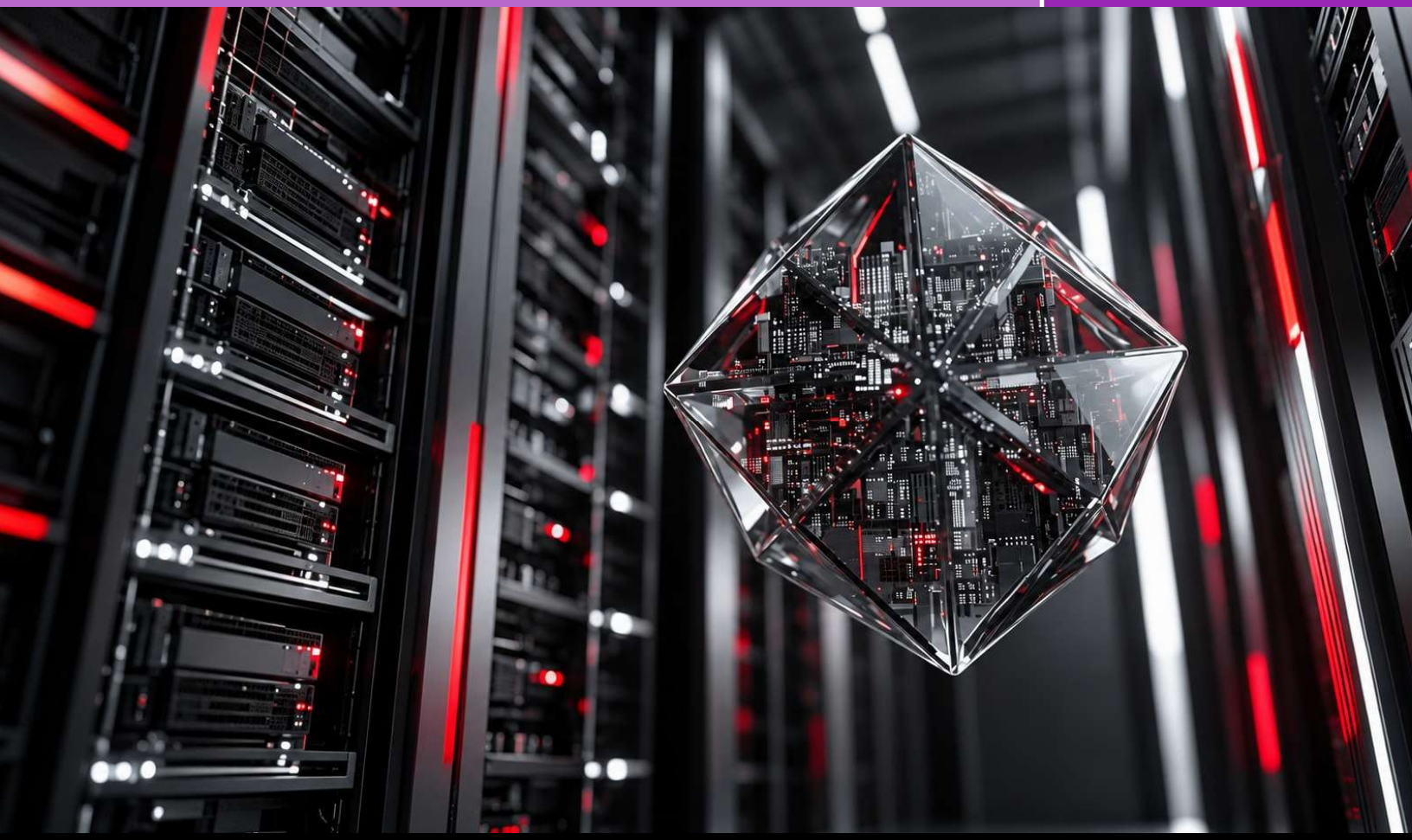
WHITE PAPER



Rethinking AI-Scale Data Center Validation

10 Reasons 1.6T Demands a New Class of Test Systems

KEYSIGHT

AI Workloads Demand Robust, End-to-End Validation

Modern AI-driven workloads are fundamentally changing the operating conditions inside data centers. Extreme bandwidth demands, sensitivity to tail latency, and highly dynamic traffic patterns place sustained pressure on Ethernet networks in ways traditional validation approaches were never designed to handle. As a result, development teams across the industry face increasing risk. Networks that appear to function in the lab can fail to deliver predictable performance in production, leading to extended job completion times, inefficient GPU utilization, and costly rework.

What makes these failures particularly challenging is that they are rarely visible during traditional testing. Links may meet compliance specifications and components may perform correctly in isolation, yet problems emerge once dense fabrics are deployed running real, heavy AI workloads. Understanding how and why Ethernet validation must evolve is now essential to ensuring reliable network performance for AI training and inference infrastructure.

