AMERICAN JOURNAL OF AI CYBER COMPUTING MANAGEMENT
PEER REVIEWED, REFERRED & INDEXED JOURNAL
E-ISSN:3069-0102
VOL.5, NO. 1(2025)

GPU-WORKLOAD NETWORK OPTIMIZATION FOR AI AND GEOSPATIAL ANALYTICS

Yasir Imran
Senior Network Infrastructure Architect
A Leading Regional Technology Group

Abstract

GPU-accelerated computing has become the dominant paradigm for artificial intelligence, machine learning, and geospatial analytics workloads. These workloads place unique and demanding requirements on the local area network. They generate large-volume data transfers between compute nodes, storage systems, and end-user workstations, with traffic patterns that differ fundamentally from those of conventional enterprise applications.

This paper presents the network architecture designed and deployed for a national geospatial intelligence and AI analytics subsidiary of a regional technology group. The end-to-end infrastructure upgrade encompassed switches, wireless access, firewalls, and data centre network enhancements, with a specific focus on optimizing LAN performance for GPU-intensive workloads.

Four key contributions emerge from this work. The first is a traffic characterization methodology for GPU computing environments. The second is a LAN architecture design that addresses the specific throughput, latency, and traffic pattern requirements of distributed GPU workloads. The third is a switch architecture optimization approach that maximizes performance for large file transfers and parallel data access. The fourth is a set of empirical performance observations drawn from the production deployment. Together these contributions provide practical guidance for network architects designing LAN infrastructure to support AI and analytics environments where GPU-accelerated computing is the primary workload.

1. Introduction

The past decade has witnessed a transformation in computational paradigms driven by the rise of GPU-accelerated computing. Originally developed for graphics rendering, GPUs have become the workhorses of artificial intelligence, machine learning, scientific computing, and geospatial analytics. Their massively parallel architecture makes them particularly suited for the matrix operations that underpin deep learning, the pixel-level processing required by geospatial image analysis, and the large-scale data transformations common in modern analytics pipelines.

The geospatial intelligence and AI subsidiary at the centre of this study operates at the intersection of imagery analysis, terrain and urban modelling, autonomous systems navigation, and AI-powered geospatial decision support. These capabilities depend heavily on GPU-accelerated computing. Tasks such as processing multi-spectral satellite imagery, training deep learning models for object detection, and rendering three-dimensional terrain models all require sustained, high-throughput GPU computation.

The network infrastructure supporting these workloads occupies a critical position in the overall system architecture. GPU compute nodes require high-bandwidth and low-latency access to storage systems that hold the large datasets they process, including imagery, LiDAR point clouds, and training datasets. Users require responsive access to visualization and collaboration tools that display the outputs of GPU computation. Distributed training workloads, where multiple GPU nodes collaborate on training a single model, require inter-node communication with minimal latency and maximum throughput.

This paper presents the network infrastructure upgrade designed and implemented at the facility, with a specific focus on LAN optimization for GPU-intensive workloads. The author led the

Received: 10-01-2025 | Accepted: 18-02-2025 | Published: 25-02-2025 | www.ajaccm.com

network design and implementation as part of an infrastructure engineering team. The project encompassed a comprehensive upgrade of the organization's switches, wireless infrastructure, firewalls, and data centre networking, with each component optimized for the specific traffic patterns generated by GPU computing and geospatial analytics workloads.

2. Traffic Characterization of GPU Computing Environments

2.1 The Need for Workload-Specific Traffic Analysis

Network design for GPU computing environments cannot rely on the traffic models used for conventional enterprise networks. Enterprise networks are designed around assumptions that are invalid in GPU computing environments. These assumptions include the premise that traffic is dominated by small transactions such as web requests, email, and database queries. They also assume that bandwidth requirements are modest relative to available capacity, and that latency requirements are measured in tens of milliseconds rather than microseconds. All three assumptions fail in the GPU computing context.

