

## Ethernet Won't Collapse the Four Enterprise Networks Into One

Access, Front-End, Back-End, and Scale-Up Networks May Converge on Ethernet, but Network Teams Own Only Half of the Map

Analyst(s): Tom Hollingsworth

Publication Date: October 13, 2026

Document #: AINTH202609

### Key Points

- The enterprise network is now four systems – the access network, the front-end data center fabric, the AI back-end fabric, and the GPU scale-up domain. Each is defined by the payload it carries.
- Ethernet already carries the first two, has taken most of the AI back end, and is contesting the scale-up domain through the ESUN and UALink specifications.
- Protocol convergence won't produce one network. The back-end and scale-up versions of Ethernet behave nothing like a front-end fabric, and network teams fully own only two of the four.

### Recommendations

Networking vendors should build for four networks with four buyers, because the protocol may converge while the purchase decisions don't.

1. **Sell Into the Rack Decision, Not After It:** The scale-up interconnect and much of the front-end design, are locked in when the GPU system is chosen. Vendors that wait for the network team's refresh cycle will find those choices already made by whoever signed for the compute.
2. **Make the Handoffs the Product:** The boundaries between front-end, back-end, and scale-up fabrics are where training and inference problems surface, and they're the one area every buyer shares. Telemetry, troubleshooting, and automation that span those interfaces give vendors a durable role even in networks they don't supply.
3. **Treat Familiar Tooling as an Entry Point, Not a Guarantee:** ESUN and UALink lean on Ethernet operating systems, gNMI, and Redfish, which gives incumbent network

vendors a credible path into the rack. That path only pays off with partnerships among GPU and system builders, because the scale-up fabric still ships bundled with compute.

## What You Need to Know

- For roughly two decades, most enterprises described their footprint as one network. AI deployments have split it into four: the access network, the front-end data center fabric, the AI back-end fabric that links GPU servers, and the scale-up domain that links GPUs inside a rack or pod.
- The Ultra Ethernet Consortium published its [1.0 specification in June 2025](#), adding packet spraying, packet trimming, and link-level retry for back-end fabrics. OCP's ESUN workstream, driven by Meta and Microsoft, approved [ESUN 1.0](#) in early 2026, which [replaces 28 to 48 bytes of IP and UDP headers with a 4-byte header](#) for scale-up traffic.
- The UALink Consortium released its [1.0 specification](#) in April 2025, defining 200G lanes for pods of up to 1,024 accelerators, and followed with version 2.0 in April 2026. The consortium expects [1.0 silicon in labs in 2H 2026 and products in 2027](#).
- NVIDIA rates NVLink at [1.8 TB/s per GPU](#) in Blackwell NVL72 racks and 3.6 TB/s in Vera Rubin, bidirectional figures that work out to roughly 14.4 Tb/s per GPU in one direction. NVIDIA's networking revenue reached [\\$10.98 billion in fiscal Q4 2026](#), up 263% year over year, and CFO Colette Kress [attributed that growth primarily to NVLink scale-up switches](#).

## Analysis

Ethernet is likely to end up carrying most of the four enterprise networks, and that still won't make them one network. The access network, the front-end fabric, the AI back end, and the GPU scale-up domain share a protocol name, but they don't share a purpose, a failure model, or an owner. The further up that stack you go, the less of it the network team actually controls.

## Payload Draws the Line Between the Four Enterprise Networks

The cleanest way to separate these networks is to ask what each one moves. The access network moves user sessions, and the front end moves application packets. The back end moves memory transfers between GPU servers working on the same job, and the scale-up domain moves memory reads and writes between the GPUs themselves. Storage fabrics and out-of-band management networks are additional networks to build, but each behaves like the front end or the back end, so they are assumed to be included in these four enumerated types.

As the payload shifts from sessions to memory, loss tolerance drops from generous to nearly zero, and latency budgets shrink from milliseconds to well under a microsecond. A dropped

frame on Wi-Fi costs someone half a second of frozen video. A congested back-end link stalls a collective operation, and a cluster of very expensive GPUs sits idle waiting on the slowest participant. That's why the Ethernet that won the back-end sprays packets and trims payloads. It's still Ethernet from a protocol perspective, just taught to behave like a very different network.

