THE LIPPIS REPORT

The Resource for Network/IT Business Decision Makers

CISCO NEXUS® 9000 SERIES

Lippis Report Research Note
October 2013

© Lippis Enterprises, Inc. 2013

Cisco's Nexus 9000 Re-defines Software-Defined Networking

It was back in February of 2013, during the Open Networking User Group, or ONUG, hosted by Fidelity Investments in Boston, that one of its board members told me "We could wake up in the morning and Cisco will have an open networking solution that changes the industry." Well, November 6th 2013 was that morning. Cisco acquired Insieme Networks and with it, addressed networking's biggest complaints that have been voiced far and wide; that is, data center networking is too over-subscribed, ridged and not flexible to support on-demand workload creation and movement. These and other complaints are perhaps best articulated in an October 2010 blog by James Hamilton, VP and Distinguished Engineer on the Amazon Web Services Team, titled Data Center Networks Are In My Way. Since then Insieme's engineering team has redefined networking, as we know it, manifested in a product portfolio that will not only change networking but the IT industry. There are multiple value propositions embedded in the new Cisco product line, which I'll cover here in the Lippis Report over the next few quarters. But for this Lippis Report Research Note, we review the new Nexus 9000 series of data center switches, which Cisco promises is the most port dense and power efficient plus fastest packet forwarder and programmable data center modular switch in the industry. The Nexus 9000 series represents a familiar starting point on the journey toward a new era in software-defined networking.

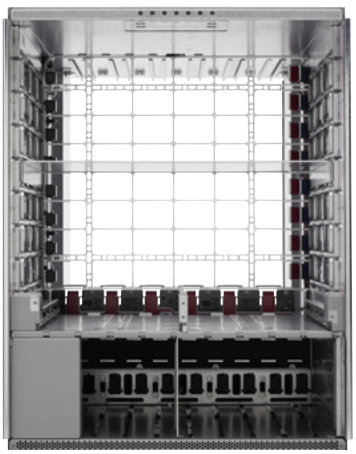
Cisco recently announced its new Nexus 9000 family of data center switches, which comprises the Nexus 9300 series fixed switches and the Nexus 9500 series modular switches. Of particular interest is the Nexus 9508, which is impressive in terms of performance, power efficiency, 10/40GbE and future 100GbE port density, programming environment and orchestration attributes. But while most companies announce products long before first customer ship with long road maps of when product features are available, Cisco's Nexus 9000 series is ready in Q4 of this calendar year. In fact, the independent and open industry Lippis/Ixia team have recently completed the process of testing the Nexus 9508 at 288 40GbE capacity for layer 3 unicast plus IP multicast forwarding latency and congestion management via RFC 2544, 3918 and 2889, respectively. The Nexus 9508 test report is in production and will be available here shortly. In the meantime, we review the Nexus 9000 product architecture, and how it will change data center network design.

Nexus 9500 High 10/40GbE Density with Wire Speed Non-Blocking Performance

The first modular switch of the family—the Nexus 9508—boasts impressive density and speed metrics. It supports an industry-leading 288 non-blocking 40GbE ports in its eight-slot chassis format. Each one of the eight slots within the chassis is populated with a 36-40GbE line card, equating to 288 40GbE. In addition, the Nexus 9508 supports 1,152 non-blocking 10GbE ports with 40GbE-to-10GbE breakout cables. More on the various modules below, but for now the Nexus 9000 delivers the same port density as other competitors, such as the Arista 7500E.

Chassis Architecture

There are many unique attributes of the Nexus 9500 series, such as its chassis design. The Nexus 9508 is an eight-slot chassis, meaning it supports eight line card slots—an



© Lippis Enterprises, Inc. 2013

important distinction as these slots are used solely for line cards versus a mix of line cards and supervisor modules. Two supervisor modules are located underneath the line card modules, offering full 1-for-1 redundancy. Under the supervisor modules are eight power supply slots capable of supplying some 3kWatts of AC power each to the Nexus 9500; however only two are needed to power a fully populated chassis. An additional two power supplies can be used for redundancy in a 2-plus-1 or 2-plus-2 configuration. The remaining four slots provide power headroom for later migration to support 100GbE and another generation of ASIC. If history is any guide, then the chassis will be in operation for well over a decade supporting multiple generations of ASIC.