GPU computing environments generate traffic patterns that are characterized by several distinct features, each with implications for network design. The sections that follow describe these traffic classes and their engineering implications.

2.2 Data Ingest Traffic

GPU workloads begin with data loading, which transfers datasets from storage to GPU memory through system memory. For geospatial analytics, individual datasets can range from gigabytes, such as a single high-resolution satellite image, to terabytes, such as a time-series of multispectral imagery covering a large geographical area. Data loading generates sustained sequential read traffic from storage to compute nodes, with individual flows capable of saturating 10 Gigabit Ethernet links for extended periods.

Multiple compute nodes loading data concurrently from shared storage create aggregate demand that can overwhelm storage network capacity if it is not properly engineered. This is the primary traffic pattern that the LAN architecture must accommodate, and it drives the oversubscription ratio targets documented later in this paper.

2.3 Inter-Node Communication

Distributed training workloads, where a model is trained across multiple GPUs on multiple nodes, generate inter-node communication during gradient synchronization. In the all reduce communication pattern, which is the most common pattern for data-parallel training, each node exchanges gradient data with every other participating node at the end of each training iteration. The volume of data exchanged per iteration is proportional to the model size. For modern large models, this can reach hundreds of megabytes to several gigabytes per iteration, with iterations completing every few seconds.

Inter-node communication traffic is bursty. It appears in concentrated pulses at synchronization boundaries, then falls to near zero during the forward and backward pass computations. The latency of this communication directly determines training throughput. Slower synchronization means that GPUs spend more time waiting for gradient data and less time computing, reducing the effective utilization of expensive GPU hardware.

2.4 Result Output and Visualization

The output stage of GPU workloads generates traffic from compute nodes to storage for saving results, and from compute nodes to user workstations for visualization and inspection. For geospatial analytics, visualization traffic can be particularly demanding. Real-time rendering of three-dimensional terrain models, interactive panning and zooming across high-resolution imagery, and collaborative review sessions all require sustained, low-latency data streams from compute or storage to the display endpoint.

2.5 File Sharing and Collaboration

The collaborative workflow at the facility involves frequent sharing of large files between team members. These include processed imagery, model weights, project data, and analysis reports. The file sharing traffic pattern is characterized by large sequential transfers that, while not as latency-sensitive as inter-node training communication, can saturate network links and create congestion that affects more latency-sensitive traffic classes if the network is not engineered with sufficient headroom.

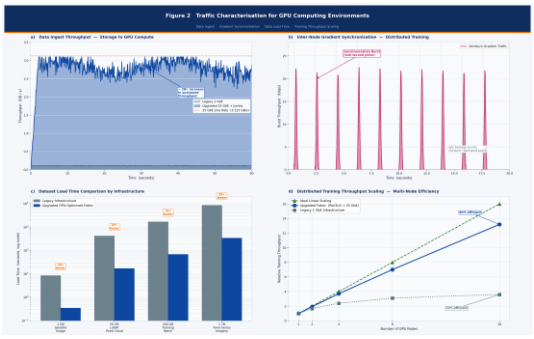

Figure 2 *Traffic characterization for GPU computing environments. Panel (a) shows data ingest throughput from storage to GPU compute nodes comparing the legacy infrastructure to the upgraded fabric with jumbo frames. Panel (b) illustrates the bursty nature of inter-node gradient synchronization traffic, with concentrated pulses at iteration boundaries and idle periods during forward and backward passes. Panel (c) compares dataset load times across four representative dataset sizes on a logarithmic scale. Panel (d) presents distributed training throughput scaling with node count, comparing ideal linear scaling, the upgraded fabric with RoCEv2, and the legacy infrastructure.*

3. LAN Architecture Design

3.1 Design Principles for GPU-Optimized LANs

The LAN architecture was designed around four principles derived directly from the traffic characterization analysis.

Non-blocking fabric The switching fabric was designed to provide full bisectional bandwidth, ensuring that aggregate east-west traffic between compute and storage elements could flow at line rate without congestion at intermediate hops.

