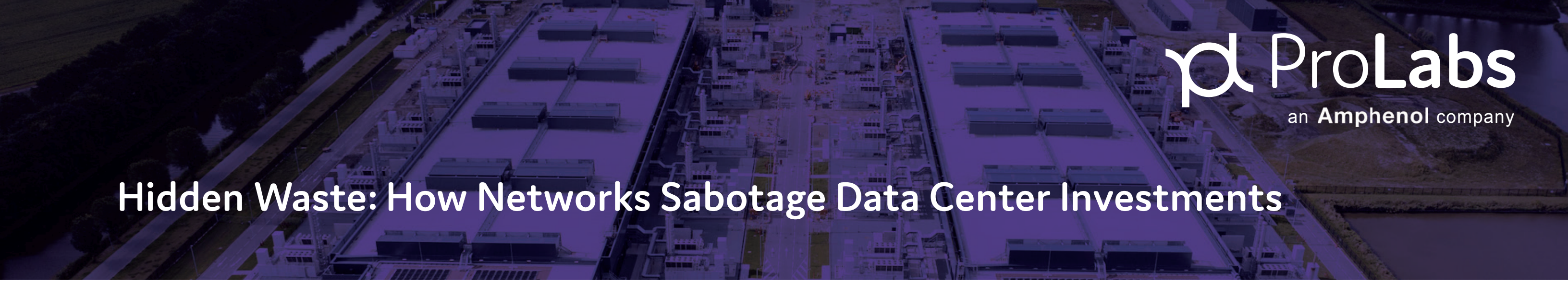


Hidden Waste: How Networks Sabotage Data Center Investments

With expectations of increasing server performance, comes incremental performance pressure. This has now become the norm of data center operations.

But are you overlooking something critical?

At ProLabs, we've observed a troubling reality: while organizations invest millions in cutting-edge servers, storage, memory, and applications, underperforming networks often become the hidden bottleneck that can waste valuable resources.

As data centers face mounting challenges, the consequences of network inefficiencies have never been more severe - transforming what might seem like minor performance issues into significant operational and financial liabilities.

The stark truth is that every millisecond of latency, every packet dropped, and every bandwidth constraint ripples through the entire infrastructure ecosystem, effectively negating any investments made elsewhere. Removing inefficiencies in network optimization isn't just good practice - it's essential to prevent the systemic waste of resources that underperforming networks inevitably cause.

Computing Resources: Powerful but Paralyzed

With high-performance servers comes multiple CPU cores, specialized GPUs, and vast amounts of HBM and RAM representing significant capital investments.

However, these powerful computing resources sit idle or underutilized when network bottlenecks prevent data from flowing efficiently. Consider an AI training environment where multi-million-dollar GPU clusters wait milliseconds - which can quickly accumulate into minutes, hours and days - for data because network throughput can't keep pace.

Or cloud services where customer experience suffers because application response times are dominated by network latency rather than actual processing time. These scenarios represent pure resource waste, with expensive computing assets delivering only a fraction of their potential value.

At ProLabs, we've seen cases where network upgrades effectively "unlock" computing investments that were previously held hostage by network constraints.

Energy Consumption: Paying for Power Without Productivity

Perhaps the most waste caused by underperforming networks is energy consumption. Data centers consume enormous amounts of electricity—often 10-20 megawatts for a mid-sized facility—yet network bottlenecks mean this power is frequently used without proportional productivity.

When high-performance computing resources idle while waiting for data to navigate inadequate and misunderstood networks, they continue consuming significant electricity. Existing cooling systems work at full capacity regardless of actual computational throughput.

The result is an infrastructure drawing power at near-maximum levels while delivering a fraction of its potential output. Analysis shows that optimizing network performance can reduce effective per-transaction energy usage by up to 45% [2]—not by reducing raw power consumption, but by dramatically increasing the productive work performed per kilowatt-hour. In an era of rising energy costs and sustainability mandates, this inefficiency has become financially and ethically untenable.

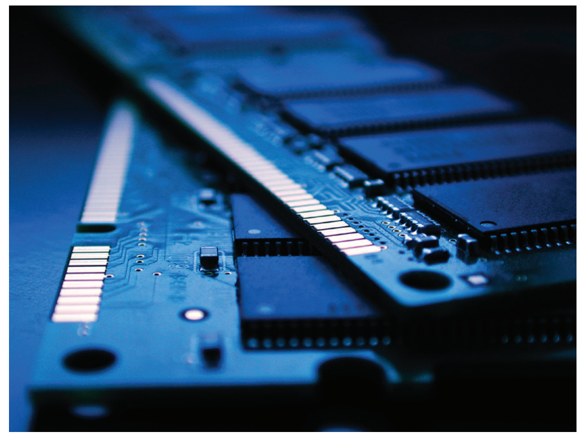
Storage Systems: Capacity Without Accessibility

Organizations invest heavily in high-performance storage systems, from all-flash arrays to sophisticated object storage platforms. And with data centers adopting the latest PCIe Gen5 SSDs, with higher speeds and capacities, there is an expectation that storage has to be bigger and faster, to cope with growing demands.

Yet any investments are squandered when network constraints create access bottlenecks. Storage that can internally deliver millions of IOPS becomes effectively no better than legacy systems when network paths can't carry this throughput. This limitation is particularly evident in backup and recovery scenarios. We've observed organizations with petabyte-scale environments where recovery time objectives (RTOs) are completely dominated by network transfer rates rather than actual storage performance.

