Software Defined Fabric™

Highlights

Software Based, Automated Data Center Fabric

- Autonomous Leaf/Spine switch provisioning and management
- Supports hardware level network virtualization and isolation
- Self-healing (automated remediation)
- Automated software updates and upgrades

Standard Linux Operating System

- Standard upstream kernels without any patches
- Same s/w orchestration model and management tools across networking, compute and storage
- Ease of rollback[2] to previous version of the OS and simple deployment updates

Supports multiple networking white boxes

Open Networking and APIs

- Netconf API and YANG Modeling
- Orchestration agnostic; plugins for OpenStack and K8s

Autonomous Provisioning

- Zero-touch provisioning fabric
- Minimal to no human intervention
- Provisions in minutes instead of hours or days

Fully Self-Forming and Self-Discovering

- Automated discovery of network topology

Performance

- Full non-blocking folded CLOS topology
- Throughput of up to 6.4 Tbps per switch
- Ultra-low latency <500 ns

Fully Programmable Fabric

- Uses the industry standard P4 programming language
- Enables addition of new features and services in runtime[2]
- Avoids vendor lock-in
- Eliminates wait times for chip upgrade cycles

Continued ›

Product Oveview

Software Defined Fabric™ is a fully automated, industrialized software solution for networking white boxes, and designed specifically with the vision of virtualizing the data center. Designed for hyperscale, as well as distributed data centers, the solution is targeted to clients in the Data Center Operator (DCO), Telco, Enterprise, Cloud and Gaming industries. As a fully programmable fabric, it allows advanced networks functions to be deployed at scale with maximum performance and lowest cost.

Software Defined Fabric[1] brings forth a new paradigm in networking, leveraging automation, programmability, open networking standards and white box hardware. The solution offers the following key high-level benefits:

- A high performance and adaptable fabric architecture that provides a significant increase in throughput and power efficiency and reduction in latency
- Distributed Kubernetes/OpenShift based control plane for better resiliency
- Future-proof; programmable control and data planes
- Operational consistency; uses same standard Linux distribution as that used by compute and storage nodes in the data center and is managed as a "server node" to enable unified OS updates of the datacenter
- Enhanced support for network "slicing" enabling allocation of physical resources into multiple autonomous and secure (vFabrics)
- Optimized platform for VM and Container based technologies and workloads
- A native IPv6 based networking environment fully capable of supporting both IPv6 and IPv4 traffic
- Fully automated fabric with advanced self-forming and self-discovery capabilities, zero-touch provisioning of the virtual networking and virtual components with automated software updates

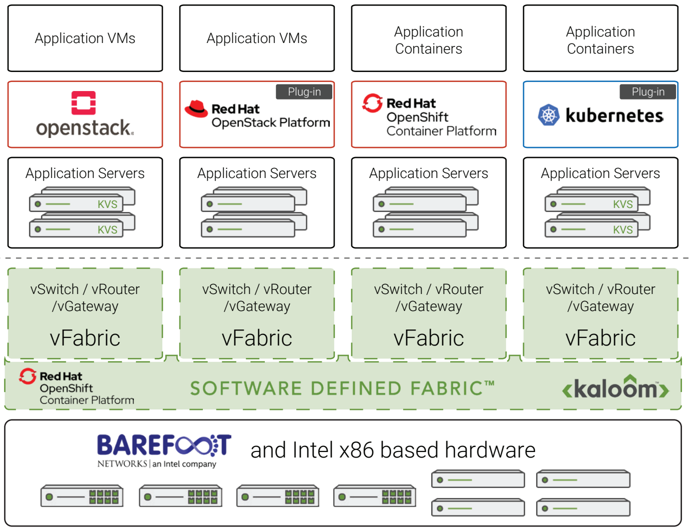

Figure 1: Software Defined Fabric - High Level Block Diagram

Virtual Environment Optimized

- Native support for virtual machine and container-based workloads

Robust, Resilient and Fully Containerized Control Plane

High Scalability

- Native IPv6, multi POD, cloud and multi data center solutions

Supports Kaloom Virtual Switch (KVS)[1] Software

- Virtual Switch offload; optimized virtual switch replacement for OVS-DPDK
- Minimizes compute resources, reduces user space to user space latency, increases data throughput

