Embracing 400G and 800G

Laying the Foundation for Next Generation Networks with 7050X4 and 7060X5

Introduction

As next generation applications and distributed data workloads continue to grow, high speed compute nodes are becoming increasingly mainstream. These emerging applications, and the compute and storage on which they are built, demand faster access to more data; and drive the need for larger scale data center networks with higher bandwidths to maximize scalability and performance, while minimizing job completion times.

Until now, general compute has only required 10G or 25G connectivity, while limited numbers of high end applications tended to require up to 50G or 100G. Today, standard compute nodes are themselves becoming capable of 50G or 100G data rates, while high end compute and storage systems are migrating towards higher speeds such as 200G and 400G.

As data centers evolve, the growth of high speed hosts at the network edge inevitably dictates the strategic direction of the overall data center network, driving a need for more fabric bandwidth, with 400G capable spine switching at present, and 800G in the future.

For existing facilities, operating on longer refresh cycles, it is imperative to support the coexistence of both existing 1G/10G/25G connections based on PCIe Gen3 and the emerging requirement for 50G, 100G and 200G, driven by nodes equipped with PCIe Gen4 and future PCIe Gen5 network adapters. It is also important to take advantage of the consolidation made possible by higher bandwidth densities to collapse multi-tier networks for cost, performance and efficiency.

For cutting edge, greenfield deployments, native 200G/400G/800G solutions enable high radix, high performance designs with increased bandwidth and port density at the lowest power and cost per Gbps.

Arista's 7050X4 and 7060X5 family of fixed systems, together with the 7358X4 and 7388X5 compact modular platforms, provide compelling options to build out these next generation data centers, offering multi-generational compatibility, along with high density 50G, 100G, 200G, 400G and 800G solutions with scalable resources, low latency and the rich feature set of Arista EOS. The X4 and X5 families maximize investment protection while presenting opportunities for significant cost savings and efficiency improvements for organizations planning upgrade cycles or greenfield deployments.

In this paper we will review some of the key use cases of high speed compute networks, made universally accessible through Arista's broad portfolio of platforms that support multiple generations of technology, each designed to fit different workload types and scale requirements.

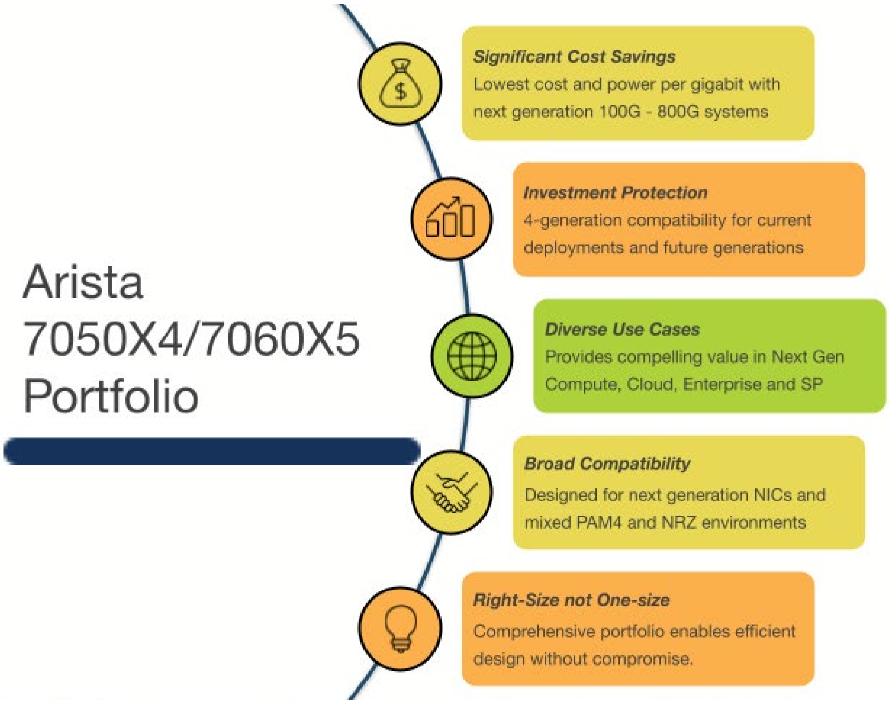

Arista 7050X4/7060X5 Portfolio
Significant Cost Savings
Lowest cost and power per gigabit with next generation 100G - 800G systems
Investment Protection
4-generation compatibility for current deployments and future generations
Diverse Use Cases
Provides compelling value in Next Gen Compute, Cloud, Enterprise and SP
Broad Compatibility
Designed for next generation NICs and mixed PAM4 and NRZ environments
Right-Size not One-size
Comprehensive portfolio enables efficient design without compromise.