One of the key design attributes is that the Nexus 9500 has no mid-plane for line card-to-fabric module connectivity, enabling unobstructed front-to-back airflow, thus contributing to its power and cooling efficiency; more on this below. There are three fan trays, two system controllers and six fabric modules accessible from the back of the chassis. The bottom line is that the Nexus 9500 Chassis was designed for reliability and scale well into the future.

Integrated Merchant and Custom Silicon Line Cards

A first in network switch design is the use of both merchant and custom silicon to compliment and add innovation upon, which is what Cisco (previously Insieme Networks) engineering achieved for line cards. This mix of merchant and custom silicon enables price point advantage plus a means to engineer added value beyond that of switches based solely on merchant silicon. The proof is in the following three line cards and the many more to follow in the quarters ahead.

There will be three line cards available for the Nexus 9500: 1) 48 1/10GbE SFP+ with 4-40GbE QSFP+, 2) 48 1/10GbT with 4-40GbE QSFP+ and 3) 36-40GbE QSFP+ full line rate. A fully populated Nexus 9500 provides over 30 Tbps of switching capacity.

The 1/10GbE line cards provide 640 Gbps of line rate capacity. Note that the 4 40GbE can be configured as both uplinks and downlinks.

The 40GbE line card is based on QSFP+ form factor. Cisco claims that the Nexus 9000 offers non-blocking performance at all packet sizes from 64 to 9216 bytes, and this claim has been validated in a separate independent test report. The 36 40GbE line cards are a non-over-subscribed design, eliminating small packet size wire speed forwarding limitation others may face. From a network design perspective, the above line card options offer a configuration of 384 1/10GbE plus 32 40GbE ports for mixed 1/10/40GbE environments without using QSFP+ breakout cables. This configuration is very helpful for 10GbE-40GbE migrations. Alternatively, network engineers may wish to use QSFP+ breakout cables on the 40GbE line cards, providing additional flexibility and 10GbE density.

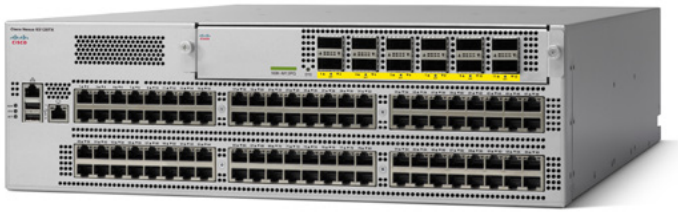
Power Efficiency

The Nexus 9500's innovative hardware architecture is designed for extremely high efficiency in keeping with today's market requirement of green computing. As mentioned earlier, the Nexus 9500 does not have a mid-plane for connecting line cards to fabric modules, allowing for completely unobstructed front-to-back airflow. The line cards contain only two to four ASICs with no buffer bloat and the aforementioned-mix of 28nm Cisco and 40nm Broadcom custom and merchant ASICs, respectively. In terms of power, the Nexus 9000's power supplies are platinum rated and are said to exhibit 90% to 94% power efficiency across all workloads. On a power per port basis, Cisco claims that the Nexus 9500 consumes typically 3.5W per 10GbE port and 14W per 40GbE port, which would make the Nexus 9500 the most power efficiency modular switch in the data center market. We calculate that it would only cost $5,475.26 annually to power a 288 40GbE fully loaded Nexus 9500.