Embedded with Kaloom's High-Performance Virtual Router (vRouter)[1]

- Reduces CAPEX; eliminates the need to purchase and deploy physical routers or expensive software based virtual routers running on compute servers
- Maximizes performance; data forwarding within a given vFabric done at the Software Defined Fabric's line rate and not in a software appliance
- Support for EVPN
- Support for BGP and OSPF[2]

Integrated with Kaloom's Virtual Gateway (vGW)[1] embedded within the vRouter

- Provides VxLAN gateway functionality
- Enables multi-tenancy
- Enables connectivity of L2 domains over a L3 network

flowEye™ In-band Network Telemetry and Analytics[1] solution support

- Real-time in-band network telemetry of actual (not sampled or synthetic) traffic flow data collected via hardware or software-based sensors
- Analytics on aggregated collected data for display via different dashboard views
- Data visualization with 100ms precision

Segment Routing (SRv6) support[2]

High Availability

- Fabric/vFabric redundancy
- Active-Active Control and Data Plane
- N+M redundancy

Scale-Out Optimization

- Equal cost Multi-Pathing (ECMP) for Traffic Balancing
- Support for MLAG
- Multi-fabric expansion capability[2]

For more detailed information please refer to *https://www.kaloom.com/product-collateral*

Software Defined Fabric Topology

Software Defined Fabric is built on a 3-stage folded CLOS and fully non-blocking topology (see figure 2) also referred to as a Leaf and Spine architecture where "Spine" represents the middle stage of the CLOS topology and "Leaf" switches represent the input and output stages.

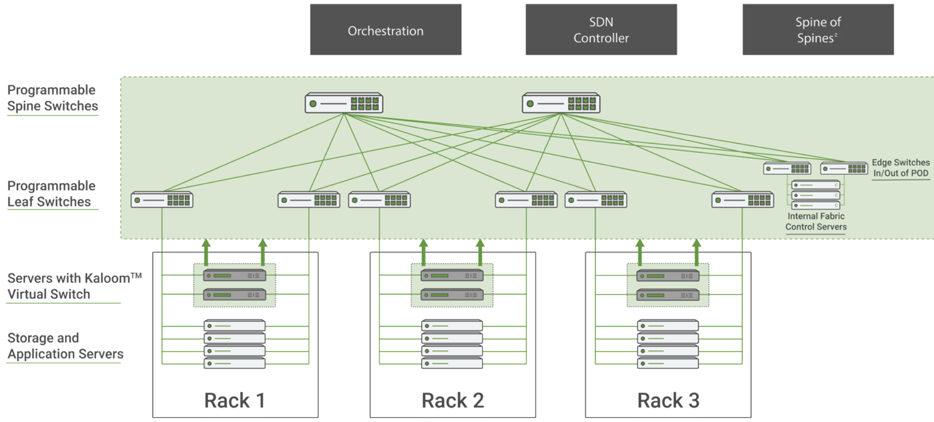

Figure 2: Software Defined Fabric Topology

Figure 2 illustrates several application servers connected to the leaf switches, with each server being connected to two separate leaf switches for redundancy. Leaf switches may have multiple paths to every server and connect to all the spine switches for maximum connectivity options. The Software Defined Fabric is controlled by a cluster of Controllers together with distributed control plane functionality in each network node, to create a distributed Kubernetes based control plane for the fabric. The controllers have one serving as the Master Controller to control quorum and the remaining nodes used to deploy networking services (i.e., vRouters, etc.).

Figure 3 depicts the physical view of a typical POD deploying the Software Defined Fabric as a networking solution and contains the following components:

Spine Switches – Connects all the leaf switches. Each leaf switch is connected to all spine switches in the upper layer.

Edge Leaf Switches – Provide the interface to exit the Software Defined Fabric and forward the user data traffic towards the Wide Area Network (WAN). They can be co-located with other Leaf Switches

Application Servers – Connect to the Software Defined Fabric leaf nodes and run various data center workloads.

Fabric Controller – Runs in the controller servers to manage and control all the nodes of the Software Defined Fabric.

Leaf Switches – Connects the application servers via one or multiple paths.