Why Ethernet Validation Must Evolve

AI is reshaping data centers into tightly coupled compute fabrics where high-speed Ethernet networks must deliver unprecedented throughput and predictably low latency. Modern deployments now depend on 1.6T Ethernet, 224G optical lanes, advanced congestion control, and automation as a unified system. This level of integration creates a demand for optimized interoperability across components and layers that extend beyond the scope of traditional validation methods.

As AI workloads amplify east-west traffic, validation challenges shift from individual components to the behavior of the system as a whole. Interconnects, cables, switches, and GPUs no longer operate independently; each becomes a potential point of failure within a tightly coupled fabric. Small physical-layer inconsistencies across electrical and optical domains can cascade into fabric-wide instability once real workloads are applied, even when basic lane-by lane tests appear to meet standards specifications.

Without more comprehensive validation, teams risk deploying networks that appear ready in the lab but break down at scale in production. In AI environments, where transactions occur millions of times per second across thousands of synchronized endpoints, even minor inefficiencies can escalate into significant performance and reliability issues. This is why Ethernet validation must evolve from isolated compliance checks to an end-to-end testing discipline that exposes risk early and protects value of large-scale AI infrastructure investments.

Ten Reasons to Rethink Your AI Validation Approach at 1.6T

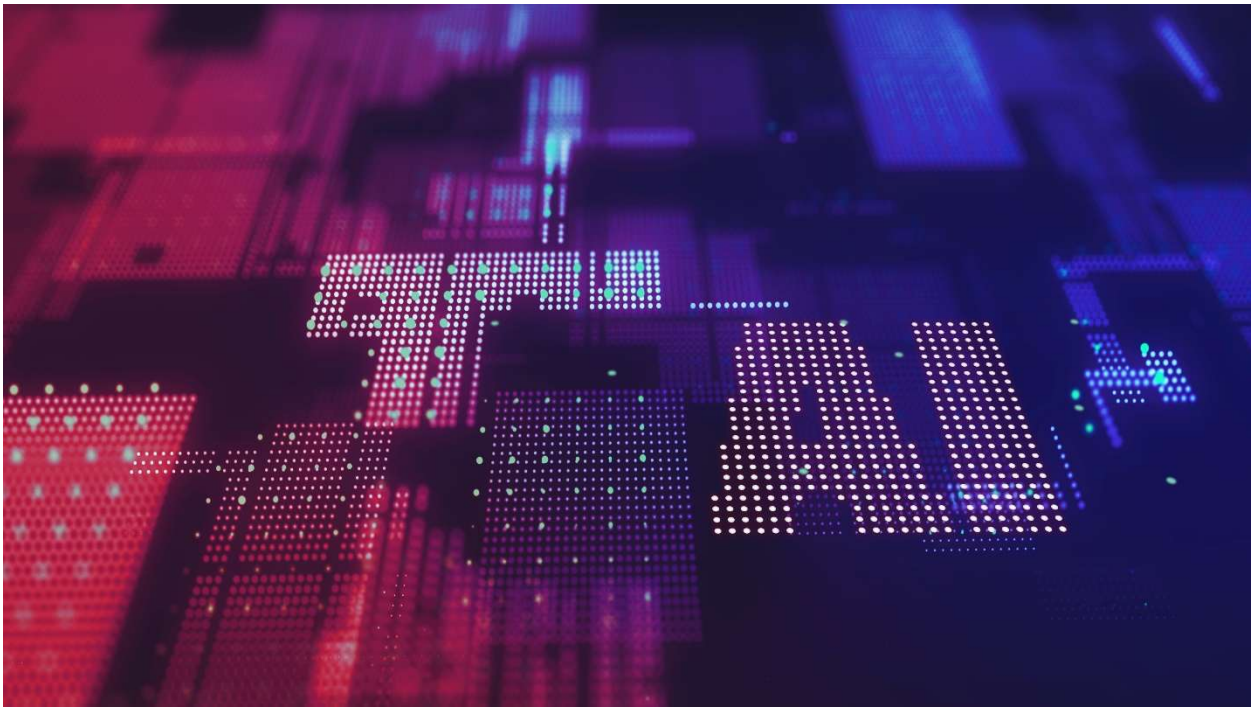
While the need to evolve Ethernet validation is driven by many factors, the ten challenges that follow represent the most pressing issues facing hyperscalers, networking equipment manufacturers, and component developers as AI infrastructures scale. Together, they define the considerations required to move from isolated testing toward an end-to-end validation discipline capable of ensuring reliable, predictable network behavior at scale.