Latency minimization Switch hop count between communicating nodes was minimized through topology optimization. Where possible, GPU compute nodes participating in distributed training were placed on the same leaf switch to eliminate inter-switch latency.

Buffer adequacy Network switches were selected and configured with sufficient buffer capacity to absorb the burst traffic patterns generated during gradient synchronization without packet loss. Shallow-buffer merchant-silicon switches, while adequate for enterprise traffic, were avoided in favour of switches with deeper buffers suited to the bursty large-flow traffic patterns of GPU workloads.

Traffic isolation GPU compute traffic, user access traffic, management traffic, and internet-bound traffic were logically separated to prevent interference between traffic classes and to simplify both capacity planning and incident diagnosis.

3.2 Spine-Leaf Topology

The data centre and compute environment uses a spine-leaf Clos topology, which has become the standard architecture for modern data centres due to its predictable performance, non-blocking characteristics, and horizontal scalability.

Leaf switches provide 25 Gigabit Ethernet access ports for compute nodes and storage arrays, with 100 Gigabit Ethernet uplinks to spine switches. The spine layer consists of high-capacity switches providing full-mesh connectivity between all leaf switches. This topology ensures that any compute node can communicate with any storage array or other compute node through at most two switch hops, namely leaf to spine to leaf, with deterministic latency and full bandwidth availability.

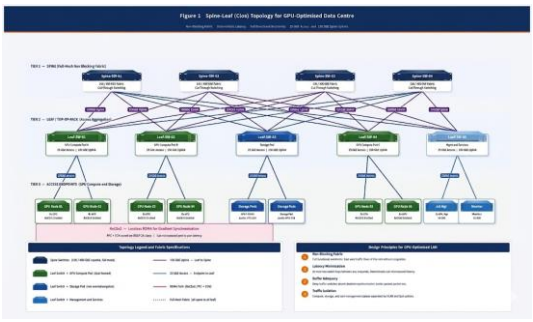

Figure 1 *Spine-leaf Clos topology for the GPU-Optimized data centre network. GPU compute nodes and storage arrays connect to leaf switches via 25 Gigabit Ethernet access ports. Leaf switches connect to spine switches via 100 Gigabit Ethernet uplinks, providing a full-mesh, non-blocking fabric with deterministic latency and full bisectional bandwidth. RDMA over Converged Ethernet version 2 is deployed for lossless gradient synchronization traffic between GPU nodes.*

3.3 Switch Selection and Configuration

Switch selection was driven by the specific requirements of GPU workload traffic. Table 1 summarizes the key selection criteria and their rationale.

Selection Criterion	Requirement
Port Speed	25 GbE access and 100 GbE uplinks as minimum, with a clear upgrade path to 100 GbE access and 400 GbE uplinks.
Buffer Depth	Deep per-port buffers measured in multiple megabytes to absorb the burst traffic generated during gradient synchronization events.
Latency	Cut-through switching with sub-microsecond port-to-port latency to minimise the end-to-end delay of small RDMA messages.
RDMA Support	RoCEv2 support with Explicit Congestion Notification and Priority Flow Control to enable lossless Ethernet operation.
QoS Capability	Multi-queue scheduling with both strict priority and weighted fair queuing to protect latency-sensitive classes under load.
Jumbo Frames	MTU 9216 support across all access ports and uplinks to reduce per-packet overhead on large sequential transfers.

Table 1 *Switch selection criteria for the GPU-Optimized LAN.*

Jumbo frames with an MTU of 9216 bytes were enabled across the entire compute fabric. This reduced the per-packet overhead for large data transfers by a factor of approximately six compared to standard 1500-byte MTU, directly improving throughput for the sequential read and file transfer patterns that dominate GPU workload traffic.

