

# JUNIPER AND WEKA POWER AI DATA CENTERS SOLUTION BRIEF

Get The Fastest Performance, The Highest Reliability, And An AI/ML Network Infrastructure That Adapts To Your Data Needs.

## Challenge

AI/ML workloads are computationally intensive, requiring a full-stack data center infrastructure of compute, storage, and networking that is optimized for performance. GPU and tensor processing units (TPU) throughput is critical for customer experience, and poorly designed networking and storage can negate performance gains.

## Solution

Delivering a pre-validated solution, Juniper and WEKA provide scalable, high-performance, AI-optimized data center solutions. Combining the scale, performance, and capacity of Juniper's Ethernet-based AI data center fabrics and the WEKA® Data Platform for flexible and efficient storage optimizes data center infrastructure to accelerate AI performance.

## Benefits

- High-performance, efficient data center designs for the best AI performance
- Multivendor data center infrastructure for choice, flexibility, and economics
- AI lab-tested solutions with Juniper Validated Designs (JVDs)

## The challenge

AI/ML server infrastructure places huge demands on the entire data center stack, including the network, the data storage system, and even the power and cooling infrastructure. For optimal server performance, the network and storage systems must be efficient and reliable. When network issues occur, job completion times (JCTs) will be extended, or worse, jobs may need to be restarted. Storage issues can result in slow checkpoints, longer JCTs, or data stalls. In other words, if the network and storage systems are not optimized, the performance gains in GPU and TPU technologies will not be realized, costing time and money.

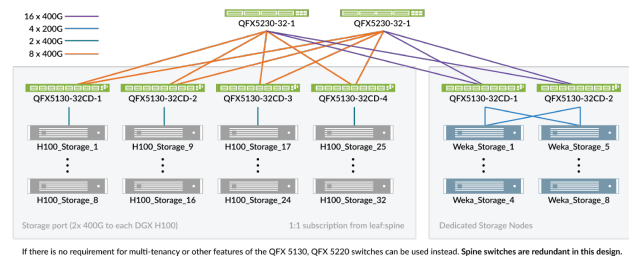


Figure 1: Components of the Juniper and WEKA solution

## The Juniper and WEKA solution

Next to IP, Ethernet is the world's most dominant networking technology. With low costs, a full developer ecosystem, and port speeds soon to reach 1.6 Tbps, Ethernet is well suited to handle the increasing demand for AI and ML technologies. As the world's premier networking vendor powering eight of the Fortune 10 companies, Juniper Networks is leading the investment and innovation in Ethernet solutions to handle the unique and rigorous challenges of AI/ML data center networking.

High-performance AI/ML data center networking relies upon speed, capacity, and lossless, low-latency connectivity. The Juniper QFX series of switches delivers high-density, 400G and 800G Ethernet solutions built upon the latest Broadcom technology. With non-blocking, lossless performance and advanced traffic management features, QFX switches are a game changer for both GPU interconnect and dedicated storage fabrics. When used with the WEKA Data Platform for high-performance data access, the combined QFX and WEKA solution optimizes GPU performance and efficiency for AI/ML training and inference.

WEKA has developed an innovative approach to the enterprise data stack, specifically designed for the AI era. The WEKA Data Platform represents a new benchmark for AI infrastructure, featuring a cloud and [AI-native](#) architecture that offers unparalleled flexibility, allowing deployment across on-premises, cloud, and edge environments. This platform enables seamless data portability, effectively transforming outdated data silos into

dynamic data pipelines. As a result, it significantly accelerates GPUs, AI model training, and inference, as well as other performance-intensive tasks. This enhanced efficiency not only optimizes workload performance but also reduces energy consumption and carbon emissions—contributing to a more sustainable computing environment.

### Why use Juniper in my AI/ML storage fabric?

As part of a three-pronged solution, compute, networking, and storage are interdependent for overall performance of AI/ML training and inference models. While much focus has been placed on Nvidia GPU performance gains, the performance demands placed on storage continue to increase at a rate that is commensurate with chip technology.

To optimize storage performance and eliminate bottlenecks that could impact JCT, Juniper's 400G and 800G QFX series of switches deliver the cost efficiencies and openness of Ethernet while matching the performance of other proprietary standards. Tested together and documented as a Juniper Validated Design, the QFX along with WEKA storage and Nvidia compute are pre-validated by Juniper test engineering to simplify the design and deployment of high performance, stable AI [data centers](#) that are fast, simple, and cost-effective.