Figure 1. Key advantages of the Arista 7050X4 and 7060X5 product portfolio

High Speed Connectivity for Compute and Storage

The advent of the next generation of compute nodes is one of the primary drivers of the evolution of data center networks. Most existing networks provide 10G/25G connectivity to the hosts, along with a number of 100G uplinks into the data center leaf-spine network.

Existing 10G and 25G interfaces based on SFP form factor each use a single signaling lane running at either 10G or 25G, offering the lowest cost and highest density solution, which is critical when deploying many thousands of servers. By contrast, current implementations of 40G and 100G, combine four parallel underlying signals together to create a higher speed interface. The added complexity naturally increases the physical size, cost and power consumption of transceivers and cables, and the switches and NICs they connect to, limiting broad adoption.

With the advent of higher speed system buses built on PCIe Gen4 and Gen5 technologies, supporting 56G PAM4; the 50G, 100G and 200G network interfaces are expected to grow to dominate compute connectivity within the next few years. Supporting large numbers of high speed host network connections economically clearly requires a step change in the underlying technology to become viable.

The solution comes by way of increasing the single-lane signaling rate to 50Gbps using technology inherited from 400GbE. Individual 50G lanes are backwards compatible to 10G/25G but may also be combined to create 100G, 200G or 400G interfaces, offering a common architecture to address brownfield migration and greenfield deployment needs.

The Arista 7050X4 portfolio, comprising the 7050X4 and 7358X4 series, is designed to bridge the gap between the 1G/10G/25G and 50G/100G/200G/400G generations perfectly with the introduction of the SFP-DD and DSFP form factors providing investment protection for high volume connectivity with 10G/25G/50G/100G quad-speed support, as well as dense 200G/400G systems for high end applications and upstream connections to match emerging 200G and 400G data center fabrics. The new platforms add advanced features such as MACsec encryption and quad-speed support, while reducing power consumption, lowering the cost per Gbps, and retaining the common 1RU and 2 RU form factors for optimal investment protection, four-generation compatibility and forklift-free network evolution.

Network adapters supporting speeds of 50G, 100G and 200G based on 56 Gbps PAM4 technology are available from all large NIC vendors, providing options for 50G, 100G and 200G connections in the ubiquitous SFP and QSFP56 form factors, offering the best economics, power consumption and connector density for volume applications.

