Advanced Networks for Artificial Intelligence and Machine Learning Computing

Scaling Fiber Networks to Meet Tomorrow's Data Center Demands

Executive Summary

This document explores the critical considerations linked to data centers optimized for AI workloads. By highlighting the growing computational power required by large language models (LLMs), the paper seeks to inform readers on the necessity for advanced networking and innovative physical layer solutions.

What you will learn:

01

Energy consumption

Due to the energy required by high-performance hardware and the complexity and size of the datasets needed for LLM training and inference, AI data centers present significantly higher power demands compared to traditional hyperscale architectures.

02

Cooling solutions

AI workloads generate significant heat, necessitating advanced thermal management methods such as direct-to-chip and immersion cooling – traditional air-cooling methods cannot meet the cooling requirements of high-density AI data center environments.

03

Physical space requirements

To accommodate very large systems with specialized hardware and cooling systems, AI data center size – both in terms of physical footprint and cubic meters – has grown and continues to grow, with future projects indicating even greater space requirements.

04

Network topologies

Choice of network topology defines a system's data flow efficiency and readiness for rapid scalability. With the aim of minimizing latency and maximizing bandwidth, operators must select between purpose-designed topologies for optimized results.

05

Backend network (BENW) and frontend network (FENW)

From sharing model updates during training to low-latency connections between accelerators, discover the essential load balancing and network control mechanisms behind the dynamic demands of AI workloads.

06

Scalability

Looking to the future, we consider a scalable Clos network supporting hundreds of thousands of endpoints. How can operators ensure robust performance and fault tolerance?

This overview of AI data center infrastructure, hardware requirements, and capabilities provides the groundwork for a forthcoming comprehensive exploration of in-depth technical considerations.