1. Legacy Test Systems May Fall Short at 1.6T

Challenge

AI training workloads drive east-west traffic at data rates and speeds far beyond previous network limits and the capabilities of legacy test systems. Technologies such as 1.6T Ethernet with PAM4 modulation and 224 Gb/s SerDes are rapidly becoming standard in next-generation AI fabrics. Figure 1 shows Ethernet roadmap continues to accelerate, reaching 800 GbE today, with 1.6 TbE in development and 3.2 TbE ahead.

Legacy test platforms which cap at lower lane rates cannot saturate modern fabrics. As a result, they often underrepresent burst behavior or real-world workload dynamics. This shift requires new validation approaches capable of generating sustained line-rate traffic at scale to expose congestion effects, performance bottlenecks, and stability risks before deployment.



Ethernet Speeds

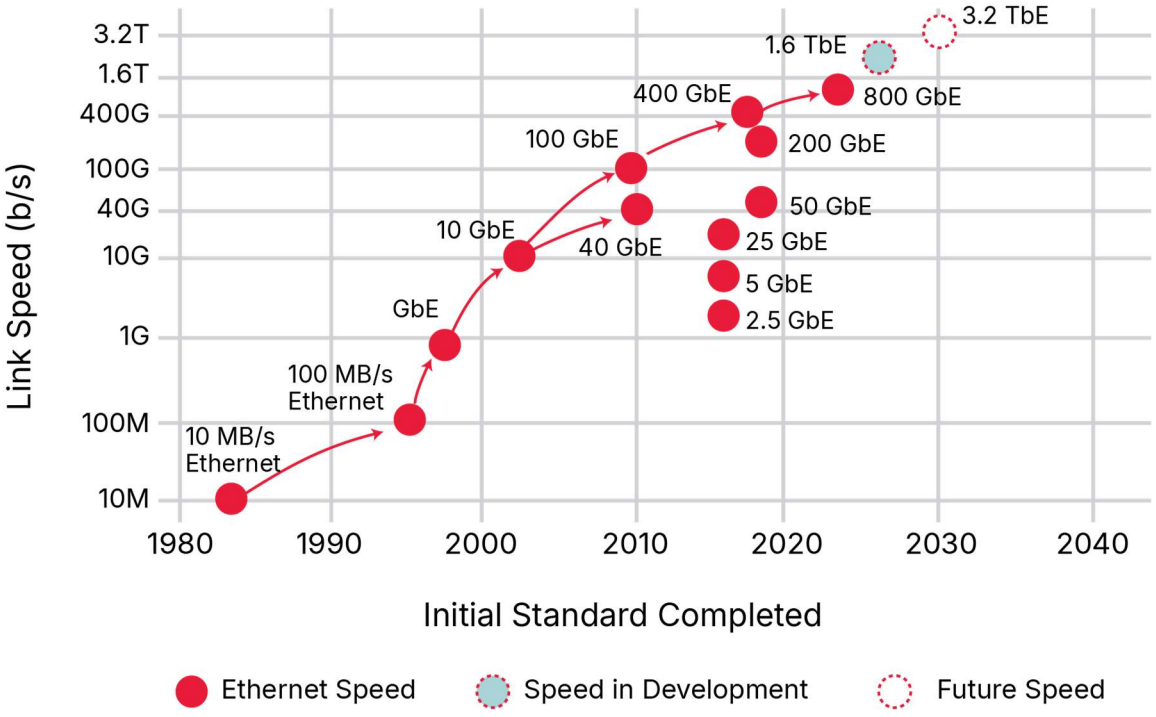


Figure 1. Ethernet link speeds continue to scale rapidly, from 10 Mb/s to 800 GbE today, with 1.6 TbE and 3.2 TbE on the horizon.

Solution

Effective validation at 1.6T requires solutions capable of high-density, ultra-high speed Ethernet traffic generation and analysis. These solutions must drive true line-rate performance across large spine and leaf fabrics. Ethernet traffic generators should sustain traffic at full interface speed, while maintaining accurate timing and packet behavior.

To fully characterize fabric behavior, effective validation must extend beyond peak throughput to include fabric-level benchmarking and congestion characterization. This includes visibility into how traffic behaves across multiple hops, where bottlenecks form, and how congestion propagates across interconnected switches and links. Together, these capabilities allow teams to validate not only how fast the network can run, but whether it can hit and hold line rates under realistic, sustained, and stress operating conditions.

2. Tail Latency, Not Average Latency, Determines AI Performance

Challenge

In tightly synchronized AI training workloads, job completion is governed by the slowest GPU. Long-tail delays can therefore stall progress across the entire workload. Tail latency, not average latency, determines system efficiency which is why even small delays can ripple across the fabric, slow coordinated operations and extend job completion times.