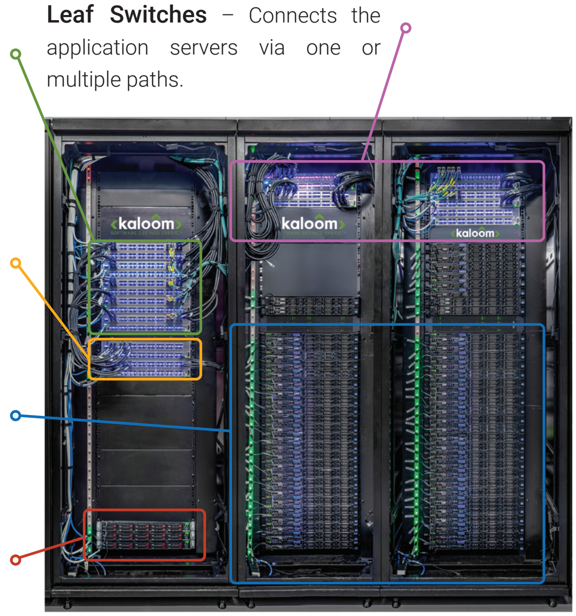

Figure 3: Software Defined Fabric Physical View

In the application server racks, each end-user server used for storage or computing applications is connected to two separate leaf switches for redundancy. The POD can be expanded by adding several new application server racks connected to the fabric through additional leaf switches.

The Leaf and Spine Switches are based on the Intel Tofino chipset (previously Barefoot Networks). The Intel Tofino device is fully programmable using the P4 Programming language.

Software Defined Fabric Topology - Hardware Guidelines[3]

The following information in table 1 highlights some of the key high-level hardware guidelines regarding the main components of the Leaf and Spine architecture of the Software Defined Fabric.

Component	Hardware Guidelines
Leaf and Spine Nodes	Switch models validated and certified (or pending) to be interoperable with Kaloom software **Accton (Edge-core)[3,4]** • Accton (Edge-core) 100BF-65X, with 65 100G ports for Fabric connectivity • Accton (Edge-core) 100BF-32X, with 32 100G ports for Fabric connectivity • Accton (Edge-core) 100BF-32Q, with 32 100G ports for Fabric connectivity Ports: • 32/65 QSFP28 ports, each port supports 4 x 10/25 GbE per port using splitter cables • Management ports: 1 x RJ-45 serial console, 1 x RJ-45 1G BASE-T management, 1 x USB Type A storage Performance: • Switching capacity: 3.2/6.4 Tbps • Forwarding Rate: 4.7/9.5 Bpps • Jumbo frames support up to 9000 Bytes • Packet Buffer Size: 22MB for 100BF-65X, 100BF-32X, and 20MB for 100BF-32Q integrated packet buffers Software: • Pre-loaded with Open Network Install Environment (ONIE) software installer **Inventec[2,4]** • D10056 Datacenter Switch Ports: • 48x25 GbE SFP 28 + 8x100GbE QSFP28 Performance: • 2B Packets per Second • 2.0 Tbps Throughput • Line Rate L2/L3 Forwarding • 22MB Packet Buffer Software: • Pre-loaded with Open Network Install Environment (ONIE) software installer The Spine and Leaf nodes in the Software Defined Fabric require switches that use: Key Components Switch Silicon: Barefoot Tofino (Intel) - Programmable ASIC CPU Modules: • Intel® x86 Broadwell-DE, Pentium D-1517 • 4/8/16 GB SO-DIMM DDR4 RAM • 16 MB SPI Flash • 8 GB NAND Flash • 32 GB M.2 SSD
Controller Nodes	The Controller nodes in the Software Defined Fabric use standard Intel Xeon based server hardware with the following minimum requirements: • Dual CPU; Intel® Xeon processor family with at least 8 cores per CPU • 128 GB RAM or more • 300 GB SSD or larger • Intel Ethernet Network Adapter XXV710-DA2 (2 x 25GbE ports) for fabric connectivity • Dedicated 1 GB Ethernet port for host management connectivity • Dedicated BMC Ethernet port for out-of-band management connectivity
Application Servers	An Application Server's only requirement is that it must have at least one Network Interface Card (NIC) that supports either 50 GbE, 25GbE or 10GbE ports. The following NIC models have been validated to work with Software Defined Fabric. • Intel® Ethernet Network Adaptor XXV710-DA2 – (2 x 25GbE ports) • Intel® Ethernet Converged Network Adaptor X710-DA2 – (2 x 10GbE ports) • Intel® Ethernet Converged Network Adaptor X520-DA2 – (2 x 10GbE ports) • QLogic® FastLinQTM QL45212-DE Series Intelligent Ethernet Adapters – (2 x 25GbE ports)

Table 1: Hardware Guidelines

Supported Standards

The following information in table 2 provides a list of the standards to which Kaloom is compliant.

