Turnkey Datacenter Leaf-Spine Integration Specification: Data Center Spine-Leaf Fabric Design Using 400G and 100G Switching

PLATFORM INTRODUCTION

Data Center Spine-Leaf Fabric Design Using 400G and 100G Switching represents a turnkey integration specification for next-generation data center architectures, purpose-engineered to deliver deterministic, ultra-low-latency any-to-any connectivity across hyperscale, cloud, and enterprise environments. This platform family provides a fully non-blocking, fixed-form-factor switching solution that unifies 400 Gigabit Ethernet spine tiers with 100 Gigabit Ethernet leaf tiers under a single operational model, eliminating the oversubscription bottlenecks and protocol complexity inherent in legacy three-tier designs. The architecture is validated for large-scale BGP EVPN-VXLAN fabrics, enabling seamless Layer 2 and Layer 3 overlay services with hardware-accelerated encapsulation and decapsulation at line rate.



DATACENTER RACK PREREQUISITES

The deployment of a 400G/100G spine-leaf fabric requires disciplined physical and environmental readiness to ensure rated performance and long-term reliability. Each rack position must provide dual independent power feeds from separate PDUs rated for the platform's maximum power draw, with hot-swappable redundant AC or DC power supplies installed in a 1+1 configuration. Thermal design assumes front-to-back or back-to-front airflow matched to the rack's hot-aisle/cold-aisle orientation; inlet temperatures must remain within 0°C to 45°C at operating altitude up to 3,000 meters. Structured cabling should use OM4 or OM5 multimode fiber for 100G SR4 distances up to 100 meters and single-mode fiber for 400G DR4/FR4 reaches of 500 meters to 2 kilometers. Rack rails, cable management arms, and grounding lugs must be installed prior to chassis mounting. All spine and leaf nodes should be

provisioned with out-of-band management connectivity via dedicated 1G RJ45 or SFP management ports, isolated from the in-band data plane.

LINK AGGREGATION PROTOCOLS

The fabric supports a comprehensive suite of link aggregation and multi-pathing protocols designed to maximize bisectional bandwidth and resiliency. IEEE 802.3ad Link Aggregation Control Protocol (LACP) enables active-active bundling of up to 32 member links per port channel across leaf-to-spine and leaf-to-server connections. Equal-Cost Multi-Path (ECMP) routing with 128-way hardware load balancing distributes traffic across all available spine nodes using flow-based hashing, supporting both Layer 3 and Layer 4 hash inputs. For overlay environments, the platform integrates BGP EVPN with VXLAN encapsulation, providing control-plane learning and ARP/ND suppression to reduce broadcast flooding. Additional supported protocols include MLAG (Multi-Chassis Link Aggregation) for dual-homed server connectivity, PFC (Priority Flow Control) and ECN (Explicit Congestion Notification) for lossless RoCEv2 transport, and BFD (Bidirectional Forwarding Detection) for sub-second failure detection across fabric links.

KEY FEATURES

- Non-Blocking Fabric Architecture: Delivers full line-rate forwarding across all

400G and 100G ports simultaneously, ensuring zero internal oversubscription and predictable latency under sustained load.

- Hardware-Accelerated VXLAN: Supports up to 256,000 tunnel endpoints with line-rate encapsulation/decapsulation, enabling scalable multi-tenant overlay networks without CPU intervention.

- Dynamic Load Balancing: 128-way ECMP with resilient hashing and flowlet-based rebalancing minimizes flow polarization and adapts to topology changes without packet reordering.

- Advanced QoS and Buffering: Per-port and per-queue dynamic buffer allocation with 64 MB shared packet memory supports lossless RoCEv2, strict priority scheduling, and Weighted Random Early Detection (WRED).

- Comprehensive Telemetry: Streaming telemetry, sFlow, and in-band network telemetry (INT) provide real-time visibility into fabric health, latency, and microburst events.

- High Availability: Redundant hot-swappable power supplies and fan trays, plus graceful restart and non-stop forwarding, ensure sub-second convergence during maintenance events.

- Automated Fabric Provisioning: Zero-touch provisioning (ZTP) and declarative intent-based configuration reduce deployment time and eliminate human error in large-scale spine-leaf rollouts.

VERIFIED SPECS

Each platform in the 400G/100G spine-leaf family is validated against rigorous performance benchmarks, including zero-loss forwarding under 100% line-rate load, latency below 800 nanoseconds port-to-port, and full compliance with IEEE 802.3bs (400GbE) and IEEE 802.3cd (100GbE) standards. Switching capacity scales from 12.8 Tbps in 1RU leaf configurations to 25.6 Tbps in 2RU spine configurations, with packet buffer sizes ranging from 32 MB to 64 MB depending on SKU. All models support jumbo frames up to 9,216 bytes, IPv4/IPv6 dual-stack routing, and MPLS L3VPN for inter-DC connectivity. Power consumption is optimized through intelligent fan speed control and energy-efficient silicon, with typical operational draw between 350W and 1,200W per chassis.

Parameter	**Specification**
Form Factor	1RU Leaf Switch / 2RU Spine Switch
Switching Capacity	12.8 Tbps (Leaf) / 25.6 Tbps (Spine)
Port Configuration	Leaf: 32x100G QSFP28 + 8x400G QSFP-DD; Spine: 64x400G QSFP-DD
Packet Buffer	32 MB (Leaf) / 64 MB (Spine) Shared
Forwarding Latency	< 800 ns Port-to-Port
Power Supply	Dual Hot-Swappable Redundant AC/DC (1+1)

Airflow	Front-to-Back or Back-to-Front (Field-Replaceable Fan Trays)
Management Ports	1x RJ45 Console, 1x 1G SFP Management, 1x USB
VXLAN Tunnel Endpoints	Up to 256,000
ECMP Paths	128-Way Hardware Load Balancing
Operating Temperature	0°C to 45°C at Sea Level to 3,000m
Compliance	IEEE 802.3bs, 802.3cd, 802.1Q, 802.3ad, RoHS-6, CE, FCC, UL

STACKING DEPLOYMENT LOGIC

The fabric architecture supports two complementary stacking and virtualization models to simplify management and scale operations. For leaf-tier deployments, up to eight switches can be interconnected in a logical stack using 100G or 400G stacking ports, presenting a single management plane and control plane to the network operator. This stacking model enables simplified configuration, unified software upgrades, and cross-stack link aggregation for dual-homed servers. For spine-tier deployments, the platform supports Multi-Chassis Trunking (MCT) and distributed anycast gateway, allowing multiple spine switches to operate as a single logical endpoint for VXLAN

tunnels and BGP peering. This design eliminates the need for complex protocol tuning during capacity expansion and ensures that server-facing leaf switches can be added or removed without disrupting existing traffic flows. Combined, these deployment logic models deliver a fabric that scales horizontally from a single rack to hundreds of racks while maintaining operational simplicity.