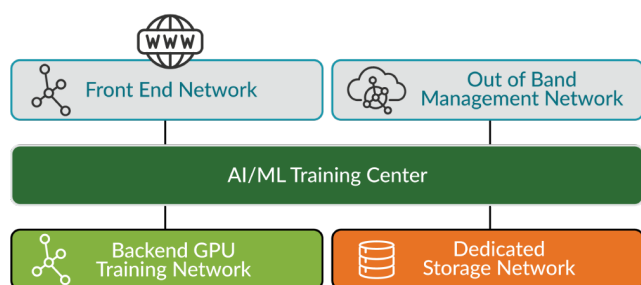


Figure 2: WEKA storage supporting both frontend and backend networks.

### Solution Components

WEKA solution components include:

#### WEKA High-Performance Data Platform for AI

The WEKA Data Platform is a software-based solution built to modernize enterprise data stacks. The software can be installed on any standard AMD EPYC or Intel Xeon Scalable processor-based hardware with the appropriate memory, CPU processor, networking, and NVMe solid-state drives. A configuration of six storage servers is required to create a cluster that can survive a two-server failure.

An AI data pipeline must be able to handle all types of data and data sizes. The WEKA Data Platform's patented data layout and virtual metadata servers distribute and parallelize all metadata and data across the cluster for incredibly low latency and high performance, no matter the file size or number.

WEKA's unique architecture is radically different from legacy storage systems, appliances, and hypervisor-based software-defined storage solutions because it not only overcomes traditional storage scaling and file sharing limitations that hinder AI deployments, but also allows parallel file access via Portable Operating System Interface (POSIX), Network File System (NFS), Server Message Block (SMB), Amazon Simple Storage Service (S3), and GPUDirect storage.

Juniper solution components include:

#### Juniper QFX Switch Series

For high-performance storage fabrics, Juniper recommends two specific switches from our broader QFX portfolio.

The Juniper [QFX5130](#) line of switches is designed to secure and automate the data center network. With 32x 400G ports, up to 51.2 Tbps (bidirectional) Layer 2 and Layer 3 performance, and with latency as low as 550 nanoseconds, the QFX5130 switches are perfect for carrying high-performance storage traffic from platforms such as WEKA.

The Juniper QFX5130 switches are ideal leaf switches within a storage fabric. These 25.6 Tbps, high-throughput switches fully support multitenancy and other enterprise features.

Meanwhile, the QFX5230 line make capable spine switches utilized in the Juniper and WEKA solution to provide 64x400G, high-performance, non-blocking transport for the QFX5130 leaf switches. Like the QFX5130 switches, the QFX5230 features up to 51.2 Tbps bidirectional throughput.

#### Example architecture

In this architecture, we are using the WEKA distributed POSIX client, combined with Juniper switches to achieve storage throughput rates exceeding 720 Gbps and read Input/output operations per second (IOPS) exceeding 18M.

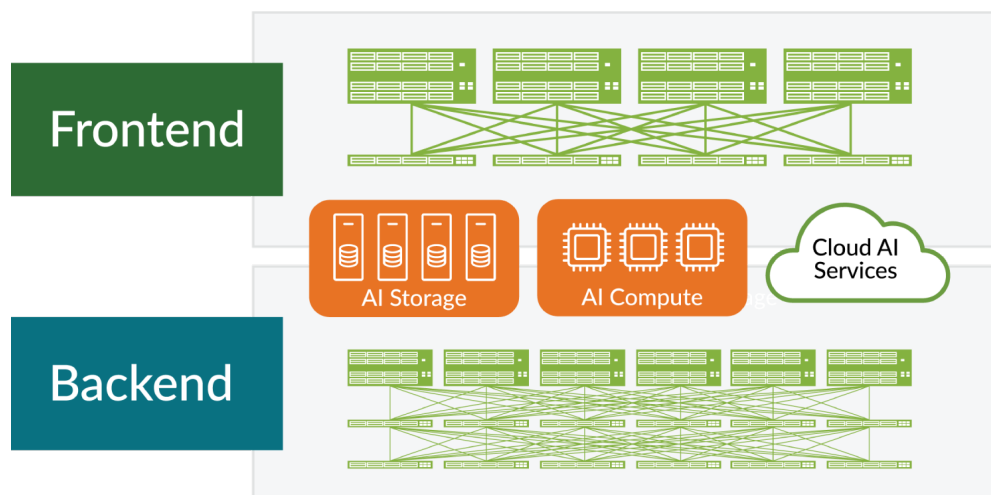


Figure 3: An eight-node WEKA cluster in a dedicated storage fabric, supporting an NVIDIA-powered AI/ML training cluster.