Traffic patterns such as microburst and in cast can rapidly build queues, making congestion response mechanisms essential to keeping latency under control. When tail latency grows unchecked, the impact is not just a slower network, but underutilized idle GPUs and wasted resources. As a result, validation must move beyond averaging metrics.

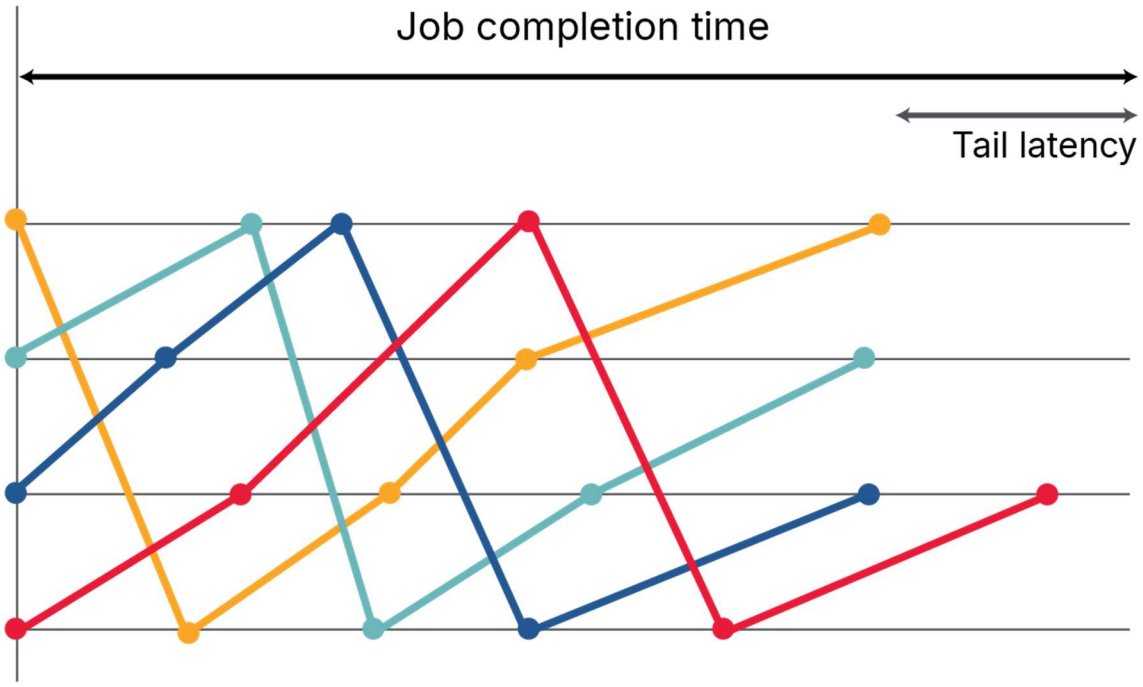


Figure 2. Job Completion Time. Image depicts 4 GPUs processing and sharing information across the cluster. Tail latency is when the last GPU has completed processing time which in the figure above it is the red node.

Solution

Addressing ultra-low latency at scale requires validation solutions that can measure latency behavior under realistic, high-load traffic conditions, rather than relying on synthetic or lightly loaded tests. Solutions should provide visibility into full latency distributions, allowing teams to understand how performance degrades at the tail as traffic patterns shift and congestion develop.

Equally important is end-to-end congestion characterization. Explicit Congestion Notification (ECN) allows network devices to signal congestion before packets are dropped. Data Center TCP (DCTCP) defines how traffic responds to those congestion signals to slow transmission and maintain low latency. Validation should also confirm proper Quality of Service (QoS) behavior, which prioritizes different traffic classes when networks are under load. Effective validation solutions need to be able to test ECN marking, DTCTCP response, and QoS enforcement across switches and NICs.

Together, these capabilities determine whether latency-control mechanism functions as intended when networks are pushed to scale. Without validating how congestion is signaled, how traffic responds, and how queues are managed across the fabric, long tail delays can quietly erode performance and stall AI workloads. Effective end-to-end latency validation is therefore essential – not just to measure performance, but to optimize GPU utilization and training efficiency under real operating conditions.

3. Massive Topologies Introduce Emergent Behavior That Small Tests Miss

Challenge

Large deployments often span multiple racks and include thousands of GPUs and NICs, creating environments with overlapping traffic classes and complex failure domains. As these topologies grow, coordinating behavior across devices becomes increasingly difficult, and even small anomalies can propagate across the larger fabric. In addition, large topologies must balance tail latency, bandwidth, and power consumption for reliable network performance.