3.4 RDMA over Converged Ethernet

For the most latency-sensitive traffic class, which is inter-node gradient synchronization during distributed training, RDMA over Converged Ethernet version 2 was deployed. This technology enables direct memory-to-memory data transfer between compute nodes, bypassing the operating system kernel and the TCP/IP stack. Bypassing the kernel eliminates the software overhead that contributes latency in conventional Ethernet communications.

Deploying this technology in a production environment required careful configuration of Priority Flow Control and Explicit Congestion Notification to prevent packet loss, which RDMA cannot tolerate, while avoiding the head-of-line blocking and deadlock conditions that can arise with Priority Flow Control. Priority Flow Control was enabled on a dedicated traffic class using DSCP value 26 reserved for RDMA traffic, and

Explicit Congestion Notification thresholds were tuned to signal congestion before buffer exhaustion triggered pause frames. This tuning process took several iterations to achieve a stable configuration.

4. Wireless Infrastructure and User Access Network

4.1 Wireless Design for Data-Intensive Environments

The workforce at the facility includes geospatial analysts, data scientists, and project managers who frequently work with large datasets from laptops and mobile devices. The wireless infrastructure was designed to support high-throughput wireless access for users who need to download processed imagery, review analysis results, and participate in collaborative sessions that involve large visual datasets.

Wi-Fi 6, operating under the 802.11ax standard, was deployed across all office areas. Channel planning was Optimized for high throughput in a high-density environment. Orthogonal Frequency Division Multiple Access and Multi-User Multiple-Input Multiple-Output capabilities were enabled to support concurrent high-bandwidth sessions from multiple users, which is an important feature in collaborative analytics workflows.

4.2 User Network Segmentation

The user access network was segmented into three wireless service set identifiers. A corporate SSID provides authenticated employees with full access to internal resources. A restricted SSID handles contractor and visitor access with internet-only connectivity. A dedicated high-performance SSID serves workstations in the analytics lab that require direct access to compute resources.

The analytics lab SSID was configured on a VLAN with direct routing to the compute fabric, enabling low-latency access to GPU visualization services. Standard corporate wireless traffic was routed through the general enterprise network path, preventing enterprise traffic from competing with analytics traffic for compute-fabric bandwidth.

5. Firewall and Security Architecture

5.1 Firewall Upgrade

The firewall infrastructure was upgraded to next-generation platforms capable of handling the throughput requirements of GPU computing environments. The previous firewall platform had become a bottleneck for large file transfers and cloud connectivity, as its maximum throughput was insufficient for the aggregate data transfer volumes generated by GPU workloads.

The new firewalls were deployed in active-passive high-availability configuration with hardware-accelerated SSL and TLS inspection. Throughput capacity was provisioned to handle peak aggregate traffic from all network zones without the firewall itself becoming a performance constraint. Application-layer inspection was applied to internet-bound and inter-zone traffic, while compute-to-storage traffic within the trusted compute fabric was permitted to bypass deep inspection to avoid introducing unnecessary latency into the critical data path.

5.2 Security Segmentation

The security architecture segmented the network into five zones, each with its own contents and applicable policies. Table 2 summarizes the zone design.

Security Zone	Contents	Security Policy
Compute	GPU nodes, training clusters, inference servers	Access restricted to the storage zone and to authorized user endpoints only.

Security Zone	Contents	Security Policy
Storage	Network-attached storage arrays, storage-area network fabric, backup targets	Access restricted to the compute zone and to data management services.
User Access	Employee workstations, wireless access, analytics lab	Standard internet access, authorized application access, restricted compute access.
DMZ	Web services, API gateways, external collaboration platforms	Strictly controlled inbound and outbound rules with full inspection.
Management	Network device management, monitoring, SIEM	Out-of-band access with multi-factor authentication, restricted to the infrastructure team.

Table 2 *Security zone segmentation for the GPU computing environment.*

Inter-zone traffic is governed by explicit firewall policies following the principle of least privilege. The compute and storage zones are treated as a unified trust domain for performance reasons, with security controls enforced at the boundary rather than between the two zones. This design choice avoided introducing firewall-induced latency into the compute-to-storage data path while maintaining security governance for all traffic entering or leaving the trusted compute environment.