At the network layer, we are using Juniper QFX5130 switches acting in a leaf node capacity, and Juniper [QFX5230](#) switches in the spine role. Connectivity from the WEKA nodes to the leaf-switches is 200G, and 400G from the leaf switches to the spine switches. Connectivity from the GPU hosts to the leaf switches is 400G with either a single or two 400G connections utilized (Figure 3 shows two connections in use). The network capacity from the GPU server NICs to the spine is 25.6 Tbps. From the spines to the storage leaf switches, it is 12.8 Tbps, a 2:1 oversubscription to the storage leaf switches. The WEKA nodes themselves, have 12.8 Tbps of bandwidth from the servers to the leaf switches. The spine switches have 16 additional ports free on each to accommodate future expansion.

In this architecture, the network is a flat IP fabric with no requirement to implement Data Center Quantized Congestion Notification (DCQCN) or other network congestion avoidance and management features. This makes the WEKA cluster easy to deploy, ensures low-latency, full bandwidth, and provides the highest performance and throughput speeds.

### Combining WEKA with Juniper QFX Switches

Bringing the WEKA Data Platform together with Juniper Apstra and Juniper switches creates an easy to deploy, easy to manage, end-to-end storage solution for your AI clusters. With comprehensive enterprise features (no CLI required), unchallenged performance and ease of management, the WEKA Data Platform plus Juniper solutions provides best-in-class performance—without the complexity and trade-offs of proprietary solutions.

### Juniper Apstra

Use Juniper Apstra to easily provision VXLAN tunnels for storage traffic, providing security and data segregation for multitenancy.

Apstra intent-based networking software automates and validates the design, deployment, and operation of data center networks, from Day 0 through Day 2+. Tag storage interfaces as WEKA and easily use configlets for any specific network class of service or other required customizations.

### Summary—a network infrastructure ready for AI demands

Leverage the power of distributed storage by using the WEKA distributed storage system, and combine it with Juniper QFX 400G switching to ensure that your storage performance achieves record-breaking storage throughput, combined with a low-latency, rock-solid, IP-based network from Juniper.

### About Juniper

Juniper Networks believes that connectivity is not the same as experiencing a great connection. Juniper's AI-Native Networking Platform is built from the ground up to leverage AI to deliver exceptional, highly secure, and sustainable user experiences from the edge to the data center and cloud. Additional information can be found at Juniper Networks ([www.juniper.net](http://www.juniper.net)) or connect with Juniper on [X](#) (Twitter), [LinkedIn](#), and [Facebook](#).



### About WEKA

WEKA is architecting a new approach to the enterprise data stack built for the AI era. The WEKA® Data Platform sets the standard for AI infrastructure with a cloud and AI-native architecture that can be deployed anywhere, providing seamless data portability across on-

premises, cloud, and edge environments. It transforms legacy data silos into dynamic data pipelines that accelerate GPUs, AI model training and inference, and other performance-intensive workloads, enabling them to work more efficiently, consume less energy, and reduce associated carbon emissions. WEKA helps the world's most innovative enterprises and research organizations overcome complex data challenges to reach discoveries, insights, and outcomes faster and more sustainably – including 12 of the Fortune 50. Visit [www.weka.io](http://www.weka.io) to learn more or connect with WEKA on LinkedIn, X, and Facebook.

See why WEKA has been recognized as a Visionary for three consecutive years in the

Gartner® Magic Quadrant™ for Distributed File Systems and Object Storage – get the report.

## Reference Links

[WEKA Architecture Guide](#)

[Checkpointing for Resiliency and Performance in AI Pipelines](#)

[Juniper Apstra Data Center Automation](#)

[QFX Series Switches](#)

[Juniper 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**