

100-Gigabit Ethernet for High-Performance Computing with QFX10000 Switches

Leverage fast, affordable, and familiar Ethernet for HPC storage interconnect

Challenge

Leverage the economics and familiarity of Ethernet for high-performance compute, big data, and other data-intensive workloads.

Solution

- QFX10000 line of Switches

Benefits

- Deploy high-performance, highly available, and flexible storage interconnect for HPC workloads
- Lower CapEx and OpEx with affordable, familiar Ethernet for storage interconnect
- Reduce the need for specialized InfiniBand skills

High-performance computing (HPC) is critical to finding answers to complex questions across government, academia, and industry. In HPC, supercomputers and advanced processing techniques are used to solve complex computational problems at lightning fast speeds through computer modeling, simulation, and data analysis. With uses ranging from curing diseases and making air travel safer to improving weather forecasts and advancing manufacturing techniques, HPC enables countless breakthroughs and innovations.

Ethernet and InfiniBand have long been used as interconnect technologies for HPC clusters. InfiniBand was supposed to kill Ethernet, solving all of the connectivity problems between servers and storage, but InfiniBand never became a universal interconnect. In fact, Ethernet is still the most widely deployed network technology at the TOP500 supercomputer sites, according to InsideHPC.¹ InfiniBand may offer speed and latency advantages, but it's expensive, controlled by a single vendor, and requires a specialized IT skill set.

The Challenge

Not only has Ethernet remained a viable choice for HPC interconnect, but its use is growing as 100-Gigabit Ethernet (100GbE) delivers the necessary speed, stability, and scale with cost advantages. 100GbE is gaining momentum as the preferred HPC storage interconnect of choice, particularly for applications that aren't latency-sensitive, or for big data analytics and other network-intensive workloads. With networking accounting for as much as a third of the cost of a cluster, high-density 100GbE switches that are capable and cost-effective are compelling for both storage and compute interconnect applications.

High-speed data transfer over Ethernet continues to advance. Ethernet continues to get faster, and industry development of 400GbE is well underway. Traditional latency issues have been addressed within the Ethernet protocol, such as by remote direct memory access. IP over Ethernet solutions, such as RoCE and iWARP, provide all of the benefits common to Ethernet networking, including easy-to-implement traffic load sharing, security via access control lists (ACLs) and 802.1X, and traffic separation with VLANs and MPLS VPN.

The Juniper Networks HPC Storage Interconnect Solution

Juniper Networks® QFX10000 line of Switches is an affordable, flexible, and simple way to interconnect HPC and high-capacity, big data storage systems. Based on purpose-built Q5 ASICs, the QFX10000 line of modular Ethernet switches delivers unparalleled intelligence and analytics, providing deeper insights into application performance.

Features and Benefits

The QFX10000 line offers key advantages for HPC storage and compute interconnect, including:

- **Massive scale and performance even over long distances.** Organizations can leverage the QFX10000 line of Switches to create a high-performance fabric

¹ "HPC Networking Trends in the TOP500," InsideHPC, January 4, 2017, <https://insidehpc.com/2017/01/hpc-networking-trends-top-500/>



with guaranteed quality of service (QoS) and dynamic congestion management. QFX10000 Switches deliver up to 96 Tbps of throughput today and are scalable up to 200 Tbps in the future.

The memory architecture of the QFX10000 line reduces the risk of packet loss to near zero, which is critical for HPC and similarly demanding applications. The QFX10000 line uses a Hybrid Memory Cube (HMC), a new class of three-dimensional memory that's purpose-built for massive scale and performance. Before, it wasn't possible to design systems that could scale on multiple vectors, including I/O bandwidth and logical system scale. Due to limited memory technology, only one of the system's dimensions could be scaled, either I/O or the memory reserved for packet buffering and forwarding information base (FIB) tables.

The combination of Juniper's Q5 forwarding ASIC and HMC technology breaks that barrier, enabling both I/O and packet buffer and lookup memory to be scaled at the same time.

- **High availability.** The carrier-grade QFX10000 line provides unmatched reliability. Juniper Networks Junos® operating system running on the QFX10000 Switches has also been enhanced for greater scalability, modularity, and programmability, ensuring that these high-performance switches deliver an unparalleled pace of innovation.
- **Lower CapEx and OpEx.** 100GbE is less expensive than InfiniBand, and organizations can see a savings of more than 40 percent per port. But lower CapEx is just the start. Because Ethernet is ubiquitous, it's well understood by IT professionals and well supported by a broad range of tools. That lowers operating expenses (OpEx) over the long term. Organizations need fewer IT staff with specialized, expensive InfiniBand skills, and instead, can leverage existing Ethernet skills.
- **Simplified HPC networking.** Ethernet offers the promise of a single, converged network infrastructure to support HPC, storage, and user traffic. With Juniper, organizations only need to consider the OpEx associated with Ethernet, rather than maintaining multiple network technology infrastructures.

A Proven Solution

The National Center for Atmosphere Research (NCAR) uses Juniper's HPC interconnect solution as part of its mission to study climate change and global warming. Scientists use NCAR's supercomputer to perform some of the world's most data-intensive calculations for climate modeling, providing information that can profoundly impact communities by helping to improve computer models in ways that can better inform future evacuation efforts or aid in the dispatch of recovery teams.

Juniper solutions, including the high-performance, high-density QFX10008 Switch, serve as the foundation of the supercomputer network. The network is designed to meet the demanding capacity and bandwidth required by researchers to process climate data and conduct predictive weather analysis while sharing results with colleagues around the world. The system is expected to perform 5.34 quadrillion calculations per second, making it one of the top performing supercomputers in the world.²

Summary—Simplify HPC Interconnect

Organizations can simplify their HPC clusters and lower costs by using 100GbE for compute and storage interconnect, and future technology advances will deliver even higher performance, availability, and flexibility. With QFX10000 Switches, organizations have a high-performance, carrier-grade connectivity to meet the growing demand for HPC and big data analytics.

Next Steps

For more information about the Juniper Networks line of QFX10000 Switches, please visit www.juniper.net/us/en/products-services/switching/afx-series/afx10000/ or contact your Juniper representative.

About Juniper Networks

Juniper Networks challenges the status quo with products, solutions and services that transform the economics of networking. Our team co-innovates with customers and partners to deliver automated, scalable and secure networks with agility, performance and value. Additional information can be found at Juniper Networks or connect with Juniper on [Twitter](https://twitter.com/juniper) and [Facebook](https://www.facebook.com/juniper).

² <https://www.hpcwire.com/off-the-wire/juniper-networks-provide-networking-infrastructure-new-ncar-supercomputer/>

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
Fax: +1.408.745.2100
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
Fax: +31.0.207.125.701