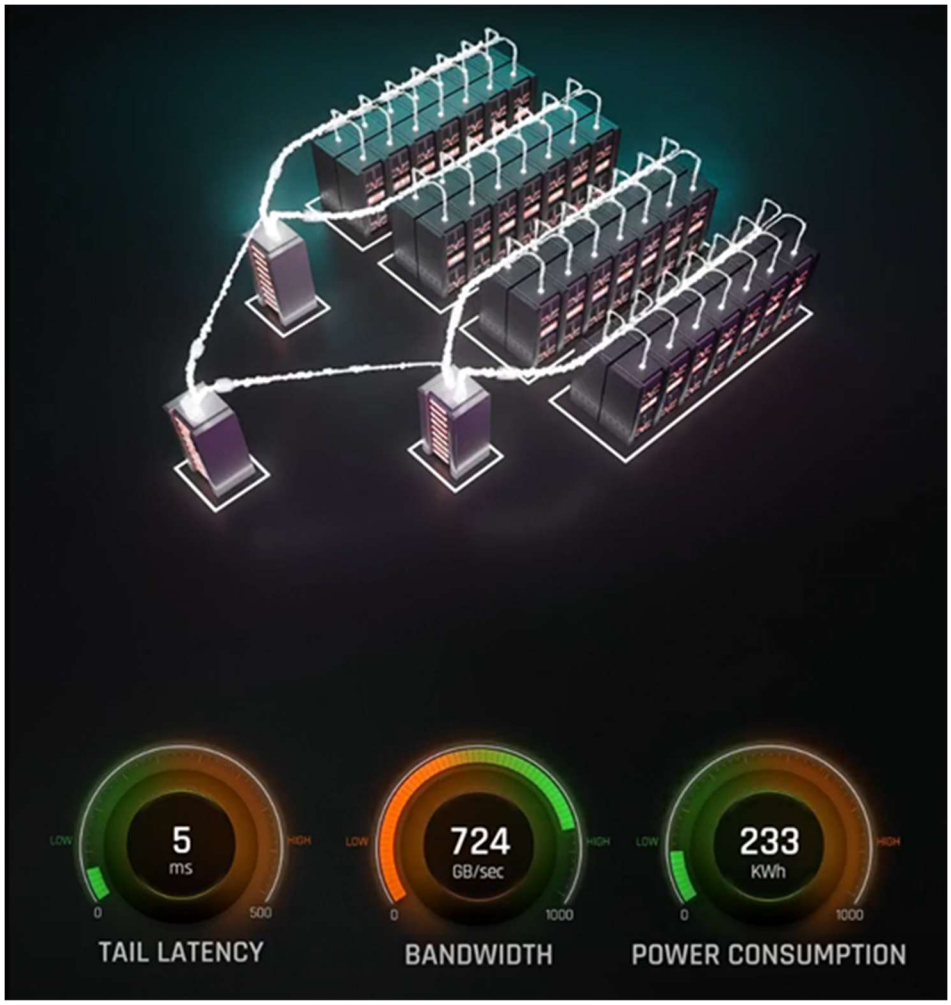


Figure 3. Tail latency, bandwidth, and power consumption are key metrics for evaluating and optimizing data center performance.

Solution

Validating massive scale-out topologies requires solutions that can emulate large, high-port-count environments and accurately represent how traffic behaves across multi-rack fabrics. These solutions should support high-density scale-out emulation, allowing teams to stress networks at realistic sizes without simplifying topology or traffic patterns.

Beyond raw scale, solutions must enable topology modeling and workload simulation, so architects can evaluate how different design choices affect congestion, resilience, and performance before deployment. This includes the ability to explore alternative fabric layouts, traffic distributions, and scaling scenarios in a controlled environment. Finally, robust validation solutions should support fault injection and orchestration across racks, making it possible to observe how failures propagate and confirm that multi-rack coordination and recovery function correctly as the deployment grows.

4. Synthetic Models Fail to Represent Real AI Workloads

Challenge

AI clusters generate highly dynamic, mixed traffic patterns, that include compute synchronization, storage activity, checkpointing, and management of communication. These flows interact and overlap, creating uneven loads and short-lived congestion that rarely appears in simplified, synthetic lab test scenarios.

Performance issues often emerge from interactions among multiple traffic types rather than from any single flow in isolation. Networks that validate well under uniform traffic can behave unpredictably in production, leading to instability and inefficient use of costly compute resources. Effective validation should reflect real production traffic scenarios to assess system-level performance and stability.

Solution

Validating realistic, mixed traffic patterns requires solutions that can generate and control diverse traffic profiles representative of real production workloads, rather than relying on uniform or synthetic streams. Solutions should support customizable flow mixes and traffic ratios, allowing parameter synchronization, storage, checkpointing, and management of traffic to be exercised simultaneously.

To expose stress conditions that commonly occur in large clusters, solutions must also model in cast and outcast scenarios, including short-lived surges and transient congestion events that can destabilize queues and scheduling. In addition, the ability to step through corner cases and burst conditions is critical to verifying stability beyond average behavior. Finally, full validation should account for telemetry and monitoring overhead, ensuring observability mechanisms do not degrade performance under sustained or traffic surges.