The result is extended downtime during critical recovery operations—sometimes costing tens of thousands of dollars per minute—despite massive investments in "fast" storage systems.

Memory Resources: Expensive and Underutilized



Modern distributed applications rely heavily on memory caching, in-memory databases, and rapid data movement between memory regions. When network performance lags, these memory resources remain partially filled or contain stale data, undermining their core purpose.

In financial services, where microseconds can mean millions, trading platforms invest in enormous memory resources to maintain competitive advantage, only to have much of this benefit negated by network latency and jitter. The waste extends beyond the hardware itself to the opportunity cost of transactions not executed optimally.

The AI Amplification Effect: Magnifying Network Deficiencies

The explosive growth of AI workloads has dramatically amplified the consequences of network underperformance. Machine learning training and inference involve massive data movements that stress network infrastructure in unprecedented ways:

- **Training data transfer:** Moving petabytes of training data from storage to compute nodes
- **Real-time inference:** Delivering model results with strict latency requirements
- **Mixed environments:** Mixed data sets being transmitted over legacy network infrastructure

When networks can't meet these demands, the results are severe:

- Training jobs that should take hours extend to days
- Model accuracy suffers as batch sizes are reduced to accommodate network limitations
- Real-time applications fail to meet SLAs despite sufficient computational resources

Most concerning is how network bottlenecks in AI workloads create a destructive feedback loop: constrained networks lead to longer training times, which increase the overall energy consumption and cost, while simultaneously reducing the ROI on expensive GPU investments that sit idle.

Security Consequences: Performance Compromises Creating Vulnerabilities

Network performance challenges often lead organizations to make architectural compromises that introduce security vulnerabilities. When faced with application performance issues stemming from network constraints, you may find teams implementing workarounds that undermine security protocols :

- Disabling encryption to increase performance
- Relaxing microsegmentation to reduce hop counts
- Bypassing inspection points to improve throughput
- Creating direct network paths that circumvent security controls

These compromises represent another form of waste, and the squandering of investments in security infrastructure and the introduction of risks that may lead to costly breaches. With regulatory standards, and their impacts on penalties and becoming more stringent – with NIS2 and DORA for the financial sector specifically –prevention of workarounds to increase network performance are no longer an option – they are a necessity.

A comprehensive approach to network performance eliminates the perceived need for these dangerous trade-offs.

The Path Forward: Network Optimization as Resource Reclamation

Facing these mounting challenges, forward-thinking organizations are adopting a "leave no stone unturned" approach to network infrastructure, treating network optimization as a form of resource reclamation rather than merely a technical exercise.

1. Comprehensive Network Visibility
   You can't optimize what you can't measure. Modern network performance tools provide granular visibility into traffic patterns, bottlenecks, and microbursts that traditional monitoring might miss. This visibility is the foundation for identifying where network constraints are wasting computational, storage, and memory resources.

2. Strategic Transceiver and Cabling Optimization
   Often overlooked, transceivers and cabling represent critical components that can dramatically impact network performance. Organizations frequently retain older, slower, or less reliable connectivity components while upgrading other infrastructure elements, creating severe mismatches in capability.

   Our compatibility testing shows that upgrading from generic to precision-engineered transceivers can reduce bit error rates by orders of magnitude, immediately improving effective throughput and reducing retransmissions that waste both network capacity and computational resources.

3. Topology Reassessment and Modernization
   Legacy network topologies designed for north-south traffic flows often create artificial constraints when applied to modern east-west dominated workloads. Leaf-spine architectures, disaggregated networking, and software-defined approaches can dramatically improve resource utilization by aligning network capabilities with actual traffic patterns.

4. Protocol Optimization and Tuning
   Default protocol configurations rarely deliver optimal performance in specific environments. TCP window sizes, buffer allocations, and flow control mechanisms need environment-specific tuning to maximize throughput and minimize latency. As a result, it effectively "finds" network capacity that was previously wasted due to suboptimal protocol configurations.

Network Excellence as Resource Stewardship

As data center challenges continue to grow in scale and complexity, the organizations that thrive will be those that recognize network performance as the critical factor that either enables or constrains the value of all other infrastructure investments.

Every element of the network stack—from physical cabling to protocol configuration—deserves careful consideration to ensure it isn't quietly undermining the massive investments made elsewhere in the technology ecosystem. At ProLabs, we have decades of experience with data centers and on-prem to ensure that networking infrastructure contributes positively to overall resource efficiency.

Our commitment to providing high-performance, precision-engineered, compatible alternatives to OEM and NEM components gives organizations the flexibility to optimize their networks without being constrained by artificial vendor limitations or unnecessarily expensive proprietary solutions.

In an environment where performance demands are coupled with progress, and efficiency and regulatory mandates tighten, leaving "no stone unturned" in network optimization isn't just a technical best practice—it's a fundamental that directly impacts the bottom line, reputation and environmental footprint of every data center and large server operation.

Speak to one of our experts, about how you can achieve efficient performance, through smart network decisions.

Footnotes:
1. Hitachi Energy, AI driving a 160% increase in data center demand and $57B investment in 2024.
https://www.hitachienergy.com/news-and-events/blogs/2025/03/ai-load-impact-on-data-centers-adapting-to-the-future-of-infrastructure
2. National Renewable Energy Laboratory (NREL), Server efficiency as transactions per second per watt.
https://www.nrel.gov/docs/fy25osti/12345.pdf