Standards	
IETF RFCs	• RFC 8200 Internet Protocol, Version 6 (IPv6) Specification • RFC 826 Address Resolution Protocol • RFC 8201 Path MTU Discovery for IP version 6 • RFC 4861 Neighbor Discovery for IP version 6 (IPv6) • RFC 4862 IPv6 Stateless Address Autoconfiguration • RFC 4941 Privacy Extensions for Stateless Address Autoconfiguration in IPv6 • RFC 2464 Transmission of IPv6 Packets over Ethernet Networks • RFC 4291 IP Version 6 Addressing Architecture • RFC 4443 Internet Control Message Protocol (ICMPv6) for the Internet Protocol Version 6 (IPv6) Specification • RFC 8106 IPv6 Router Advertisement Options for DNS Configuration • RFC 2328 OSPF Version 2 • RFC 3101 The OSPF Not-So-Stubby Area (NSSA) Option • RFC 4271 A Border Gateway Protocol 4 (BGP-4) • RFC 4760 Multiprotocol Extensions for BGP-4 • RFC 1997 BGP Communities Attribute • RFC 2796 BGP Route Reflection An alternative to full mesh IBGP • RFC 2842 Capabilities Advertisement with BGP-4 • RFC 2385 Protection of BGP Sessions via the TCP MD5 Signature Option • RFC 3065 Autonomous System Confederations for BGP • RFC 3315 Dynamic Host Configuration Protocol for IPv6 (DHCPv6)[2] • RFC 5340 OSPF for IPv6[2] • RFC 5880 Bidirectional Forwarding Detection (BFD)[2] • RFC 5881 Bidirectional Forwarding Detection (BFD) for IPv4 and IPv6 (Single Hop)[2] • RFC 7348 Virtual eXtensible Local Area Network (VXLAN) • RFC 6241 Network Configuration Protocol (NETCONF) • RFC 8040 RESTCONF Protocol[2] • RFC 8341 Network Configuration Access Control Model[2] • RFC 5277 NETCONF Event Notifications • RFC 7950 The YANG 1.1 Data Modeling Language • RFC 6991 Common YANG Data Types • RFC 8342 Network Management Datastore Architecture (NMDA) • RFC 8345 A YANG Data Model for Network Topologies • RFC 8349 A YANG Data Model for Routing Management (NMDA Version) • RFC 7432 BGP MPLS-Based Ethernet VPN[2] • RFC 8343 A YANG Model for Interface Management • RFC 8365 Ethernet VPN (EVPN) as a Network Virtualization Overlay (NVO) solution • RFC 8525 YANG Library • RFC 8528 YANG Schema Mount[2] • RFC 6022 YANG Module for NETCONF monitoring
IEEE	• IEEE 802.1 AB Link Layer Discovery Protocol • IEEE 802.1Q VLAN Tagging • IEEE 802.3ad Link Aggregation with LACP • IEEE 802.3ba 40 and 100 Gigabit Ethernet Architecture • IEEE 802.3bm 40 and 100 Gigabit Ethernet Physical layer • IEEE 802.3ab 1 Gigabit Ethernet • IEEE 802.3ae 10 Gigabit Ethernet • IEEE 802.3by 25 Gigabit Ethernet • IEEE 802.3cd 50 Gigabit Ethernet • IEEE 802.3bs 200 and 400 Gigabit Ethernet[2]

Table 2: Supported Standards

Notes & References

[1]For more detailed information please refer to ***https://www.kaloom.com/product-collateral***

[2]Roadmap item

[3]For more detailed information regarding Hardware Guidelines including requirements for Deployer Host, Management Switch, Console Server, Application Server OS, Cables and Transceivers. etc. please contact your Kaloom sales representative.

[4]Please refer to manufacturer's data sheet for complete product specifications.

 Document Version - 1.0 Publication Date: 11.12.2019