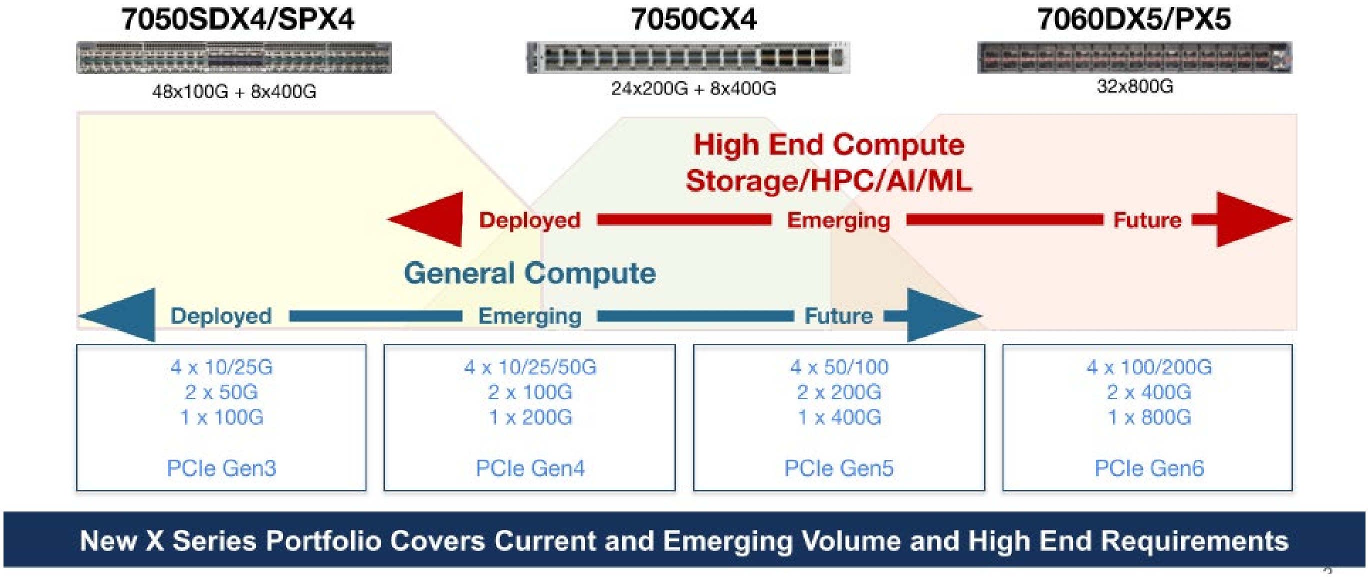


Figure 2. Multi-generation support for deterministic future proofing of networks

Common Deployments - Choices, choices, and more choices

Providing choice has always been a key part of Arista's philosophy and the 7050X4 and 7060X5 family of products epitomize choice through a wide range of form factors, providing solutions for common connectivity requirements. Supporting flexible speeds ranging from 10G to 800G, and multiple common interface types, customers get full control and flexibility in tuning the design and implementation of their networks for their specific environments. In this section, we will take a closer look at some of the typical use cases that each product facilitates.

General enterprise compute networks currently deploy 1G/10G/25G connections to the server estate with 100G upstream connections to the data center spine layer. As PCIe Gen4 and Gen5 based compute is deployed, these needs will evolve to 10G/25G/50G/100G connections for hosts and 400G upstream links to the data center fabric. The Arista 7050SDX4-48D8 and the 7050SPX4-48D8 offer flexible SFP and QSFP connectivity that support both the current and future scenarios, providing multi-generational co-existence and fork-lift free upgrades for both compute and data center fabric.

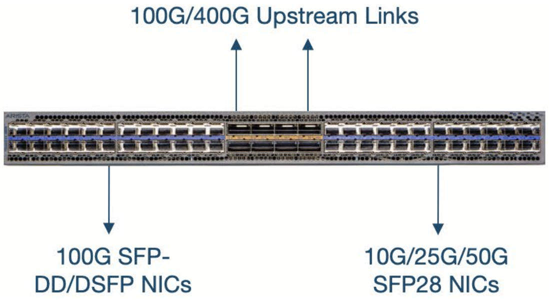


Figure 3. Host/network connection combinations on the 7050SDX4-48D8 (48 x 100G SFP-DD and 8 x 400G QSFP-DD)

As networks organically evolve, the need for devices that can fulfill multiple roles correspondingly increases. The 7050CX4M-48D8 is one example with its range of 8 x 400G upstream network side connections and 48 x 100G QSFP downstream connections, that can take on the role of a high end leaf switch for QSFP28 based 50G or 100G hosts, or be deployed as a consolidated spine for 100G connectivity down to the leaf switch and integrate into the 400G links of the super-spine data center fabrics. With the addition of integrated MACsec support on all the ports, strong encryption can also be applied to sensitive links, either to servers or to the network core.

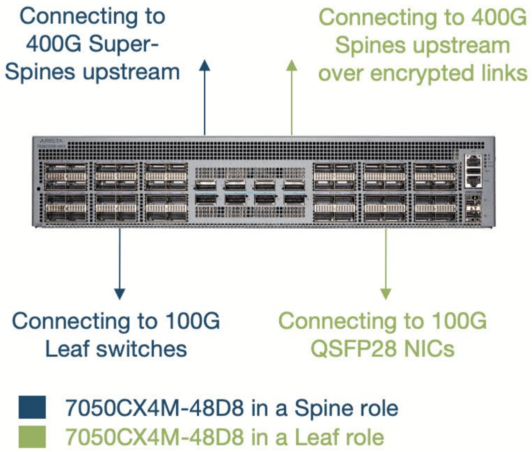

Connecting to 400G Super-Spines upstream
Connecting to 400G Spines upstream over encrypted links
Connecting to 100G Leaf switches
Connecting to 100G QSFP28 NICs
🟦 7050CX4M-48D8 in a Spine role
🟩 7050CX4M-48D8 in a Leaf role

Figure 4. Multiple options of deploying the 7050CX4M-48D8 (48 x 100G QSFP and 8 x 400G QSFP-DD)

In storage or cluster networks, hosts commonly connect using 40G/100G or bundles of 25G interfaces, while switches use 100G uplinks to the fabric. As 100G/200G based storage or processing nodes are deployed, the fabric uplinks must themselves evolve to 200G/400G. The 7050CX4-24D8 is purpose built for these use cases with it's dense native 200G QSFP56 support, providing flexibility of 50/100/200G interfaces commonly found on high bandwidth NICs for compute and storage.