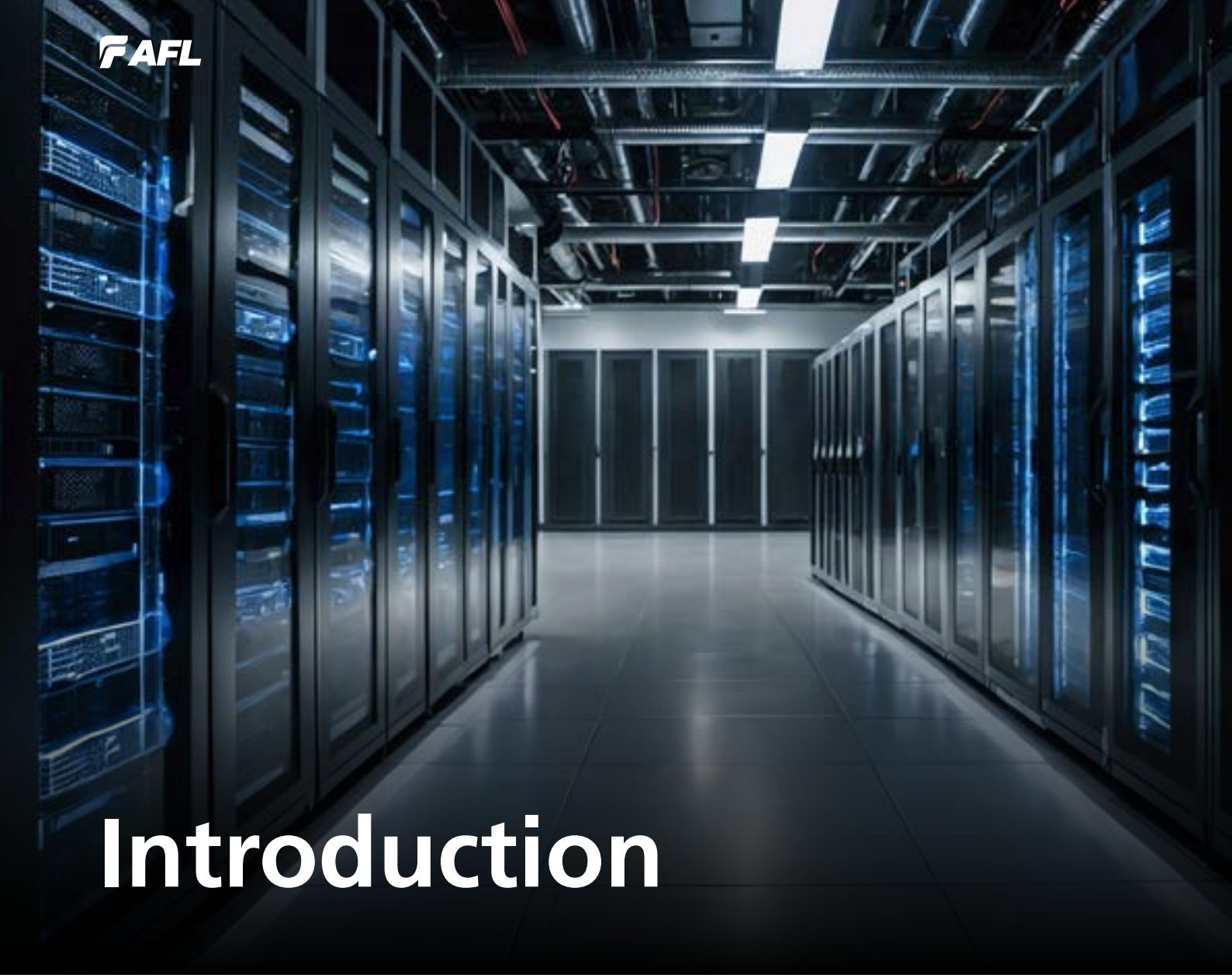
Written by

Alan Keizer
Senior Technology Advisor, AFL

Ben Atherton
Technical Author, AFL

“The emergence of generative AI, with its exceptionally large models and truly extraordinary computing requirements, has driven greater and more rapid change in data center networking in the past two years than I have seen in the previous decade.”

Alan Keizer
Senior Technology Advisor, AFL



Introduction

The surging demand for artificial intelligence (AI) and machine learning (ML) technologies presents data center operators with unique challenges in terms of increasing, optimizing, and maintaining network efficiency. To keep pace, modern data center architectures must explore new and seamless ways to adopt advanced AI technologies and hardware.

The unprecedented computational power and energy resources linked to the rise of large language models (LLMs) cannot be overlooked, requiring a deeper understanding of the energy dynamics, cooling solutions, and network topologies essential for AI data center efficiency.

By closely examining multiple performance-related factors, industry leaders can better equip the data center operators of tomorrow with the necessary tools and wisdom to support the next generation of AI innovations.

This white paper explores the intricacies of AI data center networking, highlighting the significant differences between traditional infrastructures and data centers optimized for AI workloads.

What's Different About an AI Data Center Network?

Large Language Models (LLMs) are systems trained on data to recognize patterns, discern sentiment, and generate human-like language in response to prompts. LLM creation follows a two-step process. First, the training phase involves AI models learning from datasets by adjusting parameters to improve accuracy. Next, during the inference phase, trained models apply the knowledge learned from training to make predictions - or decisions – in response to new data.

LLMs provide the natural language processing capability within the broader AI ecosystem. To train the requisite LLMs for AI data center networks requires immense power and computational resources. For example, today's leading-edge GPUs used to train clusters comprising over 100,000 GPUs can each consume 1,200 to 1,500 watts. This results in total data center power in the range of 300 megawatts moving towards a gigawatt. Let's breakdown what's different about an AI data center network.

Energy consumption

Energy is power over time, expressed as kilowatt-hours (kWh). Energy consumption is a critical differentiator that sets AI data centers apart from traditional data centers. The combination of high-performance hardware and the computational demands of training and inference drives the need for massive amounts of power. Large Language Models (LLMs), which can have billions or even trillions of parameters, require immense energy resources. As models scale, the energy required for both training and inference also increases. This relationship between power and scale underpins the energy dynamics in AI data centers.

Training phase

Factors influencing energy consumption during the training phase include hardware efficiency, dataset size, and model complexity. The training phase can be divided into two main components:

Data processing

This involves cleaning and preparing data before training. For example, the Common Crawl dataset, used to train models like GPT-3, comprises 9.5 petabytes of data.

Iterative computation

The power and hardware required during this phase varies based on dataset size and model complexity. As model parameters grow, the demand for computational resources increases, leading to greater energy consumption over time – advanced accelerators require 3-to-10 times the power but result in hundreds to thousands of times the energy consumption.

Inference phase

Once trained, large model energy consumption levels remain high, particularly in scenarios demanding real-time calculations – generally, inference is less computationally intensive than training, but still consumes substantial amounts of energy, especially in relation to high-frequency requests. This highlights the ongoing energy demands – and efficient hardware considerations – of maintaining responsive AI services.

Power and cooling

AI workloads use 300–1,000 times more power than traditional CPU-powered data center operations. The additional power generates heat, requiring sophisticated thermal management. Leading-edge AI/ML systems use some form of liquid cooling (i.e., immersion or direct-to-chip).

For high-performance components, direct to chip is the most common cooling method – the latest NVIDIA GB200 accelerator is only available with direct-to-chip cooling. While immersion methods may offer the highest potential per rack, unresolved environmental issues and equipment compatibility considerations may hinder wider adoption.

Air cooling

Traditional air-cooling methods use air to dissipate heat and cool data center components. Solutions include heat sinks (passive devices that increase surface areas to enable greater heat transfer to the surrounding air), heat pipes (used to transfer heat to cooler areas) and active cooling fans (used to force air toward components). Heat dissipation via air cooling is insufficient for AI workloads. Heat removal up to 20 kW per rack can be achieved.

