# Dell Networking Z9500 fabric switch:
## Data center use cases with “Pay-As-You-Grow” licensing

**This document is for informational purposes only and may contain typographical errors and technical inaccuracies. The content is provided as is, without express or implied warranties of any kind.**

© 2014 Dell Inc. All rights reserved. Dell and its affiliates cannot be responsible for errors or omissions in typography or photography. Dell, the Dell logo, Dell Networking, Dell Active Fabric, Dell Networking Z9500, are trademarks of Dell Inc. Other trademarks and trade names may be used in this document to refer to either the entities claiming the marks and names or their products. Dell disclaims proprietary interest in the marks and names of others.

July 2014| Rev 1.0

# Contents

# Overview

One of the issues IT organizations face today is how to right size the networking fabric to meet both current and future requirements. In the traditional networking architecture, organizations will purchase two chassis for redundancy, then add line cards to meet their current needs. They then have the ability to add line cards as they grow. This is financially inefficient since there is a major capex expense up front in order to purchase the entire chassis. Companies may then find their current requirements could be as few as a single line card, greatly underutilizing the chassis, yet needing to absorb the full upfront capex for unpredictable future requirements.

Using fixed form factor switches as an alternative, customers can use a utilizing Dell Active Fabric with a leaf/spine architecture in order to overcome the limitations of a chassis-based, multi-tiered data center. This is a cost-effective method for customers to scale out their data center without the impact of a large upfront capital expense imposed by the traditional chassis-based data center topology. It also flattens the data center, putting networking closer to computing functions essential for modern workloads. Dell Active Fabric is proven to reduce costs by up to 70%, and operational expenses by up to 30% over traditional chassis-based architectures[1].

One drawback of a fixed form factor switch fabric is the established number of ports and port sizes (e.g. 1/10/40GbE). As a customer's networking needs grow, additional boxes can be added to meet demand. Although this is a more flexible, cost-effective method to the traditional chassis-based architecture, it adds the additional expense and time requirements for deploying the larger topology. Particularly, as businesses inevitably face the need to rapidly scale to support hybrid cloud infrastructure deployments and other modern workloads, networking port requirements may rapidly increase beyond an IT organization's current needs.

# Evolution to a "Pay-As-You Grow" fabric for network expansion

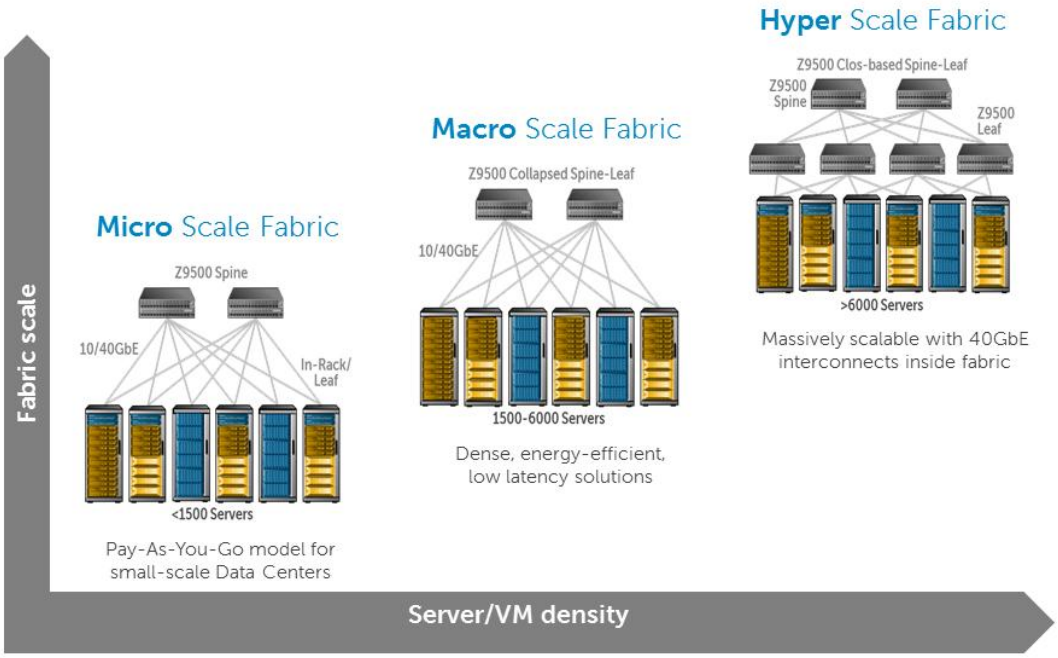
Dell has solved these issues with an innovative scaling model called "Pay-As-You-Grow" licensing. This method introduces three different SKUs within a single fixed form factor switch. The Dell Networking Z9500 fabric switch is the first switch on the market to utilize a 3 SKU licensing approach, and includes 132 40GbE ports with the flexibility to breakout up to 528 ports of 10GbE.