Together, these capabilities ensure validation reflects how networks behave in production, not just how they perform under idealized conditions. This is critical to preventing instability, avoiding wasted compute cycles and maintain predicable performance as AI workloads expand.

5. Power and Thermal Stress Can Degrade Network Stability at Scale

Challenge

As AI deployments scale in size and complexity, power and thermal conditions increasingly influence how networks behave in production. High-speed optics, accelerators, and dense cabling significantly increase power draw and heat dissipation across racks, pushing data centers closer to their physical limits and potentially degrading performance in ways that are not obvious under lab conditions.

Elevated temperatures, constrained airflow and reduced power margins can quietly degrade throughput and destabilize links over time. Without understanding how networks behave as thermal and power conditions change, teams risk deploying systems that perform well initially but become unpredictable and fragile as utilization increases.

Solution

Validating power and thermal resilience requires solutions that can drive sustained, high-load traffic while correlating network performance with environmental stress conditions. Effective solutions should enable teams to observe how throughput, latency, and stability behave as temperature increases, airflow becomes constrained, or power margins tighten. These conditions are common in dense, high-performance deployments.

To accurately reflect real operating scenarios, validation must go beyond short-duration tests and support extended, line-rate operation that exposes thermal throttling effects and power-related degradation over time. Visibility into how performance metrics shift under sustained load allows teams to identify emerging risk early and ensure that networks remain stable and predictable before production deployment.

6. Interconnects Constrain System Reliablity at 224 Gb/s

Challenge

At today's link speeds, interconnects are no longer passive components but active determinants of system reliability. As data rates climb and tolerances narrow, the physical connection between devices increasingly defines whether an AI fabric operates reliably or degrades under load.

This raises the stakes for teams building AI infrastructure because interconnect issues are often intermittent and difficult to diagnose. Small variations introduced during installation, handling, or normal wear can destabilize links only after deployment, triggering unpredictable behavior that ripples across tightly coupled systems. Without confidence in interconnecting integrity under real operating conditions, networks risk costly downtime, hidden performance loss, and the "mystery flaps" that erode trust in production environments.

Why Interconnect Validation Matters

Interconnect validation help prevents "mystery flaps" in production—intermittent link failures that may go undetected in the lab but emerge under real-world traffic stress.

Solution

Ensuring interconnect integrity at high SerDes rates requires validation approaches that move beyond isolated signal-integrity measurements and assess functional link behavior under realistic conditions. Validation solutions should enable rapid, repeatable testing of cables and connectors while links are active, allowing teams to evaluate stability, error behavior, and margin under conditions that reflect handling, installation, and deployment stress.

These solutions must provide visibility into BER, link margins, and signal stability as physical factors such as vibration, bend radius, and connector wear are introduced. Equally important is the ability to validate performance outside controlled lab environments, ensuring that interconnects maintain reliable operation once deployed in dense racks and production fabrics. Together, these capabilities help close the gap between laboratory characterization and real-world performance.

7. PHY Standards Compliance Does Not Guarantee Interoperability at 1.6T

Challenge

As Ethernet standards advance toward 1.6T, electrical and optical interfaces are being pushed to tighter margins with less tolerance for variation. IEEE and OIF specifications continue to evolve across PAM4 signaling, FEC schemes, and link training, raising the bar for interoperability and correctness at every layer of the physical interface.

This shift is challenging because compliance issues often emerge only as systems scale. Designs that meet baseline requirements in isolation can fail to interoperate reliably once deployed across dense fabrics with mixed components and real traffic. Without rigorous standards-aligned validation, teams risk deploying hardware that technically complies, yet behaves unpredictably under production conditions.

Solution

Readiness for emerging Ethernet standards requires validation solutions that enable comprehensive electrical and optical PHY compliance coverage aligned with evolving IEEE and OIF requirements. These solutions must assess signal integrity, error performance, and interoperability together, rather than validating each element independently during a one-time test.

Equally important is validating link training, auto-negotiation, and protocol behavior as standards mature. Automated workflows and standards-aligned reporting are essential to maintaining confidence as margins narrow and specifications evolve. Together, these capabilities allow teams to validate not just compliance, but real interoperability at scale.

8. Resilience Must Be Proven Under Stress

Challenge

AI networks are expected to maintain predictable performance even as failures, congestion events, or malformed traffic occur. At scale, isolated disruptions can propagate quickly, making reliability and resilience of core requirements rather than secondary considerations.

This matters because failures rarely occur under ideal conditions. Packet loss, link instability, policy misconfiguration, or unexpected traffic patterns can undermine throughput and stability if recovery mechanisms are not exercised in advance. Without validating behavior under stress, teams risk discovering weaknesses only after deployment, when impact and recovery cost are highest.