Liquid cooling

Liquid cooling is an increasingly popular heat dissipation solution in AI data centers – by 2026, 38.6%[1] of IT professionals surveyed expect to see liquid cooling techniques deployed in data centers. Examples include:

Direct-to-chip

Primarily implemented in large systems. Capable of managing 100kW+ of heat load per rack, adequate for today's most demanding AI workload environments.

Rear-door heat exchangers

Designed for high-density data centers with server rack densities exceeding 20 kW per rack (up to 50 kW per rack). Mounted at the back of racks, the units utilize chilled water to cool servers. In legacy data centers, deploying a limited number of AI/ML racks, operators may retrofit rear-door heat exchangers to enhance cooling efficiency without the need to overhaul the existing infrastructure.

Immersion cooling

Immersion techniques place servers into liquid-filled tanks (typically using a biodegradable, nontoxic, synthetic liquid). With efficient heat extraction and lower thermal system power consumption, immersion cooling methods represent the most efficient technique for AI data centers. However, there are potential environmental risks and potential compatibility issues between fluids and server components. Immersion technology is not yet widely adopted. Heat loads more than 250 kW can be accommodated.

AI data centers can reduce carbon emissions using AI-driven cooling strategies. Through data monitoring and predictive optimization, AI-powered cooling systems can adapt to energy-efficient configurations in response to environmental changes such as weather. For example, a standout Google statistic from 2016 showed how DeepMind (a subsidiary of Google, concentrated on developing AI technologies) was able to reduce the energy needed for data center cooling by 40%. The drive to reduce power consumption and heat generation is intensive throughout the AI/ML ecosystem.

Physical space requirements

Compared to traditional hyperscale data centers, AI data centers must have more computing in tightly integrated clusters. Large AI/ML clusters require significantly more physical space. The additional square footage not only houses the specialized hardware and cooling infrastructure necessary for optimal performance, but also provides the data center white space needed for rapid scalability.

The trend for larger AI data center footprints is expected to continue, with upcoming projects such as Microsoft and OpenAI's Stargate facility indicating power requirements of several gigawatts – potentially requiring nuclear power – with a total footprint spanning hundreds of acres.

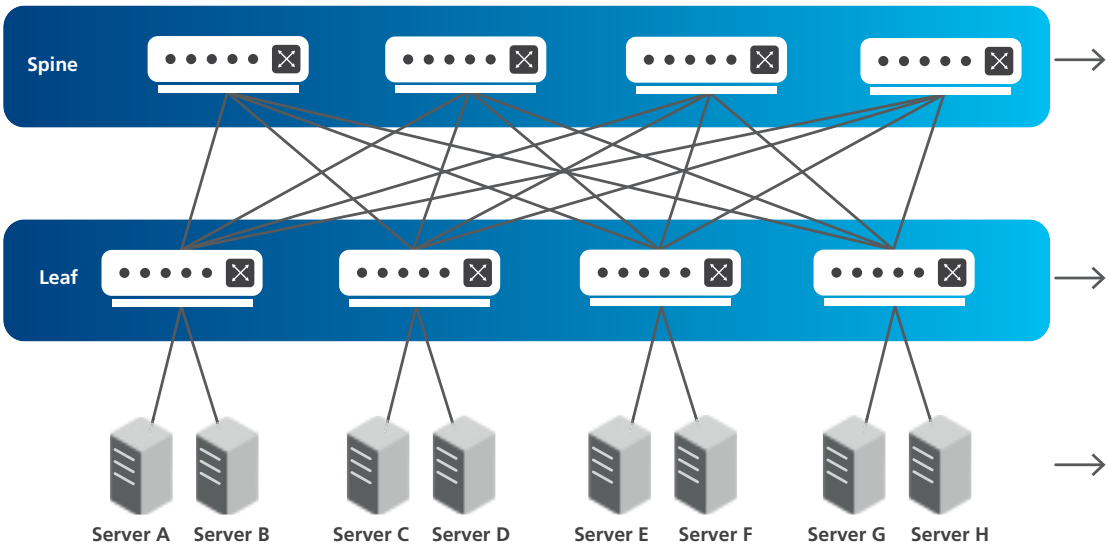
1. cio

Network Topologies

The choice of AI data center network topology is crucial not only to the efficient flow of data but to the facility's overall preparedness for rapid growth and scalability. Most switches typically have port counts of 32, 36, 64, or 72, with each port supporting multiple channels of serial communications, allowing for extensive connectivity options. There are several main types of modern, advanced network topology, including Clos, torus, dragonfly, and hybrid infrastructures. The examples below show some of the main network topologies used by AI data centers.

Clos topology (Leaf-spine)

Named after Charles Clos, who formalized the architecture in 1952, Clos topologies serve AI data centers by providing non-blocking, high-bandwidth connectivity, which minimizes congestion and supports efficient data transfer between nodes to reduce task completion times.



Structure

Clos topologies are multistage packet switching networks in a multi-layer edge, leaf, and spine configuration. Each stage uses multiple switches, connecting multiple inputs and outputs in a matrix configuration for simultaneous input/output connections.