Understanding that for smaller deployments, this many ports may exceed current requirements, Dell offers the "Pay-As-You-Grow" licensing arrangement allowing customers to license just 36, 84 or all 132 ports as three separate SKUs, all within the same switch. This allows IT organizations with relatively smaller deployments to take advantage of the Z9500's higher performance[2] and exceptional energy efficiency[3], by initially licensing a "right-sized" port amount for their current needs. It also offers the flexibility to easily increase to a higher port SKU through Dell's innovative licensing model, giving customers choice and flexibility to license just the 36 ports with the option to license 84 or all 132 ports in at a later date. This provides one of the most effective means in the market today for customers to avoid overspending for unnecessary infrastructures while ensuring a deployment already in place to easily and efficiently expand with future business needs.

[1] "Rightsizing the Enterprise Data Center Network", Gartner, March 20, 2013
[2] 10 Tbps at 40% lower latency than the Cisco Nexus 9508
[3] 1/2 the power consumption per 10GbE port versus the Cisco Nexus 6k

# Micro, macro and hyper scale deployments

With the Z9500, three different types of fabrics can be built utilizing "Pay-As-You-Grow" licensing, including a micro scale fabric, a macro scale fabric and a hyper scale fabric.




The micro scale fabric can be realized by using the lower port count SKUs on the Z9500 such as the 36 port SKU or 84 port SKU. This allows a smaller network deployment of 10 to 20 racks using two Z9500s in the core connected by Dell's virtual length trunking (VLT) technology for MLAG. It can also use Layer 3 routing technologies, aggregating two devices with a small number of rack switches to provide the head room to migrate from micro scale to macro scale while maintaining core switch density.

Macro scale fabrics can be utilized for mid-size deployments (typically 1500 to 6000 servers). IT organizations can build two Z9500s with as many as 132 ports enabled with little or no oversubscription (1:1). Reduced oversubscription can be achieved using a large number of devices, building a macro scale fabric with two devices at the core for simplified manageability. The Z9500 can also be deployed in a collapsed leaf/spine architecture with two Z9500s functioning as both core and aggregation layers, with direct connection to the server racks in a Clos formation. This topology would be ideal for a variety of demanding environments such as high-performance computing (HPC), enabling high density, high energy efficiency and extremely low latency deployments.

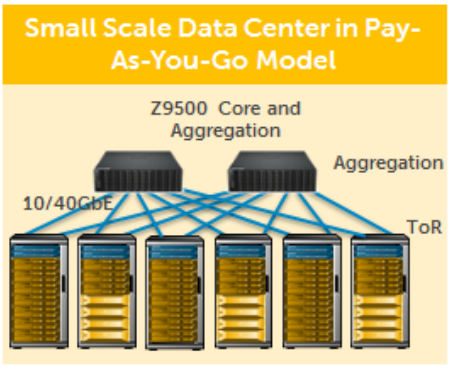
For very large networks, customers are able to build a hyper scale fabric with a high density of compute nodes. With this fabric, leaf and spine architecture is utilized, allowing scalability from 6000 to as many as 100,000 servers with all 132 ports enabled on the Z9500. This topology is highly optimized for accelerating east/west traffic, facilitating faster server-to-server communication and optimizing virtual machines (VM) migration. Utilizing the 40GbE interconnects within the fabric, the hyper scale fabric is ideal for Web 2.0, virtualization and cloud workloads.

# Use cases

Following is a list of specific use cases for the Z9500, leveraging "Pay-As-You-Grow" flexibility in combination with many of the other features of the Z9500 such as VLT, ultra-low latency, minimal jitter, large table capacities, automation features, Data Center Bridging[4] (DCB) support, L2/L3 flexibility and a wide range of additional protocol support.