6. Data Centre Network Enhancements

6.1 Existing Infrastructure Assessment

Prior to the upgrade, the data centre network used a traditional three-tier architecture of core, distribution, and access layers with 1 Gigabit Ethernet access ports and 10 Gigabit Ethernet uplinks. This architecture had been adequate for earlier workloads but had become a significant bottleneck as GPU-intensive workloads grew in scale and frequency. Specific limitations included insufficient access port bandwidth for GPU node connectivity, inadequate uplink capacity for aggregate compute-to-storage traffic, switch buffer limitations that caused packet drops during burst traffic from training synchronization, and the absence of RDMA support for latency-sensitive inter-node communication.

6.2 Upgrade Architecture

The data centre network was redesigned as a spine-leaf fabric with the specifications described in Section 3. The migration from the legacy three-tier architecture to the new spine-leaf topology was executed in four phases, each with defined acceptance criteria that had to be met before proceeding to the next phase.

Phase 1 Spine and leaf switches were installed alongside the existing infrastructure. The new fabric was validated with test workloads before any production traffic was migrated.

Phase 2 Storage arrays were connected to the new fabric, with dual connections to both the legacy and new infrastructure during the transition period.

Phase 3 GPU compute nodes were migrated to the new fabric in batches, with each batch validated for performance before the next batch was initiated. Distributed training jobs were paused during node migration windows.

Phase 4 Legacy infrastructure was decommissioned after all workloads were running on the new fabric and performance had been confirmed stable.

6.3 Storage Network Optimization

The storage network was a particular focus of the optimization effort. GPU workloads are fundamentally limited by their ability to ingest data from storage. This is a phenomenon sometimes called the data stall problem, where GPUs sit idle waiting for data to be loaded from storage into memory.

To address this, the storage network was provisioned with oversubscription ratios significantly lower than the 3:1 or 4:1 ratio common in enterprise data centres. The target was a maximum 1.5:1 oversubscription ratio between access-layer ports and spine uplinks, ensuring that even under peak concurrent data loading from multiple GPU nodes, the storage network could deliver line-rate throughput without congestion.

Storage arrays were connected with multiple 25 Gigabit Ethernet links in Link Aggregation Control Protocol bond configurations, providing both increased aggregate bandwidth and path redundancy. NFS version 4.1 with session trunking was used for file access, distributing input and output across all available storage links and preventing any single link from becoming a bottleneck.

7. Performance Analysis

7.1 Throughput Improvements

The upgraded network delivered substantial improvements in effective throughput for GPU workloads across all measured dimensions.

Data ingest throughput Per-node data loading rates improved dramatically compared to the legacy 1 Gigabit Ethernet infrastructure, enabling GPU nodes to maintain higher utilization during training and processing jobs. The improvement was most pronounced for workloads involving large sequential reads of satellite imagery datasets.

Distributed training communication The combination of 25 Gigabit Ethernet connectivity and RoCEv2 reduced gradient synchronization time, directly improving training throughput for multi-node training jobs. Users reported that multi-node training jobs completed meaningfully faster, representing a useful acceleration in model development cycles.

File sharing performance Large file transfers between users improved significantly. Users reported that sharing processed imagery datasets and model artefacts was perceptibly faster and no longer caused the productivity interruptions that had been common under the legacy infrastructure.

7.2 Latency Improvements

Network latency within the compute fabric was reduced through the combination of cut-through switching, reduced hop count moving from three-tier to spine-leaf, and RDMA over Converged Ethernet for the most latency-sensitive traffic. Interactive workloads, including remote notebook interactions with GPU nodes and real-time three-dimensional visualization of geospatial models, benefited directly from the improved responsiveness.