Usage

In data centers, the spine-and-leaf architecture connects each leaf switch to multiple spine switches. This architecture ensures minimal 'hops' between endpoints, creating a high-performance, high bandwidth, low-latency environment ideal for data centers.

Advantages

The multistage setup enables multiple pathways between endpoints for accelerated scalability and greater fault tolerance – for example, if one leaf or spine were to fail, bandwidth would degrade but network communication would still be possible.

Torus topology

Torus topologies efficiently distribute computing tasks to provide low-latency communication between nodes, enhancing performance and scalability. Note that torus is a highly specialized topology, requiring non-standard hardware and software. As such, only major AI/ML operators deploy torus topologies (e.g., Google[2]).

Structure

Each node is arranged to create a mesh-like structure, enabling connections between neighboring nodes. In higher dimensions, nodes may also connect to neighboring nodes in additional dimensions.

Usage

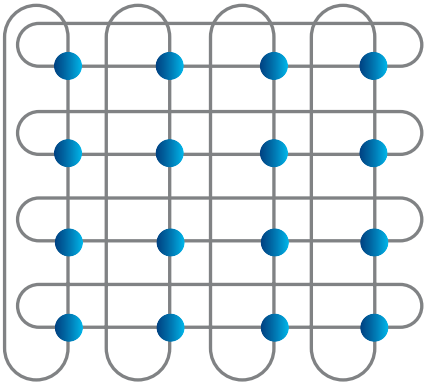
Parallel computing systems: A torus topology suits a parallel computing system, as this style of architecture supports frequent, local communication between nodes.

Advantages

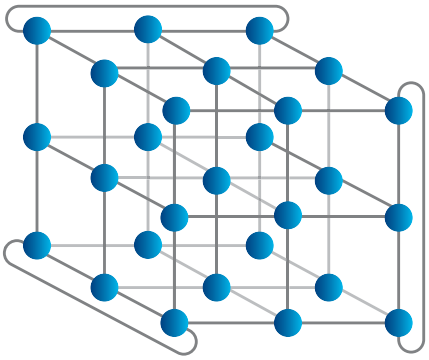
Torus topologies are relatively simplistic in design and provide the low latency required for efficient local communication.

1-D Torus (4-ary 1-cube)

2-D Torus (4-ary 2-cube)



3-D Torus (3-ary 3-cube)



Optical circuit switching

Connection patterns during machine learning (ML) training tend to be stable. The stability enables slower spine tier switching, allowing for the potential replacement of fast packet switching with slower circuit switching technologies, such as Google's Micro-Electro-Mechanical Systems (MEMS)[3].

Advantages include being wavelength, bandwidth, and protocol agnostic, facilitating mixing, matching, and evolving technologies, along with significantly lower power consumption. However, drawbacks include the current lack of commercial availability and possible concerns regarding cost and reliability.

2. Google
3. Springer

What about Hybrid Topologies?

Hybrid data center network topologies combine different elements of different structures for optimal performance. For example, a tailored topology may consist of the following elements:

Clos network for the core layer

Using a multistage switching Clos network for the core layer provides high bandwidth, low-latency, and high-speed data transfer.

Dragonfly topology for the aggregation layer

Dragonfly topologies group high-radix routers into virtual servers, minimizing global channel and reducing latency.

Torus topology for compute clusters

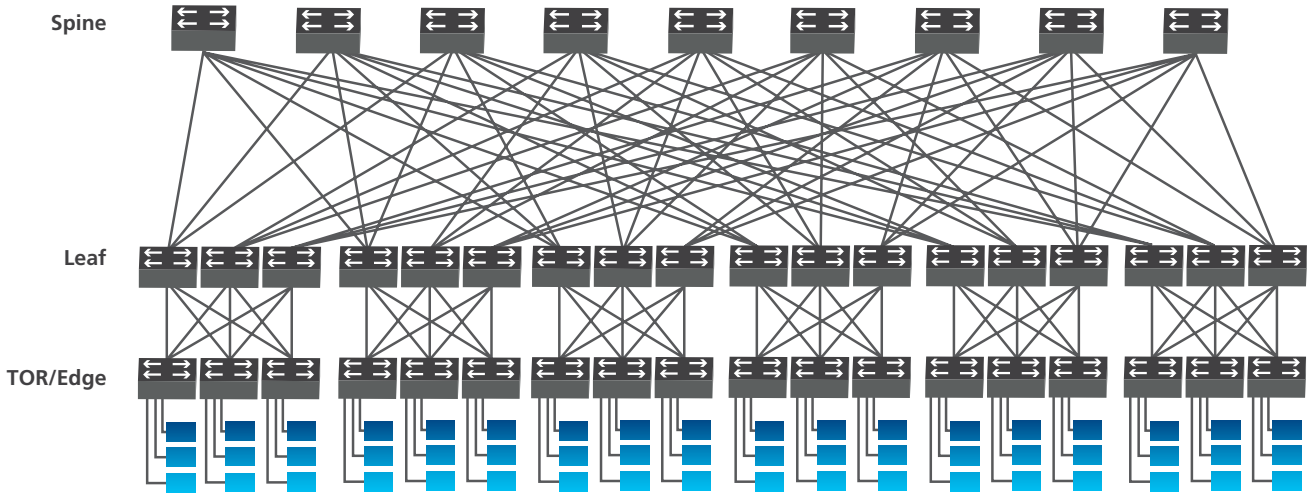
Torus topologies enhance fault tolerance and load balancing by connecting nodes in a multi-pathway grid-like structure.