Solution

Ensuring reliability, resilience, and security requires validation approaches that combine realistic traffic with controlled fault scenarios and continuous observations. Validation solutions must exercise failure recovery, policy enforcement, and stability mechanism together, allowing teams to see how the fabric responds when components degrade or abnormal conditions arise.

By correlating traffic behavior with fault events, these approaches confirm that recovery mechanisms function as intended, and that performance remains predictable under disruption. This system-level validation is essential to building confidence that AI fabrics will remain stable and secure in real-world operations.

9. Manual Validation Cannot Keep Pace with Continuous Updates

Challenge

AI infrastructures evolve continuously, with frequent configuration changes, software updates, and scaling events across multi-rack environments. Manual validation processes cannot keep pace with this rate of change.

The risk is not just slower testing, but inconsistent validation outcomes. Without repeatable and automated processes, regression can slip through unnoticed, leading to performance drift and operation instability over time.

Solution

Keeping pace with rapid change requires validation solutions that are automated, repeatable, and integrated into CI / CD workflows. Such solutions should support API-driven orchestration, enabling validation to be executed automatically as configurations change rather than relying on manual intervention.

Analytics play a critical role by detecting regressions and behavioral drifts across test runs, turning raw results into actionable insights. Pre-deployment modeling further strengthens this approach by exposing scaling risks before changes reach production. Together, automation and analytics transform validation into a continuous discipline, reducing risk while accelerating deployment velocity.

10. Validation Must Be Flexible to Scale Beyond 1.6T Ethernet

Challenge

As standards evolve, validation approaches must evolve with them. Higher SerDes lane rates and new optical form factors are already emerging, placing ongoing pressure on networks to support increasing performance and complexity.

This is challenging because static test environments quickly become obsolete. Tools and processes designed for one generation of technology can limit adoption of the next, forcing teams into costly replacement cycles or delaying innovation. Future-ready validation requires flexible, modular methodologies that can adapt across technology generations while preserving confidence at scale.

Solution

Future-proofing Ethernet infrastructures require validation solutions designed with headroom, modularity, and adaptability in mind. Modern solutions must support current requirements while providing a clear path to validate higher speeds, new form factors, and more complex architecture as standards advance.

A holistic validation methodology- spanning workload behavior, fabric performance, interconnect integrity, and physical-layer compliance – ensures long-term compatibility and reduces risk. By aligning validation workflows with technology roadmaps, teams gain the flexibility to adopt next-generation Ethernet confidently without slowing innovation.

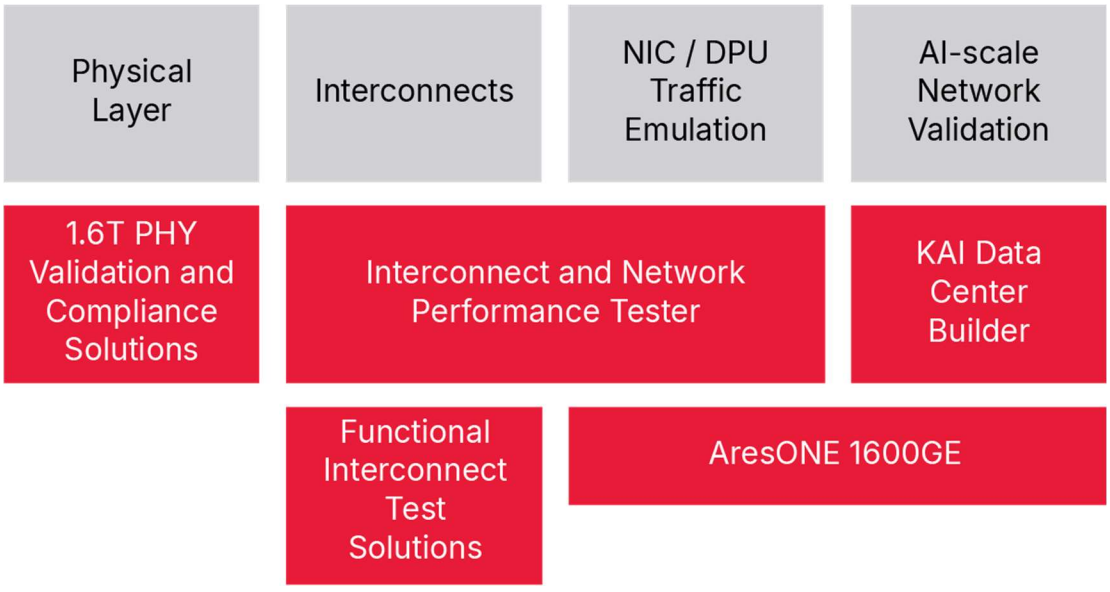


Figure 4. End-to-end validation at every layer

Validation Capabilities and Solutions

Capabilities Needed for End-to-End Validation

Taken together, the ten drivers outlined above point to a common conclusion: AI-scale Ethernet networks cannot be validated through isolated tests. Addressing these challenges requires a cohesive, end-to-end validation workflow that spans physical layer compliance, interconnect integrity, fabric-level performance benchmarking, and realistic traffic emulation.