New Top-of-Rack or Fixed Configuration Nexus Switches

While the above focused on the Nexus 9508, the Cisco/Insieme team is also introducing two new Top-of-Rack (ToR) or leaf switches that are fixed configuration switches. There's the Nexus 9396PX, which is a 48 1/10GbE SPF+ and 12 40GbE QSFP+ switch in a 2RU form factor, as well as the Nexus 93128TX, which is a 96 1/10GbE-T and eight 40GbE QSFP+ switch in a 3RU form factor. What's common across both fixed configuration switches is support for redundant FANs and power suppliers plus Front-to-Back and Back-to-Front airflow as well as dual or quad core CPU with 64GB SSD storage as the default.

© Lippis Enterprises, Inc. 2013



These fixed configuration switches are unique on multiple levels, but at their core is the support for both merchant and custom ASIC. The Broadcom Trident II chip provides the best in merchant silicon forwarding functionality while the Insieme custom ASIC provides all the key value-added functionality, such as additional buffering for mixed speed connectivity, VXLAN routing, utilizing the 40GbE ports for server connectivity or uplinks, etc. By designing the 9300 series with both merchant and custom ASIC, Cisco is not tied to the production and innovation cycle of merchant silicon, but is free to innovate and add differentiated features to its fixed configuration switches.

Programmability and the New Nexus OS

In addition to the compelling promises of the Nexus 9000's hardware, Cisco engineers designed an industry standard software platform that defines programmable networking. Cisco modernized the Nexus Operating System, or NX-OS, and updated the underlying Linux OS to a full 64-bit kernel and has opened access to the Linux shell via direct BASH access. The new software platform is comprised of three pillars: Programmability, Automation and Orchestration, plus Visibility, all in one NX-OS image.

The Nexus 9000 exposes a tremendous amount of network data that is accessible by a wide range of programmatic options, positioning the Nexus 9000 as the most feature rich software-defined networking product in the industry. For example, the Broadcom shell, ASIC counters, buffers, TCAM, CLI output, Bash, etc., are directly accessible, not only by the NX-OS console, but also through REST, RPC, NETCONF, XML and JSON APIs. Insieme has also added a new REST/RPC interface called NX-API, allowing network or DevOps administrators to provision or query the switch through structured HTTP/HTTPS REST calls. The Nexus 9000's forwarding tables may also be programmed through the Cisco onePK development environment or through OpenFlow protocols, allowing controllers such as Cisco's eXtensible Network Controller (XNC) or the Open Daylight Controller (ODL) to directly program the switch.

What's interesting about this programming environment is that Cisco supports not only the traditional man-machine metaphor in CLI, but machine-machine metaphor in XML, JSON, REST APIs, etc., so that applications can call upon network resources, and network engineers can automate redundant management tasks. Effectively, Cisco has torn down the boundaries of creating a fully programmable network that puts this power in the hands of DevOps and data center engineers. More on this in future Lippis Reports, but the programming environment that Insieme has developed promises to change not only networking but to alter data center economics and unleash a wave of new applications that leverage network value.

In addition to programmability, the Nexus 9000 promises to deliver integrated Automation and Orchestration tools, such as Puppet and OpsCode Chef agents for image/patch and configuration management, a fully integrated XMPP Pub-Sub infrastructure for management of a large number of devices, as well as native integration with OpenStack through the Cisco Neutron plugin.

The Cisco Nexus 9000 series also promises to increase overall visibility of the network by including vTracker, dynamic buffer monitoring, enhanced consistency checkers, an enhanced version of Wireshark, SMTP Email "pipe" output and Embedded Event Manager. These technologies increase device and overall network visibility of data center traffic.

Last but not least, the Nexus 9000 platforms all run on the same software binary image across both the fixed form-factor 9300 series as well as the modular 9500 series, so customers are able to deploy a complete data center network leveraging a single qualified software image.

© Lippis Enterprises, Inc. 2013

Investment Protection

The Nexus 9000 was designed from the ground up to integrate with pre-existing data center network infrastructures, preserving past investments while providing forward migration. For example, data center cabling represents a large capex spend; therefore, preserving that investment and extracting greater bandwidth is a winning design attribute. Also forward and backward network operating system capability is highly desired from a feature and skill set preservation perspective.