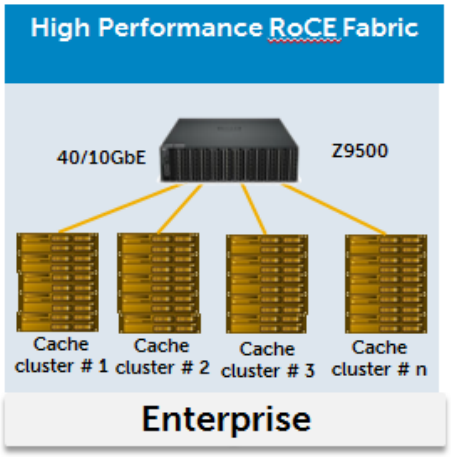
## Micro scale

### Small scale data centers



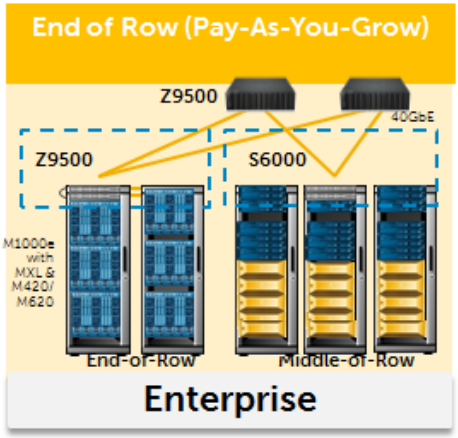

This use case is related to the small scale data center/micro scale fabric. With Dell Networking VLT, customers can build a very small Layer 2 fabric of less than 1500 servers utilizing 10G or 40G uplinks from the top-of-rack (ToR) into the core. For increased performance, oversubscription to top of rack can be eliminated for additional scalability. With this use case, customers can design for very high performance and scale while optimizing costs utilizing the "Pay-As-You-Grow" pricing model. Users can also cascade two VLT domains, including routed VLT and VLT supporting unicast and multicast routing.

### RoCE converged LAN/SAN




The Z9500 offers a few advantages when supporting Layer 3 RDMA over converged Ethernet (RoCE)[9] capabilities with DCB for Layer 2 RoCE capabilities. One advantage is it that enables a single converged fabric for supporting RoCE while utilizing the ultra-low latency of the Z9500. It also allows for predictable end-to-end RoCE performance while delivering capabilities including priority-based flow control (PFC), PFC based on differentiated services code point (DSCP), flex hashing and purpose status-based tracking – all while monitoring utilization of buffers on a port queue basis. The Z9500 also provides the ability available to use explicit congestion notifications (ECNs) in conjunction with PFC.

### End of row (EoR)

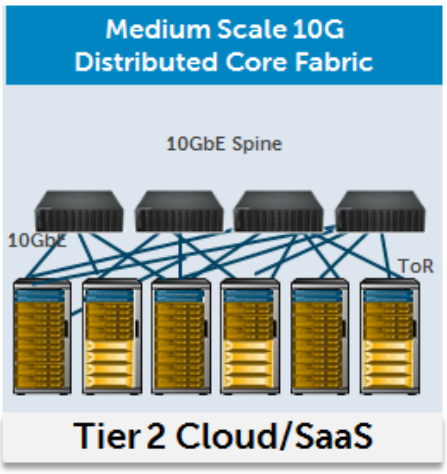



End-of-row aggregation of 40GbE links from the Dell Networking MXL switch, the PowerEdge M I/O Aggregator switch, and the Dell Networking S6000 switch delivers lower oversubscription. The "Pay-As-You-Grow" model is particularly effective for this use case in enabling companies to start small and add capacity within a row while using additional ports on the same device as needed. High density virtualization is also important with end-of-row aggregation, enabled by the Z9500's large table capacities.

[4] Available with future software update.