To meet the demands of AI scale, validation workflows need high-density traffic generation and analysis capable of sustaining true line rate while exposing burst behavior, fairness across flows, and loss under mixed conditions. On top of that, teams need latency and congestion characterization that captures full distributions and traces ECN / DCTCP signaling and queue depth dynamics when workloads shift.

Reducing risk before deployment requires system-level modeling. Digital twins help evaluate topology choices, scaling decisions, and potential congestion hot spots ahead of lab or production changes. Since physical links can become the limiting factor at high speeds, functional interconnect testing is required to correlate signal integrity of lab insights with real-world BER and stability under handling and thermal variation.

Finally, readiness at AI scale depend on automation across the validation lifecycle. Automated electrical and optical PHY compliance helps ensure alignment with evolving IEEE and OIF standards, Day-to-day reliability hinges on CI / CD-driven automation and analytics that make tests repeatable, generate evidence automatically, and flag regressions as environments evolve.

Solution Mapping

Here is a summary of Keysight solutions that address each of the ten end-to-end validation capabilities:

Reasons	Required Capabilities	Solutions
Unprecedented Bandwidth	High-density traffic generation; burst and fairness characterization	1600GE workload emulation with AresONE 1600GE
Ultra-Low Latency	Latency CDFs; ECN / DCTCP; queue depth	• Network emulation and performance benchmarking: AresONE 1600GE and KAI Data Center Builder • 1.6T traffic generation for component validation at L2 – 3: Interconnect Network Performance Tester
Scale-Out Topologies	Port density emulation; digital twin	• AresONE 1600GE • KAI Data Center Builder
Mixed Traffic Patterns	Realistic flow mixes; incast / outcast; telemetry impact	AresONE 1600GE
Power / Thermal	Sustained loads vs temp; throttling, PSU margin	AresONE 1600GE
Interconnect Integrity	BER validation and FEC performance per lane	Multiple lane validation with Functional Interconnect Test System and INPT-800GE / 1600GE
1.6T Compliance	PAM4 eye / jitter; FEC; link training; Tx / Rx	• Pro oscilloscopes (> 90 GHz) • High-performance BERTs (120 GBd) • Sampling oscilloscopes with 120 GBd clock recovery • 1.6T optical test automation software
Reliability / Security	Loss recovery; failover / storm / ACL	AresONE 1600GE
Automation / Analytics	CI / CD; APIs; reporting, regression	• AresONE • KAI DC builder
Future-Proofing	Headroom; modularity; roadmap alignment	Keysight ecosystem

Conclusion

Validating AI-driven Ethernet networks requires a shift in data center performance and reliability test workflows. In AI infrastructures, small inefficiencies, whether in latency, interconnect integrity, or congestion behavior, can cascade across thousands of endpoints, extending job completion time and leaving expensive compute resources underutilized.

The ten drivers outlined in this paper illustrate why traditional validation approaches fall short at AI scale. As bandwidth increases, traffic patterns become more dynamic, and systems grow more tightly coupled. Confidence can no longer come from validating components in isolation. Ensuring system reliability must come from understanding how the entire system behaves under load, under stress, and at operational scale.

End-to-end validation addresses this reality by spanning every layer of the system, from electrical and optical PHY compliance to fabric level performance, congestion behavior, and automated workflows that keep pace with continuous change. By validating these domains together, teams can expose risk earlier, correlate issues across layers, and assess readiness under conditions that closely mirror production.

This system-level approach enables predictable performance, faster deployment cycles, and better utilization of high-value AI infrastructure. At AI scale, end-to-end validation is no longer a nice to have – it is the foundation for building Ethernet networks that scale with confidence.

Learn more about end-to-end validation solutions for AI infrastructure at:

- https://www.keysight.com/us/en/cmp/ai-scale-out.html
- https://www.keysight.com/us/en/learn/hubs/keysight-ai-products.html
- https://www.keysight.com/us/en/learn/hubs/ai-technology.html

Keysight enables innovators to push the boundaries of engineering by quickly solving design, emulation, and test challenges to create the best product experiences. Start your innovation journey at www.keysight.com.

KEYSIGHT

This information is subject to change without notice. © Keysight Technologies, 2026, Published in USA, October 14, 2026, 7126-1056.EN