Deploying Data Center Switching Solutions

Choose the Best Fit for Your Use Case
Table of Contents

Executive Summary ......................................................................................................................................................... 3
Introduction ........................................................................................................................................................................ 3
Multivector Scaling .......................................................................................................................................................... 3
Low On-Chip Memory ASIC Platforms ....................................................................................................................... 4
  Features and Benefits of Broadcom Tomahawk-Based Platforms ........................................................................ 4
  Use Cases ..................................................................................................................................................................... 5
Limitations of Low-Memory ASIC Platforms ............................................................................................................. 5
Large External Memory ASIC Platforms .................................................................................................................... 5
  Features and Benefits of Q5-Based Platforms ........................................................................................................... 6
  Use Cases ..................................................................................................................................................................... 6
Comparison Matrix ........................................................................................................................................................ 7
Conclusion ...................................................................................................................................................................... 7
About Juniper Networks .................................................................................................................................................. 8

List of Figures

Figure 1: Multivector scaling ........................................................................................................................................... 4
Figure 2: Broadcom Tomahawk ASIC .......................................................................................................................... 4
Figure 3: Multispeed server attach in a ToR leaf switch .................................................................................................. 5
Figure 4: Juniper Q5 Silicon ............................................................................................................................................. 5
Figure 5: QFX10002 external memory architecture .................................................................................................... 6
Figure 6: QFX10000 switches in L3 fabric, Junos Fusion, and MC-LAG deployments .................................................... 7
Executive Summary
The impact of cloud computing and software-defined networking (SDN), including network virtualization, has convinced large enterprises and customers building private, public, and hybrid cloud networks to consider new approaches to help them meet their current and future data center switching needs. Data center switching platforms are constantly evolving, delivering increasingly better performance while simplifying network operations.

Two types of silicon are typically used in these data center build outs: those with memory “on-chip” (low-memory ASICs), and those with memory supported “off chip” (large, external memory ASICs). While the low-memory ASICs support high port densities at a relatively low cost and power usage, improving total cost of operation, the silicon size is fixed, restricting the ability to add memory and requiring a trade-off of features for functionality. Therefore, users planning a new data center network need to plan carefully, making sure that limited logical scaling and buffering capacity are sufficient for the application requirements.

As switching needs evolve from basic high-speed port connectivity to more complex requirements involving multidimensional scale and performance, it becomes difficult to meet them with products based on low-memory ASICs. Such use cases are best served with silicon based on an external memory architecture that allows port scaling and supports feature richness, as well as logical scaling of the forwarding information base (FIB), quality of service (QoS), and security.

This white paper will inform readers about the Juniper Networks® QFX5200 switching platforms using Broadcom Tomahawk silicon, as well as the Juniper Networks QFX10000 switches, which use Juniper’s purpose-built Q5 custom silicon. When it comes to building high-performance data centers, Juniper encourages users designing their own switching infrastructure to carefully weigh their options and choose the ASICs that best meet their specific needs, whether third-party or Juniper custom-built silicon. This white paper will also help readers understand how to position Juniper’s data center switches for feature-rich implementations.

Introduction
Consolidation, centralization, and the virtualization of compute resources are resulting in fewer, larger data centers. At the same time, increasing compute density using multicore, multisocket servers, combined with virtualization and storage convergence, is reducing the amount of physical space these consolidated data centers need. Workloads that used to take multiple racks of servers are now being delivered within a small portion of a single rack.

As a result, many data center and enterprise campus deployments today are moving towards high-density10GbE server access at the edge with a 40GbE backbone. Hyperscale data centers are looking beyond 10GbE to new standards such as 25GbE server attach with 100GbE aggregation/core backbone switches, allowing them to take advantage of server economies of scale and a 2.5X performance improvement.

The physical network topology for most enterprise data centers is a two-tier (or more) spine-leaf architecture delivered as Layer 3 or IP fabric topology, or, in some cases, as an Ethernet fabric topology. Most IP fabric-based data centers have either already deployed, or are looking to adopt, overlay technologies to achieve workload mobility.

Multivector Scaling
While developing a data center switching platform, several factors must be considered to effectively meet customer demands, including multivector scaling (Figure 1).

One particularly vexing problem (and one that can scale in multiple dimensions) is the fact that all parameters are interdependent, so changing one vector could significantly impact others; for example, adding features may cause higher latency. Or increasing buffer scale requires more memory, resulting in higher power consumption. Increasing logical scale might also impact multiple other factors like latency and power consumption. Lastly, adding tunneling features could result in lower system throughput.