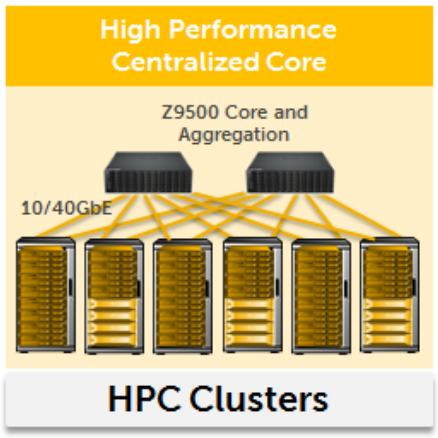
## Macro scale

### Tier 2 Cloud/SaaS



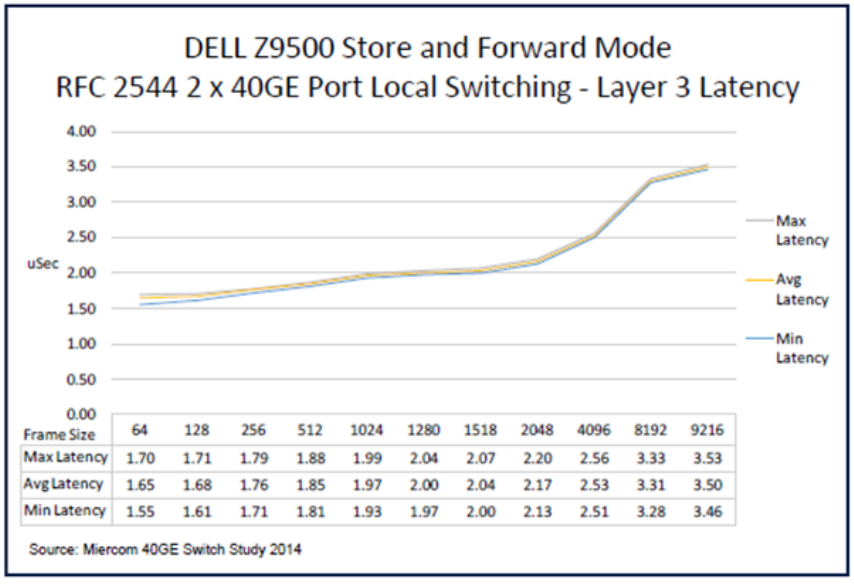

A four post design with four Z9500s for 1500 to 6000 servers, could support 10GbE core within a 1GbE server environment. Customers could also use 10GbE and 40GbE uplinks for a 10GbE server environment. The topology could be designed with BGP or open shortest path first (OSPF) with excellent scaling, both with IPv4 and IPv6. With this use case, either VLT or enhanced VLT can enable a larger VM live migration. This topology allows for strong performance at line rate for both 10GbE and 40GbE. 10GbE is available through cable breakout boxes[5], enabling 40GbE ports to be broken out into 4x 10GbE ports through a secure, highly reliable breakout cabling solution. In addition. a cable routing solution[6] is available to reduce cable clutter in front of the Z9500 switch, providing a reliable, serviceable and safe cabling solution for complex rack installations.

### High-performance computing (HPC)




With the HPC use case, customers can take advantage of significantly lower latency versus widely deployed competitive products in the market today. The Z9500 latency ranges from 600ns to 1.8µs, verified through third-party testing (figure 1).

In addition, the Z9500 has substantially low jitter rates (Figure 2). Along with the highest per-port density in the industry, the Z9500 is able to support exceptional performance for HPC applications.




| Frame Size | 64 | 128 | 256 | 512 | 1024 | 1280 | 1518 | 2048 | 4096 | 8192 | 9216 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Max Latency | 1.70 | 1.71 | 1.79 | 1.88 | 1.99 | 2.04 | 2.07 | 2.20 | 2.56 | 3.33 | 3.53 |
| Avg Latency | 1.65 | 1.68 | 1.76 | 1.85 | 1.97 | 2.00 | 2.04 | 2.17 | 2.53 | 3.31 | 3.50 |
| Min Latency | 1.55 | 1.61 | 1.71 | 1.81 | 1.93 | 1.97 | 2.00 | 2.13 | 2.51 | 3.28 | 3.46 |

***Figure 1***
*Average latency for the Dell Networking Z9500 with full fabric traversal store and forward mode ranged from 1.65 to 3.60µsec*
(Miercom Report Dell Networking Z9500, July 2014)

[5] Dell Networking Cable Breakout System Guide
[6] Dell Networking Cable Management Kit Guide