Many will use the Nexus 9000 platforms in both traditional three-tier access/aggregation/core and emerging two-tier leaf/spine architectures. Many Cisco customers will want to deploy the Nexus 9000 in data centers where other Nexus equipment is deployed to leverage that previous investment. For example, according to Cisco, the Nexus 9000 will support the Nexus 2000 series fabric extenders (FEX). Nexus 9500 modules that support both 1/10 base T and F are planned to preserve past copper and fiber-cabling infrastructure spend. The Nexus 9500 series of modular switches is a platform with shared supervisors, power supplies and modules that will be shared across the rest of the Nexus 9500 product family.

From a Nexus OS perspective, the Nexus 9000 will support a version of NX-OS 6.1(2) release, providing feature consistency with the existing Nexus portfolio of switches, but adding additional programmatic capabilities as well as much-requested features such as software patchability. This will provide a common CLI management look and feel as well as compatibility with existing SNMP MIBs, NETCONF-XML APIs, etc. Most Nexus features will carry forward, including routing protocols, management, high availability features, In-Service Software Upgrade (ISSU), Virtual Port Channels (VPC), etc., according to Cisco.

Integrated VXLAN Gateway, Bridging and Routing

VXLAN, or Virtual eXtensible LAN, is one of the most important protocols in the support of virtualized networking or overlays. The Nexus 9300 and 9500 series of data center switches will support VXLAN encapsulation and de-encapsulation in hardware. A VXLAN gateway provides VXLAN-VLAN gateway functionality; VXLAN bridging provides VXLAN-VXLAN forwarding, while VXLAN routing provides the ability to route VXLAN traffic between different IP subnets.

VXLAN gateway, bridging and routing functionality is built into the Nexus 9000 series, offering a wide range of support to manage VXLAN traffic. For the VXLAN gateway service, the Nexus 9000 series supports VXLAN Tunnel End Point, or VTEP, functionality to terminate a VXLAN-encapsulated traffic as well as initiate a VXLAN tunnel, by encapsulating the Layer 2 frame within a VXLAN header that includes the addition of a VXLAN Network Identifier, or VNID. This enables the expansion of the L2 segment space from 4,096 with VLANs today to well over 16 million segments with a 24-bit segment ID. The VXLAN gateway functionality supported across both the Nexus 9300 and 9500 enables great flexibility in the extension of VXLANs and mapping into existing VLANs on a per-switch basis, providing network engineers a degree of freedom in how virtualized network tunnels are managed and used.

VXLAN Bridging provides L2 forwarding between VNIDs natively within Nexus 9300 and 9500 switches. In addition to VXLAN Bridging, the Nexus 9000 series also adds support for integrated routing, allowing VXLAN VNIs and IEEE 802.1Q VLANs to not only bridge between each other, but also support the ability to L3 route traffic between all these different L2 segment technologies without sacrificing front-panel user ports.

So how should a network engineer think about network design with the Nexus 9000 series of switches? The integrated hardware VXLAN capability opens up completely new design possibilities, where customers can begin to look at deploying end-to-end scalable L3 fabric networks, but add VXLAN at the network edges to allow applications and workloads to flexibly communicate to each other by leveraging VXLAN hardware overlays directly on Nexus 9000-enabled switches and not sacrifice forwarding performance.

Additionally, while the Nexus 9000 will be at the center of emerging leaf/spine network designs, there are also some practical use cases that address the need for more speed and flexibility as cloud providers and enterprise IT transition from 1GbE to 10GbE and from 10GbE to 40GbE.