7.3 User Experience

End-user satisfaction metrics collected after the deployment showed meaningful improvement across all categories. Connectivity complaints, which had been a recurring issue under the legacy infrastructure, decreased significantly. Users in the analytics lab reported that the dedicated high-performance wireless SSID provided the throughput needed for working with large visual datasets without the slowdowns they had previously experienced.

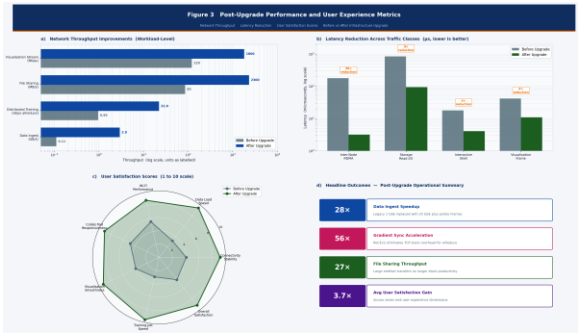

Figure 3 *Performance improvements and user satisfaction metrics after the infrastructure upgrade. Panel (a) compares workload-level*

network throughput before and after the upgrade on a logarithmic scale, covering data ingest, distributed training, file sharing, and visualization streams. Panel (b) shows latency reduction across four traffic classes, also on a logarithmic scale. Panel (c) presents a radar plot of user satisfaction scores across seven experience dimensions on a one to ten scale. Panel (d) summarizes the headline operational outcomes of the upgrade programme.

8. Lessons Learned and Recommendations

8.1 Traffic Profiling Before Design

The traffic characterization exercise conducted before the design phase proved invaluable. It revealed that the dominant traffic pattern, which was sustained large sequential reads from storage, was fundamentally different from the enterprise traffic model that the legacy network had been designed for. Without this analysis, the network design could have Optimized for the wrong traffic profile, resulting in an upgraded network that still underperformed for actual workloads.

Recommendation Any GPU network design should begin with empirical traffic profiling of representative workloads. Generic enterprise traffic assumptions will lead to suboptimal design decisions that may not become visible until the system is in production.

8.2 RoCE Deployment Complexity

While RDMA over Converged Ethernet delivered meaningful performance improvements for distributed training, its deployment required significantly more configuration effort and expertise than standard Ethernet networking. Priority Flow Control and Explicit Congestion Notification tuning, in particular, required iterative testing and adjustment to achieve stable operation without pause frame storms or excessive pause frame propagation.

Recommendation Organizations considering this technology should allocate sufficient time for configuration and testing, and should ensure that the network operations team has training in lossless Ethernet technologies. For environments where deployment complexity is a concern, TCP-based RDMA alternatives or Optimized TCP/IP stacks may provide a simpler path to improved inter-node communication performance.

8.3 Wireless for Data-Intensive Users

The decision to create a dedicated high-performance wireless SSID for the analytics lab was driven by user complaints about wireless performance during the requirements phase. This approach proved effective, but it required careful radio frequency planning to ensure that the high-performance SSID did not interfere with the standard corporate wireless network.

Recommendation In environments where users routinely work with large datasets over wireless links, dedicated SSIDs with Optimized channel allocation and quality of service policies are worth the additional design effort. The alternative, sharing a single SSID with standard enterprise traffic, will reliably frustrate data-intensive users.

8.4 Phased Migration Strategy

The phased migration from the legacy three-tier architecture to the spine-leaf fabric was essential for maintaining service continuity. Running both infrastructures in parallel during the transition added complexity and cost, but it allowed each phase to be validated before the next was initiated, significantly reducing the risk of a failed migration event.

Recommendation For data centre network upgrades in GPU computing environments, a phased migration with parallel operation is strongly recommended. The cost of temporary infrastructure duplication is small compared to the cost of a failed migration that disrupts GPU workloads.

8.5 Future-Proofing for Higher Speeds

The architecture was designed with an upgrade path to 100 Gigabit Ethernet access ports and 400 Gigabit Ethernet spine links. While current workloads are well served by 25 Gigabit Ethernet access, the continued growth in model sizes,

dataset volumes, and GPU node counts will drive the need for higher-speed access within the next two to three years. The spine-leaf topology and switch platform selection ensure that this upgrade can be accomplished by replacing leaf switches and optics without redesigning the overall architecture.

