clusters, and the simplest to operate with limited IT resources. With unique intent-based operations, an AI-native virtual network assistant, and Juniper validated designs, Juniper simplifies AI data center networking design, deployment, and troubleshooting. The solution also delivers unsurpassed flexibility to customers, avoiding vendor lock-in with silicon diversity, multivendor switch management, and a commitment to open, standards-based Ethernet fabrics.

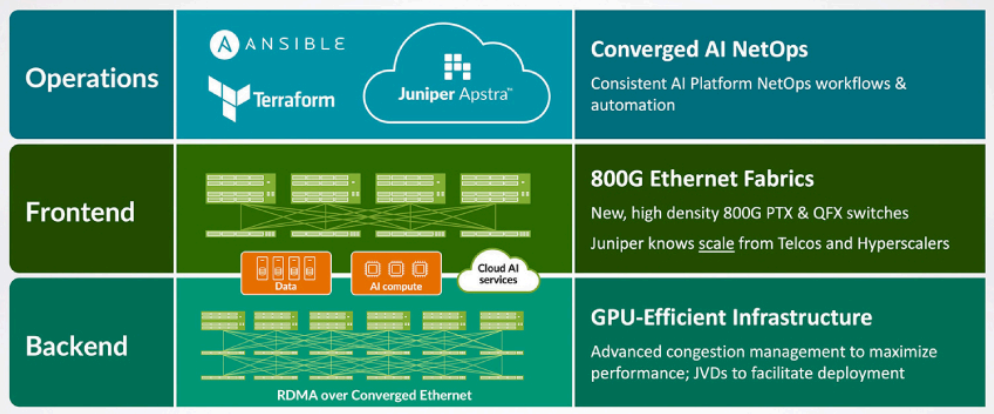

Figure 1: Juniper Networks AI Data Center Networking Solutions

Simplified Operations

Simplified Day 0/1/2+ operations use fewer resources and save time and money. Juniper software is the only multivendor data center automation platform with industry-leading, intent-based networking that accelerates time to deploy and troubleshoot by avoiding manual configuration challenges. For example, advanced load-balancing, explicit congestion notification (ECN), and priority flow control (PFC) features essential to high-performance, lossless AI networking can be applied across the fabric using Apstra without typing a single CLI command. Apstra's real-time network flow visualization pinpoints AI data center congestion hot spots with heatmaps and intent-based analytics for fast issue resolution.

As the single source of truth with closed loop assurance, Apstra offers powerful analytics to predict and proactively fix problems and avoid service disruptions. With templated blueprints, Apstra delivers reliability, consistency, and repeatability that translate into a 90% lower OPEX and an 85% faster deployment.

Through Mist AI™ with Apstra integration, Juniper is bringing AIOps to the data center, delivering additional automation and insight to optimize data center experiences. Marvis™ Virtual Network Assistant for the data center extends the rich streaming telemetry and real-time monitoring in Apstra into the Marvis dashboard to provide proactive troubleshooting and analytics across operational domains. This integration is a fundamental first step to providing future AI-driven actionable insights for multivendor data centers. Campus and branch operators using Marvis can now see data center issues on the same dashboard.

Open Flexibility

Open, flexible Ethernet solutions allow you to use proven technologies and products that avoid vendor lock-in. Juniper's new AI-optimized, high-radix 800GbE leaf and spine data center fabrics provide advanced traffic management capabilities to assure high-bandwidth, low-latency, lossless, and scalable performance over Ethernet. With a runway to 1.6 terabits and multivendor support for GPU-agnostic systems, Ethernet reduces costs, assures faster innovation, maximizes design flexibility, and avoids supply chain challenges. Support for Remote Direct Memory Access over Converged Ethernet v2 (RoCEv2) makes Juniper's Ethernet fabrics the only multi-vendor solution for DC fabric management and automation optimized for back-end AI training and front-end inference models.