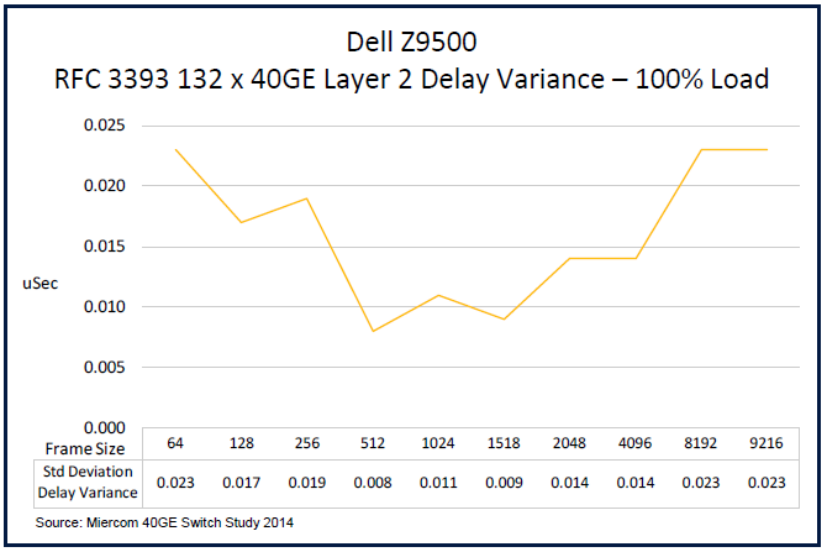
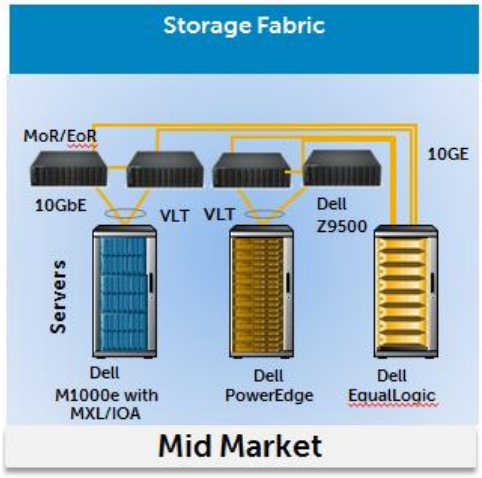



**Figure 2**
*The Dell Z9500 with eleven 12 Port 40G line cards exhibited very little variance in latency - also called "jitter" – less than one quarter of a microsecond on average for all packet sizes up to 9,216-bytes. The Dell switch configured with 132 ports was subjected to a100 percent traffic load. Layer 3 IP unicast traffic was used for the specified frame size. Tests were conducted in accordance with RFC 3393.*

For virtualized applications like VDI requiring very high capacity ARP tables, the Z9500 can support higher numbers of VMs than are posted on servers for virtual desktop infrastructure (VDI) applications. Any application that involves very high density virtualization can take advantage of the Z9500 enhanced hardware table sizes.

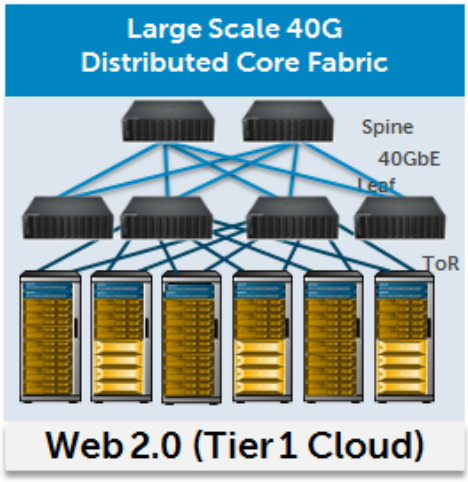
## iSCSI converged LAN/SAN




The storage iSCSI use case enables a single wire to carry LAN and SAN iSCSI traffic utilizing full DCB. This allows separation of LAN and storage traffic on the Z9500 while also leveraging some of the VLT capabilities of the Z9500 to build a resilient converged fabric. A combination of DCB and iSCSI optimizations provides very predictable iSCSI performance over a 40GbE network.