Optical Connection Switch (OCS) for inter-layer connectivity

OCS switches interconnect different layers (i.e., core, aggregation, and compute clusters), enabling efficient, dynamic optical switching.

The benefits of hybrid topologies include independent layer scaling, high availability, and cost optimization, providing robust network solutions for modern AI data centers.

What is a Fat-tree Network?



Data centers commonly use fat-tree network topologies to provide the high bandwidth and low latency essential for AI workloads, especially those involving massive data transfers for training and inference. The design ensures efficient data flow, scalability, and fault tolerance, making fat-tree topoligies ideal for the demands of modern AI applications.

Key characteristics of fat-tree networks

Balanced Connectivity

Each switch in a fat-tree network typically comprises an equal number of uplinks (connections going “up” to higher layers) and downlinks (connections going “down” to lower layers). This balance maximizes available paths for data transfer and prevents bottlenecks.

Multiple Paths

The topology features multiple redundant paths between any two endpoints, enhancing fault tolerance and allowing for efficient load balancing and congestion avoidance.

Scalability

Fat-tree networks are highly scalable. Adding more switches at each layer can expand the network to accommodate thousands or even hundreds of thousands of servers without significant redesign.

Topology structure

Fat-tree hierarchies consist of three main layers:

Top of Rack (ToR) switches (Edge layer)

Role: These switches connect directly to the servers within each rack.
Function: Aggregate and forward server traffic to the leaf switches.
Alternate names: Also known as Edge Switches.

Leaf switches (Aggregation Layer)

Role: Serve as an intermediary between the ToR switches and the spine switches.
Function: Aggregate and forward traffic from multiple ToR switches to the spine layer.
Alternate names: Sometimes referred to as Brick switches or Middle of Row (MoR) switches.

Spine Switches (Core Layer)

Role: Act as the backbone of the network.
Function: Provide high-capacity interconnections between leaf switches, enabling any leaf switch to connect to any other leaf switch with minimal latency.
Alternate names: Also known as Core switches.

Note: In some contexts, the terms for these layers may vary. The key is understanding the hierarchical relationship and function of each layer.

Operation and advantages

Equal bandwidth allocation

Uniform performance: Fat-tree networks often employ the same bandwidth and similar switch hardware at all layers. This uniformity ensures consistent performance and prevents any single layer from becoming a bottleneck.

Cost-effectiveness: Due to economies of scale, using identical switches and transceivers simplifies procurement and reduces costs.

Simplified management: Standardization allows for easier network management, monitoring, and capacity planning.

Non-blocking performance

Aggregate bandwidth: Ensuring the total bandwidth available at each layer matches or exceeds the bandwidth requirements of the layer below achieves a non-blocking performance.

Optimized routing: Effective routing protocols distribute traffic evenly across all available paths, preventing congestion and maximizing throughput**.**

Load balancing and fault tolerance

Multiple paths: The network's multiple redundant paths allow for dynamic rerouting of traffic in case of link or switch failures.

Resilience: Continuous network performance is maintained even if individual components fail, which is crucial for mission-critical AI workloads.

Implementation considerations

Standard link bandwidth

Consistency: Commonly, operators select a standard link bandwidth (e.g., 100 Gbps) for use throughout the network.

Port distribution: Switches typically comprise an equal number of ports connected "up" to the next layer and "down" to the layer below.

Routing protocols and settings:

Optimization: Operators use routing protocols such as Equal-Cost Multi-Path (ECMP) to distribute traffic evenly across multiple paths.

Workload-specific tuning: Settings may be adjusted to optimize for specific workloads, such as AI training that involves a mix of broadcast and point-to-point communications.

Workload characteristics:

AI training traffic patterns: During training, there may be broadcast messages to all workers and targeted data transfers to specific workers.

Subscription levels: In large clusters, some tasks may be confined to a subset of nodes (pods), allowing for adjusted bandwidth requirements at higher layers.

Example in AI workloads

Distributed training

Communication needs: AI models are often trained across multiple servers that need to share parameters and gradients frequently.

Low Latency and high bandwidth: Fat-tree networks provide the necessary infrastructure to handle these intensive communication patterns efficiently.