A variety of passive DACs (Direct Attach Cable) can be used directly with the QSFP56 PAM4 based NICs for speeds from 10G-200G. Conversion between QSFP56 200G-4 (4 x 50G lanes) switch port and existing 2x QSFP28 100G-4 (4 x 25G lanes) NICs is achieved with Active Electrical Cables (AEC), which convert the PAM4 signal to NRZ. Arista switches support a wide range of cables and transceivers to cover all the common use cases.

With an architecture based on 32 x 100GbE 1RU spine switches, the maximum number of compute nodes that can be supported in a single cluster is 1536 servers across 32 leaf switches. Increasing scale beyond this number requires multiple clusters that are themselves interconnected via an additional spine layer, consuming some of the spine-leaf ports and resulting in a further level of oversubscription between nodes in adjacent clusters which results in non-deterministic edge-to-edge performance.

For example, the following diagram depicts deployment of a super-spine that allows 4 clusters to be interconnected at the cost of a further 3:1 oversubscription ratio between clusters. In this model, 8 x 100GbE ports from each cluster spine connect in pairs across the four Super-Spines resulting in a smaller number of nodes per cluster:

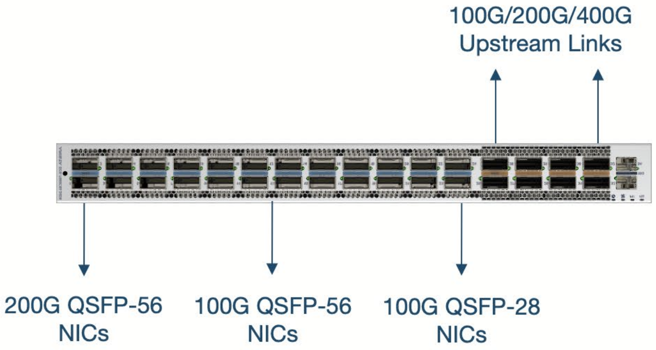


Figure 5. Host/network connection combinations on the 7050CX4-24D8 (24 x 200G QSFP-56 and 8 x 400G QSFP-DD)

The next use case applies to the most demanding applications including High Performance Compute (HPC) and Artificial Intelligence/Machine Learning (AI/ML). These next generation applications leverage ever increasing data sets and increasing numbers of clustered compute nodes communicating in fully meshed east-west traffic patterns, requiring high speed, low contention and consistent low latency networks to operate at peak efficiency.

The 7060DX5-64E, with high density 100G/200G/400G/800G support, is the platform of choice for both small clusters and large fabrics. Supporting either OSFP and QSFP-DD transceivers and cables, the 7060DX5-64E series provides up to 64 ports of 400G or 256 concurrent 100G interfaces using breakouts in a space and power efficient 1RU form factor.

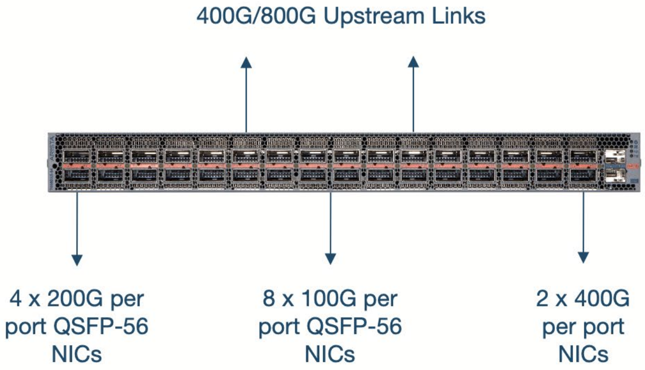

400G/800G Upstream Links
4 x 200G per port QSFP-56 NICs
8 x 100G per port QSFP-56 NICs
2 x 400G per port NICs

Figure 6: Host/network connection combinations on the 7060DX5-64E (32 x 800G QSFP-DD)

For smaller deployments, the 12.8Tbps based 7060DX5-32 offers half the port density with 32 x 400G ports supporting combinations of 400G for upstream links and 50G/100G/200G for downstream links via breakouts.

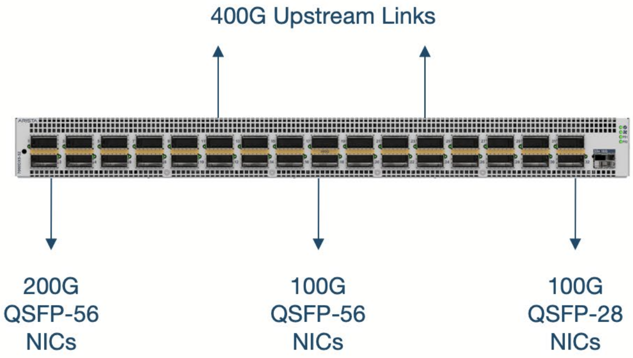


Figure 7: Host/network connection combinations on the 7060DX5-32 (32 x 400G QSFP-DD)