Providing the best possible combination of these vectors is a delicate balance, and it is difficult to achieve with most current third-party on-chip memory silicon—especially for higher layer data center deployments in the spine, edge, and core.
Switching platforms based on low on-chip memory ASICs are best suited for cost-effective, high-speed, high-density server access deployments. These platforms are ideally suited for environments where applications do not rely on the network to absorb any data buffering when network congestion occurs. The latest low-memory ASIC from Broadcom, called Tomahawk, targets cloud builders and hyperscale Web services users who want server ports and switches based on 25 Gbps lanes; these switches consume less power per bit than a 10 Gbps lane, and the incremental cost for this higher bandwidth is just 1.5X or less than the 10 Gbps speed—not 2.5X, as the speed would suggest.

Features and Benefits of Broadcom Tomahawk-Based Platforms

Some of the key features and benefits for Broadcom Tomahawk-based platforms include:

- 3.2 Tbps switching capacity with a single chip
- Multiport speeds: 10GbE, 25GbE, 40GbE, 50GbE, and 100GbE
- Configurable lower pipeline latency with sub 400 ns port-to-port operation

Platforms based on the Tomahawk ASIC can use existing cabling to link to compute and storage elements as well as across fabrics. Cloud providers are moving towards compute nodes with 25GbE attach within server racks, reducing power requirements while increasing bandwidth. Switching vendors have announced products using the latest Broadcom StrataXGS Tomahawk silicon. The new Broadcom chip supports:

- 32 ports of 100GbE
- 64 ports of 40GbE/50GbE
- 128 ports of 10GbE/25GbE on a single chip
- 64 ports of 50GbE

Use Cases

Tomahawk-based switches are mainly targeted at ultra high-scale deployments such as top-of-rack (ToR) leaf switch deployments for 25GbE/50GbE server access, with 100GbE in the aggregation/core. These platforms allow large cloud-scale customers to upgrade their network to 100GbE in a cost-effective fashion.
These switches are also well positioned for latency-sensitive deployments.

Limitations of Low-Memory ASIC Platforms

The basic set of switching requirements is growing over time. Customers are looking to scale data center fabrics for overlay support and other additional capabilities across the entire industry. Third-party silicon-based products are intended for fast forwarding and routing of typical standards-based protocols. Aspects like L2/L3 protocol support, stacking, port density, and performance will be uniform across many vendor products; any differentiation will come from the switch OS and other software components. Although low-memory ASICs are widely used in certain cloud deployments, there are a number of limitations that prevent them from meeting the demands of large enterprise customers:

- Limited amount of on-chip memory does not help in improving application performance. Applications must be designed to respond to network congestion events when Broadcom Tomahawk-based switches are used in the network.
- Limited amount of logical scale is inadequate, especially for dual stack (IPv4 and IPv6) deployments.
- There is a lack of Data Center Interconnect (DCI) features like Layer 3 Virtual Extensible LAN (VXLAN) gateway functionalities and protocol stitching.
- There are no virtual output queue (VOQ) capabilities that avoid head-of-line blocking and absorb traffic spikes.

Large External Memory ASIC Platforms

Juniper has purpose-built the Q5 ASIC to meet customer demands for multidimensional performance. The Q5 eliminates performance trade-offs across physical, logical, and overlay networks, facilitating simple, consolidated network designs and allowing customers to seamlessly evolve data center networks from existing 10GbE and 40GbE architectures to 100GbE to keep pace with evolving needs.

The Juniper Networks QFX10000 product line leverages the Q5 ASIC to deliver greater system throughput and meet the rapid and ongoing traffic growth in the data center. The Q5 ASIC delivers a unique blend of high I/O capacity (500 Gbps bidirectional), flexible pipeline, and high logical scale. This is achieved by connecting a new and innovative class of memory called HMC—a new 3D memory that provides unprecedented memory-to-ASIC bandwidth of 1 Tbps. It allows the QFX10000 to achieve high I/O, high scale, deep buffers, and high VOQ memory for different types of tables such as FIB, host, MPLS, and media access control (MAC). The Q5 ASIC is 400GbE-ready and is embedded with an on-chip analytics capability, along with a precision timing protocol and high-frequency monitoring features.
Features and Benefits of Q5-Based Platforms
Third party-based silicon switches are commonly referred to as “system on a chip” or SOC systems; all variants of the QFX10000 switches are multichip or “non-SOC” systems. The big difference between SOC and non-SOC systems is that SOC systems have no external memory—all packet memory and tables are on the chip itself.