Turnkey Solutions

Turnkey, end-to-end validated solutions ensure confidence in choice of products to expedite data center deployments. When enterprises have end-to-end network solutions, they can build high-performing AI data centers with flexibility and ease. Juniper Networks® PTX Series Routers for data centers have been extended with the new high density 800GbE PTX10002-36QDD fixed switch and new 800GbE line cards for the PTX10000 chassis. Built on Express 5 custom silicon, the largest PTX10000 chassis now supports up to 576 x 800GbE ports for high radix spine and super spine architectures. In addition, a new high-density 800GbE QFX5240 fixed platform based on the latest Broadcom Tomahawk

5 ASIC offers silicon diversity for high performance, scalable, power-efficient AI data centers.

The Juniper QFX and PTX Series platforms deliver exceptional performance, scalability, and congestion control, guaranteeing peak operation for AI clusters. In MLPerf benchmark tests, Juniper's AI Optimized Ethernet demonstrated Job Completion Time (JCT) for training workloads on par with InfiniBand, offering customers an open, high-performance fabric alternative that avoids vendor lock-in. To simplify AI deployments, Juniper is introducing new Juniper Validated Designs (JVDs). AI JVDs enable prescribed, rail-optimized, or multilayered clos fabrics to assure a complete data center solution, including switching, operations, and security. Rigorously tested, validated, and documented by Juniper lab professionals, JVDs can be used as out-of-the-box designs or as a set of guidelines to reduce risk and to properly size and budget every AI cluster.

Features and Benefits

Simplified operations: Use fewer resources and save time and money with simplified Day 0/1/2+ operations. Industry leading intent-based networking accelerates time to deploy and troubleshoot by avoiding manual configuration challenges (across any vendor's hardware).

- Single source of truth with closed loop assurance—predict and proactively fix problems to avoid service disruption
- Only AIOps for Data Center delivers additional automation and insights to optimize data center experiences
- Cost-effective solution with a 90% lower OpEx and 85% faster deployment

Open flexibility: Design your network using proven technologies and products that avoid vendor lock-in.

- Advanced traffic management capabilities assure high-bandwidth, low-latency, lossless, and scalable performance over Ethernet
- Leveraging Ethernet pushes down costs, assures faster innovation, maximizes design flexibility and avoids supply chain challenges
- Only multi-vendor solution for DC fabric management and automation maximizes design flexibility
- Interoperable with all GPUs, fabrics and switches

Turnkey solutions: End-to-end validated solutions ensure confidence in choice of products; expedite deployment times.

- Silicon diversity drives scale, performance, and flexibility. Support for Broadcom and custom silicon interconnect for thousands of GPUs at highest speeds (800 Gbps) and supports all data center fabric configurations
- Comprehensive data center security portfolio for end-to-end security
- Juniper Validated designs assure a complete data center solution, including switching, operations, and security
- Juniper's broad portfolio of switches (with silicon diversity) drives scale and performance
- Comprehensive DC Security portfolio with unified threat prevention for zero trust data center protection
- Improves data center network reliability tenfold

Solution Components

Apstra: Juniper's data center automation solution enables operators to automate the entire data center network life cycle across multivendor environments from design to ongoing operations. Apstra provides continuous validation, a single source of truth, powerful analytics, and rapid root-cause identification and remediation.

QFX Series Switches: Industry-leading line of switches delivers superior throughput and scalability, a comprehensive routing stack, the open programmability of Junos® OS, and the broadest set of EVPN-VXLAN and IP fabric capabilities.

PTX Series Routers: PTX Series Routers form the foundation of the world's largest core and WAN architectures. Powered by the latest generation of the Express family ASICs, the routers are cloud-optimized to enable smooth 800GbE migrations. They simultaneously deliver the scale, flexibility, and investment protection needed to meet today's market needs.

Marvis Virtual Network Assistant for Data Center: Marvis is the first and only AI-native virtual network assistant for the data center, delivering the best insight throughout the entire data center life cycle across any vendor's hardware. In addition, Marvis provides end-to-end visibility and assurance across all enterprise domains, from campus and branch to data center.

Junos OS Evolved: This unified, end-to-end network operating system provides reliability, agility, and open programmability for successful cloud-scale deployments.

Summary: Simplified Operations, Open Flexibility, and Turnkey Networking for the AI Data Center

Virtually every company is defining a strategy for how to best use data center infrastructure to execute corporate AI initiatives. Juniper's AI Data Center solution is the quickest and easiest way to deliver high-performing and scalable networks for AI training and inference. With multivendor, intent-based operations, an AI-native virtual network assistant, and Juniper validated designs, Juniper takes the complexity out of AI Data Center networking.

Corporate and Sales Headquarters

Juniper Networks, Inc.

1133 Innovation Way

Sunnyvale, CA 94089 USA **Phone:**

888.JUNIPER (888.586.4737)

or +1.408.745.2000

www.juniper.net

APAC and EMEA Headquarters

Juniper Networks International B.V.

Boeing Avenue 240

1119 PZ Schiphol-Rijk

Amsterdam, The Netherlands **Phone:**

+31.0.207.125.700