© Lippis Enterprises, Inc. 2013

Transitioning from 1GbE to 10GbE in the Data Center

The Nexus 9000 will offer an upgrade incentive for those who deployed the Catalyst 6500 at 1GbE in their data centers at end of row configurations. The Nexus 9000 offers better 1/10GbE port density, 40G uplinks, wire speed performance, competitive price points and no need to re-cable for either 1 or 10GbE. This will challenge the wisdom of keeping the Catalyst 6500 in the data center. In short, Catalyst 6500 data center customers will be offered the opportunity to upgrade depreciated equipment to a 1/10G access switch without the need to re-wire, which is a significant cost plus obtain all the Nexus 9000 benefits discussed above. Given the significant use of Catalyst 6500 in the past for server connectivity, this provides an easy migration option without a change in network architecture. Note that there is no backward compatibility with the Catalyst 6500, meaning Sup720 or 2T plus modules with the Nexus product lines.

For those data centers where the Catalyst 6500 provides 200 to 384 1GbE ports of copper connectivity, the Nexus 9000 can replace the Catalyst 6500 and use the same cable infrastructure and gain literally 1-for-1 1/10GbE copper ports. Therefore, Catalyst 6500 customers maintain the same cable infrastructure and upgrade to 10GbE at their pace, meaning that either an all 10GbE or a mix of 1GbE and 10GbE can be accommodated.

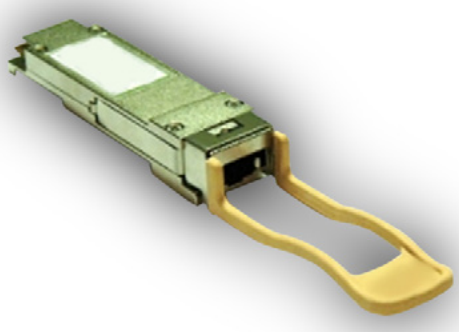
10GbE to 40GbE: Removing the 40GbE Three-Point Barrier of Entry

There are three costs that create a barrier of entry for network designers to deploy 40GbE data center infrastructure. 1) Fiber cabling upgrade requirement, which most IT executives dismiss 40GbE as soon as they wrap their mind around this cost, 2) cost of 40GbE optics and 3) 40GbE switch pricing.

Pluggable QSFP Platform to Remove Fiber Upgrade Barrier of Entry

Rather than a PHY-less design, the Insieme engineers decided to develop a QSFP+ pluggable for 40GbE optics to avoid the fiber optic cable upgrade requirement. 40GbE optics is a significant portion of capital expense, thanks to the cost of optics plus new cabling requirements. Today's 40GbE optics requires four times the amount of fiber cabling vs. 10GbE. That is, eight strands of fiber are required for 40GbE vs. two for 10GbE. The typical 40GbE MPO optical connector has 12 fiber connectors, but only eight are used (four for transmit and four for receive). Therefore, upgrading from 10GbE to 40GbE forces a cable infrastructure upgrade, which usually becomes a barrier to entry, thanks to this cable infrastructure upgrade cost.

To remove this barrier, the Insieme engineers developed a QSFP platform that supports any Cisco QSFP port and is called the Cisco 40G SR-BiDi QSFP. The SR-BiDi QSFP multiplexes two 10GbE signals into one 20GbE stream and runs two 20GbE wavelengths on the optics side, and delivers a QSFP pluggable MSA compliant electric signal to the switch module, thereby only requiring the termination of a dual LC connector as used in 10GbE optical infrastructure. That is, the SR-BiDi QSFP runs 40GbE over the same two strands of fiber that are used in 10GbE optical infrastructure. The SR-BiDi QSFP enables the re-use of existing 10GbE multi-mode fiber cable infrastructure plus patch cables as it supports the same LC connector. The SR-BiDi QSFP eliminates the cable infrastructure upgrade requirement of today's 40GbE, which can lower capex of cabling and switch hardware by as much as 70%. In short, the SR-BiDi QSFP provides a zero-cost fiber cabling upgrade path for 10GbE to 40GbE.