Scalability

Expanding clusters: As the demand for computational resources grows, additional switches and servers can be added seamlessly.

Performance maintenance: The network's design ensures that adding more nodes does not degrade performance.

Advantages of fat-tree networks

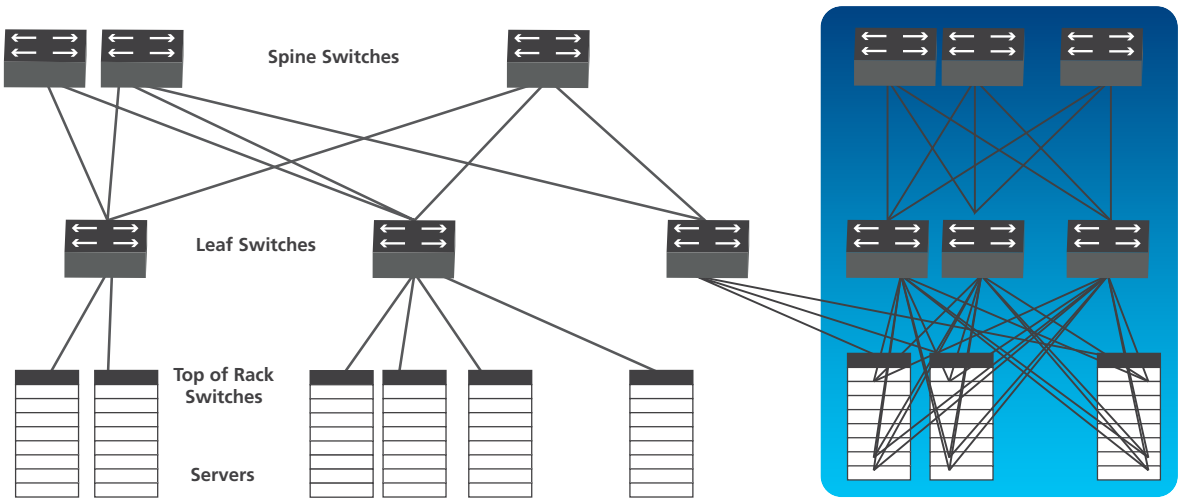
Uniform performance: Each layer can handle traffic without becoming a bottleneck, ensuring smooth data flow across the network.

Cost-effectiveness: Simplifies hardware procurement and reduces costs due to the use of standardized equipment.

Simplified management: Easier to monitor network performance and plan for expansion.

Scalability: Devices can be added to the network without complex reconfigurations, accommodating growth efficiently.

Front-end Network (FENW) vs. Backend Network (BENW)



3-Tier Clos Fabric

AI/ML Cluster with Back End Network

For AI/ML data centers, the Front-end Network (FENW) and the Backend Network (BENW) each serve essential and distinct purposes in optimizing data flow and processing efficiency.

Front End Network (FENW)

The Front-end Network (FENW) connects to every node (CPU), typically accompanied by a parallel management network that handles provisioning, orchestration traffic, and telemetry.

The FENW is the legacy hyperscale network specially designed for data management. The FENW links external connections, data storage, and various server types, facilitating seamless data access and integration.

Backend Network (BENW)

The Backend Network (BENW) connects each accelerator with up to 72 accelerators per node - typically 8 - interconnected within a scale-up network using technologies like Nvidia's NVLink and Google's ICI (while also incorporating hardware management for power distribution, cooling, status, and security networks). The switches can support multiple external connections; for instance, a 32-port switch can facilitate up to 128 links, enhancing the network's scalability.

The BENW shares model update information during training. By interconnecting multiple accelerators – including AI GPUs and TPUs – along with any associated training memory, the BENW creates a tightly-knit, low-latency network, enabling all node accelerators to share memory efficiently during compute cycles.

For instance, Nvidia's H100 DGX utilizes eight H100 accelerators, while the GB100 NVL72 features 72 GB200 accelerators. The strict latency requirements limit the maximum span to one or two racks, employing specialized network protocols and switches, often at the chip level. Google achieves optimized network performance with its Inter Chip Interconnect (ICI) technology[4].

4. Google

Connector selection

The very high density of connections per rack for AI/ML servers and switches and the total link counts push the data center designer towards multi-fiber connectors such as the MPO family and, increasingly, to the next generation very high-density connector types.

MF VSFF connectors, such as the MMC and SN-MT, usher in a new era new era of high-density, low insertion loss connectivity by combining a smaller, 16-¬ber, reduced Mechanical Transfer (MT)-style ferrule—Tiny MT (TMT) with the Very Small Form Factor duplex connector form factor, such as the MDC or SN connector.

This innovation represents a significant advancement in connector design, particularly in response to increasing counts and greater demand for higher port density. As transceivers for these connector interfaces come onto the market, MF VSFF connectors will become standard, driving speeds of 800G, 1.6T, and beyond. As such, they are likely to become the standard for On Board Optics (OBO) external connections, aligning with the roadmap for future connector designs that demand even higher density solutions.