Figure 1: The Four Enterprise Networks Compared

### Four Networks, Compared

Moving left to right, the payload shifts from people's sessions to memory, and network team control falls away.

|                            | Access                            | Front-end Fabric                                 | Back-end Fabric                               | Scale-up Domain                               |
|----------------------------|-----------------------------------|--------------------------------------------------|-----------------------------------------------|-----------------------------------------------|
| <b>What Moves</b>          | User sessions<br>Video, SaaS, IoT | Application Packets<br>North-South and East-West | RDMA Transfers<br>Collective Operations       | Memory Reads/Writes<br>Load/Store Semantics   |
| <b>Endpoints</b>           | People & Devices                  | Servers, Storage, Internet                       | GPU Servers                                   | GPUs                                          |
| <b>Domain Size</b>         | Buildings & Sites                 | Data Center                                      | GPU Cluster                                   | 72 GPUs (NVLink)<br>Up to 1,024 (UALink spec) |
| <b>Loss Tolerance</b>      | Generous                          | Moderate                                         | Near Lossless<br>One slow link stalls the job | Effectively None                              |
| <b>Latency Sensitivity</b> | Human Scale                       | Application Tolerant                             | Tail Latency Bound                            | Bus-like                                      |
| <b>Dominant Technology</b> | Ethernet & Wi-Fi                  | Ethernet Leaf-Spine<br>ECMP, BGP, EVPN           | Ethernet (RoCE, UEC)<br>and InfiniBand        | NVLink<br>UALink, ESUN emerging               |
| <b>Who Owns It</b>         | Network Team                      | Network Team                                     | Shared<br>Often arrives with the cluster      | Ships Inside the Rack<br>Bought with Compute  |
| <b>Payload</b>             | People's sessions                 |                                                  |                                               |                                               |
| <b>Network Team</b>        | Full control                      |                                                  |                                               |                                               |

Sources: NVIDIA NVLink product page (960, 72 domain size), UALink Consortium, UALink 2000, 1.0 Specification, April 2025 (1,024 accelerator pack); Dell'Oro Group AI Back-end Networks reports (back-end definition, Ethernet and InfiniBand), OCP ESUN 1.0, 2025. Loss and latency rows are qualitative. Ownership row reflects the author's analysis for typical enterprises, not survey data. Very large column footnotes are the exception.

**Futurum**

Source: Futurum Research, September 2026

## Ownership Ends Halfway Up the Stack

The network team owns the access network and the front-end fabric end-to-end. Nobody argues about who's responsible when the Wi-Fi is slow. The back-end is where that certainty starts to blur, because some enterprise network teams design it like any other fabric, while others receive it inside a cluster design they never specified.

The scale-up domain is barely a network in the sense most of us mean. UALink calls its specification a [memory-semantic interconnect](#), which is a computer's internal bus stretched to the size of a rack. NVIDIA tying its networking growth to NVLink switches shows how that bus gets purchased, as part of a [rack-scale system](#). Whoever signs for the GPUs buys the network inside them.

## The Best Case for One Network

The strongest counterargument is that everything converges on Ethernet anyway. If ESUN succeeds, one family of switch silicon and optics could span all three data center networks. Look at what ESUN gave up to get there, though. Dropping IP and UDP works because a scale-up domain's topology is fixed at deployment and needs no routing. That's Ethernet framing without IP networking, and it behaves nothing like your front-end fabric.

The deeper version goes after ownership. ESUN lists reuse of Ethernet operating systems and tooling as a goal, and UALink's manageability specification [uses gNMI and Redfish](#). Those are tools network engineers already run, and they're the most plausible path back into the rack. Familiar tools lower the barrier to operating Network 4. They don't change who buys it. As long as the scale-up fabric ships bundled with compute, the purchase decision and most of the control sit with whoever signs for the GPUs. The enterprises that build their own clusters, in financial services, pharma, and sovereign AI programs, are the carve-out, and they need four operating models instead of two.