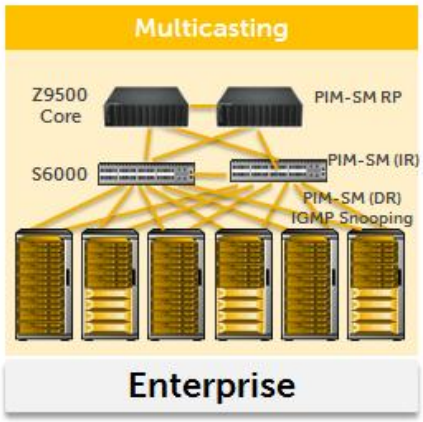
## Hyper scale

### Cloud/Web 2.0




For cloud designs scaling to greater than 6000 servers, an ideal deployment is a massively scalable leaf-spine architecture. A key issue for this use case relates to the high capacity of routing cables and compute nodes. With the high port count in the fabric, both the control plane and the routing tables need to scale. The Z9500 supports 16k IPv4 routes and through algorithm longest prefix match (ALPM)[7], the Z9500 can scale IPv4 to 128k routes and IPv6 to 32k routes. For customers building Layer 2 networks, the address resolution protocol (ARP) table is able to support up to 96k. This allows for a much larger Layer 2 domain, significantly scaling up the data plane/forwarding plane capacities. In the context of the hyper scale/Web 2.0 use case, it's also important to take advantage of automation features such as scripting mechanisms, representational state transfer APIs (REST APIs), smart script capabilities and bare metal provisioning.

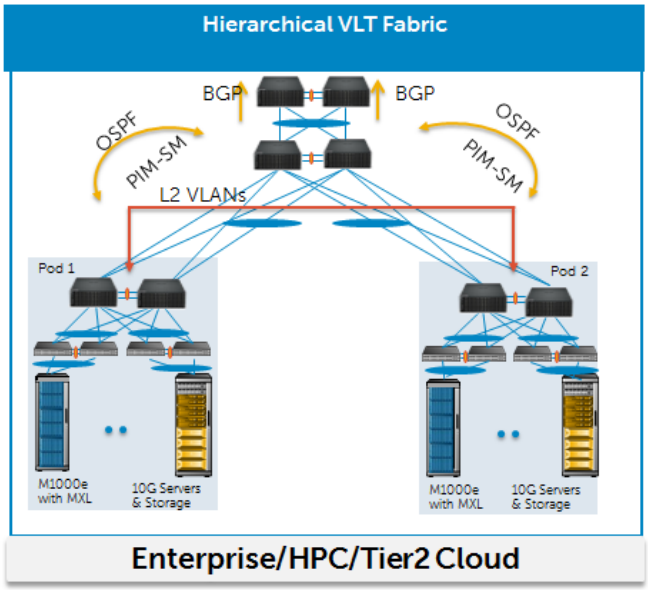
### Multicasting




With the multicasting use case, pure Layer 3 multicasting architectures can support the protocol independent multicast sparse mode (PIM-SM) rendezvous point (RP), PIM Intermediate Routers (IR), and PIM Designated Routers (DR). On the top of rack, internet group management protocol (IGMP) snooping can scale to 8k source in group pairs with capacity improvement capabilities on multicast routing. The Z9500 can support both shade trees with PIM-SM as well as source-specific trees with PIM source specific multicast (PIM-SSM). The multicasting architecture can also be done in hybrid mode, supporting a hybrid Layer 2 and Layer 3 mode with routed VLT. The RP can also be part of the VLT domain if running VLT.

### Hierarchical VLT

Hierarchical VLT fabric use cases can be deployed broadly in enterprise, HPC, and Tier2 clouds where VLT is required at every layer of the network. There are various examples of how VLT would be deployed across these use cases. One is that VLT can be applied from the rack or blade server to the top of rack/end of row. Second the VLT is able to be applied from the top of rack/end of row through the aggregation layer and into the Z9500 core. The L2 L3 boundary can be pushed all the way to the edge of a POD or inside a POD using routed VLT. With routed VLT, unicast routing protocols like OSPF or multicast routing protocols such as pretty good privacy (PGP) with IPv4, and IPv6 can be




[7] Available with future software update.

deployed. Virtual router redundancy protocol (VRRP) can also be utilized without any routing protocols as well. The Z9500 in a hierarchical VLT fabric permits both Layer 2 and Layer 3 VLANs to be stretched across PODs. This allows traffic to remain inside the POD and never exit the POD, only allowing inter-port traffic exiting the POD. This provides the performance of switching and the flexibility of routing all within a single integrated architecture.

## Networking scalability for today and tomorrow

All three fabrics (micro scale, macro scale, and hyper scale fabrics) cover a broad spectrum of the fabric market through the "Pay-As-You-Grow" model, allowing customers to optimize for 10GbE or 40GbE without having any reach limitations inside the fabric. It provides the modern IT organization an optimal method for financial flexibility and for right sizing the network for superior capex management. Through a rich set of technologies, the Z9500 also brings flexibility to a wide range of use cases for successful deployments.