Cable configuration

Data center operators must configure all cabling to optimize performance and maintain data integrity. Due to the significant bandwidth demands placed on AI data centers compared to traditional hyperscale infrastructures, AI facilities require high-quality cables, maximizing bandwidth while minimizing latency.
For this reason, qualities such as high-speed, reliable connectivity and resistance to electromagnetic interference make optical fiber cabling the ideal choice.

However, choice of cable is not the only concern. Proper cable management also includes considerations such as adopting a structured cabling system, which can mitigate clutter and improve airflow to maintain optimal operating conditions. Additionally, color-coding to simplify troubleshooting contributes to expedited incident management.

Transceiver and bandwidth choice

Choosing appropriate transceivers and bandwidth options is pivotal in meeting the high performance demands of advanced AI data center workloads. Due to factors such as high data transmission rates and simplified, plug-and-play compatibility within modern data centers (component compatibility helps prevent bottlenecks), popular transceiver choices include Small Form-factor Pluggable Plus (SFP+) and QSFP28.

Due to the ongoing and rapid evolution of AI data centers, bandwidth considerations must extend to future needs. For example, while standard 400GbE transceivers may meet current needs, adopting 800GbE presents operators with a forward-planning option.

Latency and throughput optimization

Efficient AI data center performance hinges on minimizing latency and maximizing throughput. By reducing data hops, network segmentation and high-performance switches can help resolve latency issues. Maximizing throughput requires adequate, high-speed connections and a network topology designed to manage tasks at peak demand. BENW paths are typically limited to 100m to constrain latency within the AI/ML cluster.

Quality of Service (QoS)

QoS prioritizes AI workloads to ensure network performance. QoS mechanisms serve to allocate bandwidth efficiently across high-priority tasks (e.g., real-time data processing), permitting only the most important tasks access to resources. Traffic shaping, policing, and queuing techniques commonly assist operators in managing network traffic. Appropriate QoS policies ensure AI data centers maintain consistent operations, even during periods of high demand.

Edge computing integration

By processing data closer to the source (meaning data does not need to travel long distances), edge computing reduces latency and bandwidth demands. Deploying edge computing solutions can enhance AI data center performance, particularly in relation to tasks requiring real-time processing (e.g., IoT devices, autonomous vehicles, etc.). Strategic edge nodes allow data centers to offload processing tasks, improving responsiveness. Edge computing can also play a role in data security – by enabling operators to store sensitive data closer to the origin, there is an inherently mitigated risk of data transmission interception.

Requirements for the Backend Network in AI Data Centers

In a 1:1 subscription ratio, each server or node connects to the network via a direct, dedicated link, without sharing bandwidth. This setup ensures that each server has unrivaled access to the full capacity of the network link, resulting in consistent performance, reduced latency, and high throughput.

Additional benefits for AI data centers include ease of scalability, as operators can add more direct connections as the network grows, and simplified network management by mitigating the failure-related variables associated with bandwidth sharing. However, a 1:1 ratio is not always needed, as many applications can effectively operate with shared bandwidth, allowing for cost savings and resource optimization without significantly impacting performance.

Basic requirements

Backend networks must provide non-blocking, lossless packet transmission while maintaining packet order (jitter must be minimized). Communication should not degrade processor performance – therefore, remote direct memory access (RDMA) is essential. This is a native capability of InfiniBand and can be achieved with RDMA over converged Ethernet (ROCE) protocol.

Efficient load balancing (congestion management)

Network control mechanisms and parameters must be optimized for specific training or inference models and job flow. Efficient load balancing involves the strategic distribution of computational tasks across resources (e.g., servers) to prevent single-point congestion and ensure optimal network performance.

The mechanism includes redundancy and failover measures, ensuring seamless rerouting in the event of a single-point failure to maintain service continuity. Efficient load balancing also maximizes efficient resource utilization.

Comprehensive Network Control for AI Training

Network control mechanisms and parameters must be optimized for specific training or inference models and job flow. Efficient load balancing (congestion management) prevents single-point congestion by strategically distributing computational tasks. However, load balancing is only one aspect of a much broader network control strategy:

Dynamic network demand

Network demand changes throughout training sequences, creating a variable necessitating responsive, highly adaptive control mechanisms.

Failure management

In long training sessions on large clusters, failures may occur. Effective network control requires robust monitoring (telemetry) to capture and store checkpoint model statuses, detect faults, and facilitate reloading and restarting processes.

Reliability in AI networks is more critical than in legacy hyperscale computing, especially regarding technologies such as Ethernet and InfiniBand – the networks support Remote Direct Memory Access (RDMA), enabling efficient, low-latency data transfers with high throughput.

ML training is a process necessitating synchronous subsystem operations. The failure resolution process can mean halting cluster processing, identifying the failure point, reconfiguring the accelerator array, reloading checkpointed interim model parameters, and restarting.