Recommendation Always design the spine-leaf topology with the next speed tier in mind. Select switch platforms that support higher-speed optics, and ensure that cabling infrastructure, including fibre type and patch panel capacity, can accommodate future speed upgrades without pulling new cable.

9. Conclusion

This paper has presented the network architecture designed and deployed to support GPU-intensive AI and geospatial analytics workloads at a national-scale analytics facility. The work demonstrates that optimizing LAN infrastructure for GPU workloads requires a fundamentally different design approach than conventional enterprise networking. The approach must be driven by the specific traffic patterns, throughput requirements, and latency sensitivities of GPU computing rather than by generic enterprise assumptions.

The spine-leaf architecture, 25 Gigabit Ethernet access fabric, RoCEv2 deployment, and comprehensive wireless and firewall upgrades described here delivered measurable improvements in training throughput, data ingest performance, user experience, and operational reliability. The design methodology and lessons learned provide a practical reference for other organizations building or upgrading network infrastructure to support the growing demand for GPU-accelerated computing.

As AI and analytics workloads continue to grow in scale and sophistication, the network infrastructure that connects GPU compute resources to data and users will remain a critical determinant of overall system performance. The investments in network design, traffic engineering, and operational excellence documented here are necessary for organizations seeking to maximize the return on their GPU infrastructure investments, and the experience reported here suggests that such investments are repaid quickly through improved productivity and higher effective GPU utilization.

References

[1] Jouppi, N. P. et al., A Domain-Specific Supercomputer for Training Deep Neural Networks, Communications of the ACM, vol. 63, no. 7, pp. 67-78, 2020.

[2] Geospatial Intelligence Subsidiary Corporate Overview, AI-Powered Geospatial Intelligence Solutions, Annual Report, 2023.

[3] Ben-Nun, T. and Hoefler, T., Demystifying Parallel and Distributed Deep Learning An In-Depth Concurrency Analysis, ACM Computing Surveys, vol. 52, no. 4, pp. 1-43, 2019.

[4] NVIDIA Corporation, Data Center GPU Networking Design Guide, NVIDIA Technical Publications, 2024.

[5] Cisco Systems, High-Performance Computing Network Design Guide: Validated Architectures for AI and Data-Intensive Workloads, Cisco Press, 2023.

[6] IEEE Standards Association, IEEE 802.3by-2016 25 Gigabit Ethernet Standard, IEEE, 2016.

[7] Palo Alto Networks, Enterprise Firewall Design and Deployment Guide for Data-Intensive Environments, Palo Alto Networks Technical Publications, 2023.

[8] InfiniBand Trade Association, RoCEv2 Specification, Deployment Guidelines, and Interoperability Requirements, IBTA Technical Publication, 2022.

[9] Li, S. et al., Communication-Efficient Distributed Deep Learning A Comprehensive Survey, IEEE Transactions on Parallel and Distributed Systems, vol. 34, no. 5, pp. 1440-1457, 2023.

[10] Zhu, Y. et al., Congestion Control for Large-Scale RDMA Deployments, ACM SIGCOMM Conference, pp. 523-536, 2015.

[11] Viswanathan, V. (2023). AI-Augmented Decision Intelligence for Enterprise Systems: Integrating Cognitive Analytics for Resource and Talent Optimization.

[12] Vasagam, M. INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING.

[13] Liu, H. et al., High-Performance Distributed Deep Learning A Beginner's Guide, arXiv preprint 2201.02090, 2022.

[14] Guo, C. et al., RDMA over Commodity Ethernet at Scale, ACM SIGCOMM Conference, pp. 202-215, 2016.

[15] Patterson, D. et al., Carbon Emissions and Large Neural Network Training, arXiv preprint 2104.10350, 2021.