Some of the key features and benefits of a Q5-based platform such as the QFX10000 include:

- Enhanced cell-based fabric architecture to achieve optimal load distribution
- VOQ and high-speed delay bandwidth buffer
- Extremely flexible packet processing capabilities to process multiple headers without sacrificing performance
- Enhanced cloud analytics and telemetry functions using precision timing synchronization
- Architectural flexibility (L3 spine-leaf, L3 Clos, etc.) that can be configured on a per-port basis
- Enhanced adaptive load balancing that improves efficiency and link utilization
- Deep buffers enabled by high-performance custom silicon to ensure quality of service
- Future-proof design, which enables seamless transition from 10GbE to 40GbE, 100GbE, and eventually 400GbE

Use Cases
The QFX10000 switches can be deployed in a number of different networks, including:

- Layer 3 fabrics
- Juniper Networks Junos® Fusion architectures, a simple, easy-to-deploy fabric that scales to support mid-to-large data center deployments while simplifying network management and configuration
- Juniper’s multichassis link aggregation group (MC-LAG) for L2 and L3 networks, giving customers complete architectural flexibility

These open architectures ensure that customers can innovate on top of Juniper Networks Junos operating system to accelerate the pace of innovation. Support for a wide range of MPLS, DCI, and SDN gateway technologies also makes the QFX10000 line a perfect fit for spine/edge deployments.
Comparison Matrix
The following table compares key characteristics of low memory and large external memory-based switching solutions from Juniper.

**Table 1. Low On-Chip Memory vs. Large External Memory Comparison**

<table>
<thead>
<tr>
<th></th>
<th>Low On-Chip Memory ASIC: Broadcom Tomahawk-Based Switches</th>
<th>Large External Memory ASIC: Juniper Q5-Based QFX10000 Switches</th>
</tr>
</thead>
<tbody>
<tr>
<td>Place in network</td>
<td>High speed/low buffer leaf</td>
<td>High speed/high buffer leaf and spine</td>
</tr>
<tr>
<td>Delay bandwidth buffer</td>
<td>Few microseconds</td>
<td>Up to 80 milliseconds</td>
</tr>
<tr>
<td>Virtual output queuing</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Cell-based architecture</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>L2 VXLAN overlay</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>L3 VXLAN overlay</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Port speeds</td>
<td>10GbE, 25GbE, 40GbE, 50GbE, 100GbE</td>
<td>10GbE, 40GbE, 100GbE (400GbE ready)</td>
</tr>
<tr>
<td>Latency</td>
<td>~500 ns</td>
<td>2.5-5.0 microseconds</td>
</tr>
<tr>
<td>IPv4 FIB</td>
<td>128,000</td>
<td>256,000</td>
</tr>
<tr>
<td>IPv6 FIB</td>
<td>64,000</td>
<td>256,000</td>
</tr>
<tr>
<td>Host routes</td>
<td>104,000</td>
<td>Over 2 million</td>
</tr>
<tr>
<td>IPv4 multicast routes</td>
<td>104,000</td>
<td>Up to 128,000</td>
</tr>
<tr>
<td>IPv6 multicast routes</td>
<td>52,000</td>
<td>Up to 128,000</td>
</tr>
<tr>
<td>MAC addresses</td>
<td>13,600</td>
<td>Up to 512,000</td>
</tr>
</tbody>
</table>

**Conclusion**
The networking industry is experiencing tremendous technological advances, particularly in SDN, NFV, and network virtualization. It is important that users of these technologies carefully consider their options and select those that are optimized for their particular deployment scenarios. For server access switching needs in the data center, cost and high density are generally the overarching criteria; for these cases, third-party silicon-based switching solutions will fit the bill. In comparison, Juniper’s silicon-based systems, such as the Q5-based QFX10000, provide multidimensional scale and performance with future-proofed investment protection for data center spine/edge and DCI deployments.

**About Juniper Networks**
Juniper Networks is in the business of network innovation. From devices to data centers, from consumers to cloud providers, Juniper Networks delivers the software, silicon and systems that transform the experience and economics of networking. The company serves customers and partners worldwide. Additional information can be found at [www.juniper.net](http://www.juniper.net).