Operators wishing to enhance performance and reliability must consider scaling up (adding more resources to existing nodes) and scaling out (adding more nodes). To ensure optimal communication, each accelerator requires a dedicated network connection. Connection density per rack – at both the server and switch levels – plays a vital role in maximizing throughput and minimizing latency. This process must be automated to ensure seamless recovery following a failure.

Building a Clos Network with a Thousand, Ten Thousand, or even One Hundred Thousand End Points

Endpoints may refer to network connected servers, storage devices, routers, and user devices such as laptops. Building a Clos network to support one thousand endpoints, ten thousand endpoints, or even one hundred thousand endpoints is feasible simply by adding more leaf and spine switches to the Clos.

A well-architected Clos infrastructure enables vast and rapid scalability without significant redesign. Each additional spine switch increases bandwidth, minimizes data packet hops, and reduces latency. Also, rerouting traffic to bypass failed switches enhances system performance, fault tolerance, and redundancy.

We will use as an example a BENW for a very large AI/ML cluster with 131,072 end point accelerators (GPUs or TPUs). We will select switches with 64 ports per switch and use transceivers that can operate at 800G or 2x400G. The 400G links will connect to the end points and inter-switch links will be 800G.

Switches

Designing a non-blocking, three-stage Clos topology architecture for 100,000 endpoints involves balancing the number of switches required at the ingress, middle, and egress stages:

Edge stage	Assuming 64 inputs per switch, this fabric requires **2,048 switches.**
Aggregation stage	The middle-stage switch count the edge stage count: **2,048 switches.**
Egress stage	The number of core switches is half the count of aggregation switches (i.e., **1,024 switches**)
	Total (ingress + middle + egress): A Clos topology network supporting 131,072 endpoints requires **5,120 switches.**

Optical links per layer

Continuing with the example of a Clos topology designed to support 131,072 endpoints, each of the 2,024 edge switches connects to the aggregation switches with 32 links, resulting in 65,536 optical links. The same link count applies to the aggregation to core switch link for a total inter-switch link count of 131,072.

Power consumption

A power consumption per switch of 1,500 watts is assumed. Optical links also consume power but on a much smaller scale - around 30 watts per link, assuming 15 watts per 800G transceiver and two transceivers per link.

Switch power	5,120 switches × 1500 watts per switch = **7,680 kW**
Optical links	65,536 links x 300 watts per link = **1,966kW**

Combining these figures, the total power consumption across ingress, middle, egress, and optical links comes to **9,646 kW**

Number of server racks

We assume 32 accelerators per GPU/TPU rack and 16 2RU switches per switch rack.

This gives the following rack counts:

GPU/TPU	4,096 racks
Edge switches	128 racks
Aggregation switches	128 racks
Core switches	64 racks
Total	**4,416 rack**

In Conclusion...

At its core, this document seeks to establish a link between the growing demand for advanced AI technologies and the increasing complexity of AI-optimized data centers (a forthcoming 'Part 2' will delve deeper into more advanced aspects of AI data center operations).

Compared to traditional setups, AI data centers consume significantly more power and generate massive amounts of heat. This necessitates innovation in cooling and energy management strategies, highlighting the importance of adopting efficient fiber cabling solutions to support high bandwidth while minimizing heat-related performance inefficiencies. The shift from traditional air cooling to more effective liquid cooling methods further underscores the need for robust physical layer solutions capable of accommodating advanced cooling systems.

The discussion on network topologies and backend network design emphasizes the necessity of facilitating low-latency connections between accelerators – this is where AFL stands out as a player uniquely placed to meet the essential connectivity requirements for large-scale, cutting-edge AI and ML data center builds. With fewer than a handful of global companies capable of providing such specialized, state-of-the-art connectivity, AFL's fiber network solutions ensure scalability and adaptability to not only meet, but exceed the changing network demands of tomorrow.

As data centers continue to evolve, the need for high-performance, energy-efficient optical fiber solutions becomes paramount. AFL sets an industry benchmark in offering advanced networking and physical layer solutions, inspiring data center operators to invest in future expansion plans with confidence.

In summary, this document serves as a call to action for the data center industry, providing valuable insights for progressive stakeholders eager to adopt innovative AI and ML technologies.

Founded in 1984, AFL is an international manufacturer providing end-to-end network solutions to the energy, service provider, enterprise, hyperscale and industrial markets. The company's products are in use in over 130 countries and include fiber optic cable, assemblies, and hardware, transmission and substation accessories, outside plant equipment, connectivity, test and inspection equipment, fusion splicers, and training. AFL also offers a wide variety of services supporting data center, enterprise, wireless and outside plant applications.

Headquartered in Spartanburg, SC, AFL has operations in the U.S., Mexico, Canada, Europe, Asia and Australia, and is a wholly owned subsidiary of Fujikura Ltd. of Japan.