The second barrier of entry is 40GbE optics cost, which Cisco plans to price to enable 40GbE adoption. This pricing strategy will remove both the fiber cabling and optics barrier of 40GbE entry. What's going to happen in the market is that network designers will move from 10GbE to 40GbE without having a cabling, optics and switch cost barrier. This is a flash point, triggering massive 40GbE upgrades.

© Lippis Enterprises, Inc. 2013

While the above provides a low cost and smooth transition from 10GbE to 40GbE, the Nexus 9500 with 288 40GbE ports provides this transition at scale and high density for those data centers and cloud infrastructure that require the same 40GbE density as their 10GbE infrastructure. The Nexus 9500 offers two very powerful transition options; they are: transitioning from 1-to-10 GbE in the end of row space, and second, moving from 10GbE to 40GbE in the aggregation space. Cisco, with the acquisition of Insieme, is offering a way to arbitrage data center physical media connectivity by upgrading at lower costs and higher performance.

But the Nexus 9000 series is not just about L1 connectivity and L2/L3 forwarding. These upgrade options are table stakes to one of the most fundamental transitions occurring in computer networking, and that's a programmable software-defined network. That is, the Nexus 9000 offers a L4-L7 narrative with its programming environment engineered into the series of these switches. In short, Cisco is offering a practical way to transition to higher speed data center networking through favorable economics, which also happens to include a new programmable network platform ready for the age of software-defined networking. James Hamilton may want to revisit his October 2010 blog.

© Lippis Enterprises, Inc. 2013

About Nick Lippis

Nicholas J. Lippis III is a world-renowned authority on advanced IP networks, communications and their benefits to business objectives. He is the publisher of the Lippis Report, a resource for network and IT business decision makers to which over 35,000 executive IT business leaders subscribe. Its Lippis Report podcasts have been downloaded over 200,000 times; ITunes reports that listeners also download the *Wall Street Journal's* Money Matters, *Business Week's* Climbing the Ladder, *The Economist* and The *Harvard Business Review's* IdeaCast. He is also the co-founder and conference chair of the Open Networking User Group, which sponsors a bi-annual meeting of over 200 IT business leaders of large enterprises. Mr. Lippis is currently working with clients to design their private and public virtualized data center cloud computing network architectures with open networking technologies to reap maximum business value and outcome.

He has advised numerous Global 2000 firms on network architecture, design, implementation, vendor selection and budgeting, with clients including Barclays Bank, Eastman Kodak Company, Federal Deposit Insurance Corporation (FDIC), Hughes Aerospace, Liberty Mutual, Schering-Plough, Camp Dresser McKee, the state of Alaska, Microsoft, Kaiser Permanente, Sprint, Worldcom, Cisco Systems, Hewlett Packet, IBM, Avaya and many others. He works exclusively with CIOs and their direct reports. Mr. Lippis possesses a unique perspective of market forces and trends occurring within the computer networking industry derived from his experience with both supply- and demand-side clients.

Mr. Lippis received the prestigious Boston University College of Engineering Alumni award for advancing the profession. He has been named one of the top 40 most powerful and influential people in the networking industry by *Network World. TechTarget*, an industry on-line publication, has named him a network design guru while *Network Computing Magazine* has called him a star IT guru.

Mr. Lippis founded Strategic Networks Consulting, Inc., a well-respected and influential computer networking industry-consulting concern, which was purchased by Softbank/Ziff-Davis in 1996. He is a frequent keynote speaker at industry events and is widely quoted in the business and industry press. He serves on the Dean of Boston University's College of Engineering Board of Advisors as well as many start-up venture firms' advisory boards. He delivered the commencement speech to Boston University College of Engineering graduates in 2007. Mr. Lippis received his Bachelor of Science in Electrical Engineering and his Master of Science in Systems Engineering from Boston University. His Masters' thesis work included selected technical courses and advisors from Massachusetts Institute of Technology on optical communications and computing.

© Lippis Enterprises, Inc. 2013