## Count to Four

Ethernet winning more of these networks is likely. Collapsing them into one isn't. They share a name with campus switches and not much else. For network teams, the useful work is knowing where their map ends, which networks they own, which they depend on, and where the handoffs sit. Those interfaces are where problems will surface.

## What to Watch

- **UALink Silicon Timing:** The consortium expects 1.0 silicon in labs in 2H 2026 and products in 2027. A slip extends NVLink's lead and keeps scale-up purchases tied to NVIDIA racks, while multivendor systems shipping on time would give enterprises their first real scale-up choice.
- **ESUN Adoption and Tooling Reach:** Watch whether switch vendors ship ESUN silicon and whether existing network operating systems extend into the rack. If network teams start running scale-up fabrics with gNMI-based tools, the ownership line in this analysis moves up the stack. A specification that ships only in hyperscaler fleets would leave the enterprise map unchanged.
- **NVLink Fusion Terms:** NVIDIA has [opened NVLink to partner CPUs and accelerators](#), but trade press has reported those chips [must still connect to an NVIDIA product](#). Loosening that requirement would weaken the bundling argument, while keeping it confirms the door opens only as far as NVIDIA's own silicon allows.

- **Enterprise On-Premises Buildouts:** Track whether rack-scale AI systems in financial services, pharma, and sovereign programs start flowing through network procurement. That would signal network teams gaining a vote earlier in the purchase cycle. Continued enterprise growth on neoclouds would point the other way and deepen the dependency.
- **Scale-Across Networking:** Broadcom has begun describing [a scale-across domain](#) that links entire data centers. Whether it becomes a fifth network or an extension of the back end will test whether payload remains the right dividing line.

For more information, [review the ESUN 1.0 specification overview](#).

## Other Insights from Futurum

[Scale-Across AI Networking: The Third Dimension of AI Infrastructure Design](#)

[Co-Packaged Optics: The Key to Unleashing AI Networking's Full Potential](#)

[The 1.6TbE Transition Will Reshape Data Center Architecture from the Ground Up](#)

## About Us

### About the Author

Tom Hollingsworth, Research Director, covers Networking for The Futurum Group. He is an expert in data center networking technologies with over 20 years of experience. He is CCIE Emeritus #29213. He helps clients understand the technologies behind networking and how it can enable technology shifts toward new advances. He can be reached at [thollingsworth@futurumgroup.com](mailto:thollingsworth@futurumgroup.com).

### About The Futurum Group

Every day, The Futurum Group's analysts, researchers, and advisors help business leaders worldwide anticipate tectonic shifts in their industries and leverage disruptive innovation. Unlike traditional analysts, The Futurum Group works not only in analysis and research but also takes that insight and knowledge even further, engaging all the way through the go-to-market process.

Futurum Research provides in-depth research and insights on global technology markets using advisory services, custom research reports, strategic consulting engagements, digital events, go-to-market planning, and message testing. It also creates, distributes, and amplifies rich media content that all stakeholders read, watch, and listen to.

See more details on The Futurum Group at [futurumgroup.com](http://futurumgroup.com).

## Copyright & Use License

### Copyright Notice

Copyright ©2026 by The Futurum Group, LLC. All rights, including that of translation into other languages, are specifically reserved. No part of this publication may be reproduced in any form, stored in a retrieval system, or transmitted by any method or means, electrical, mechanical, photographic, or otherwise, without the express written permission of The Futurum Group [futurumgroup.com](http://futurumgroup.com). United States copyright laws and international treaties protect this publication. Unauthorized distribution or reproduction of this publication, or any portion of it, may result in severe civil and criminal penalties and will be prosecuted to the maximum extent necessary to protect the publisher's rights.

### License Notice

This document may be distributed within the licensed organization only.

The following acts are prohibited:

Transmittal to others outside your immediate organization including partners, resellers, external consultants, etc. in any media format

Posting on a website which is accessible to others outside your immediate organization

The possession or use within an unlicensed organization

### Limitation of Liability Notice

The information contained herein has been obtained from sources believed to be reliable. The Futurum Group shall have no liability for errors, omissions, or inadequacies in the information contained herein or for interpretations thereof. The reader assumes sole responsibility for the selection of these materials to achieve their intended results. The opinions expressed herein are subject to change without notice.