Connectivity Choices

The 7050X4 and 7060X5 support all widely deployed, and next generation, optics and cable for the datacenter. Newer form-factors & media types supported include:

- 100G SFP-DD and DSFPs: A new form-factor optimized for 50G PAM-4 based NIC connectivity. All Arista SFP-DD and DSFP ports also support 10/25/50G SFPs.
- 200G QSFPs (QSFP56) that enable simple 100G to 200G upgrades for leaf-to-spine links, as well as 4 x 50G or 2 x 100G PAM4 based NIC connectivity
- 800G OSFP & QSFP-DD for seamless upgrade from 400G to 800G and flexible breakout to 8 x 100G, 4 x 200G or 2 x 500G.

Table 1: Pluggable form factor evolution in the data center, supported by Arista 7050X4 and 7060X5 portfolio

Form Factor Family	Widely Deployed	New
SFP	10G SFP, 25G SFP	50G SFP (SFP56), 100G SFP-DD and DSFP
QSFP	40G QSFP, 100G QSFP	200G QSFP (QSFP56)
OSFP and QSFP-DD	400G OSFP and QSFP-DD	800G OSFP and QSFP-DD

Introduction to the 100G SFP-DD and DSFP form factor

The SFP-DD ("SFP Double-Density") and the DSFP ("Dual SFP") are two approaches (driven by different MSA groups) to achieving the same objective: A compact, SFP-compatible form factor with 2x 50Gb/s PAM-4 electrical lanes in each direction to enable a total bandwidth of 100G / port. The basic concept is shown below:

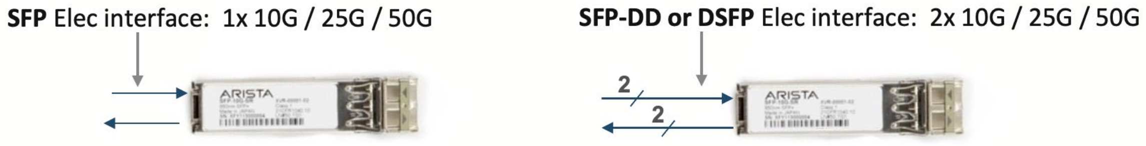

SFP Elec interface: 1x 10G / 25G / 50G
SFP-DD or DSFP Elec interface: 2x 10G / 25G / 50G
2
2

Figure 8: Comparison of SFP+/SFP28 (left) with SFP-DD/DSFP (right)

The SFP-DD specification adds a second row of contacts to the SFP electrical connector to enable the 2x 50G electrical interface. The DSFP standard approaches this slightly differently, repurposing a few of the low-speed pins of the original SFP standard to enable the 2x 50G electrical interface. The figure below summarizes the approach of the SFP-DD and the DSFP to achieve a dual-lane interface.

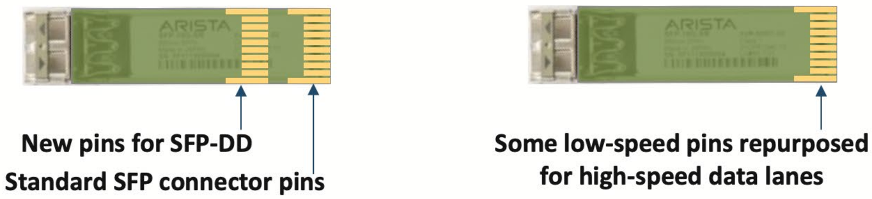

New pins for SFP-DD
Standard SFP connector pins
Some low-speed pins repurposed for high-speed data lanes

Figure 9: Comparison of SFP-DD (left) with DSFP (right)

Despite the different approaches to enabling 100G operation, both Arista's SFP-DD and DSFP systems offer quad-rate support: Each SFP-DD or DSFP port supports the use of 10G, 25G and 50G SFP optics and cables, as well as 100G SFP-DD / DSFP copper cables, enabling one system to support 4 generations of speeds and transceivers.

100G SFP-DD and DSFP copper cables for switch-to-NIC connectivity

Arista's SFP-DD and DSFP platforms and copper cables enable easy connectivity from Top-of-Rack (TOR) switches to:

- NICs with native 100G SFP-DD / DSFP ports
- NICs with QSFP56 ports at 100G-2 (PAM-4) or 50G-2 (NRZ) data rates, and
- NICs with 10G/25G/50G SFP ports.

The diagrams below illustrate TOR to NIC connectivity using SFP-DD / DSFP cables:

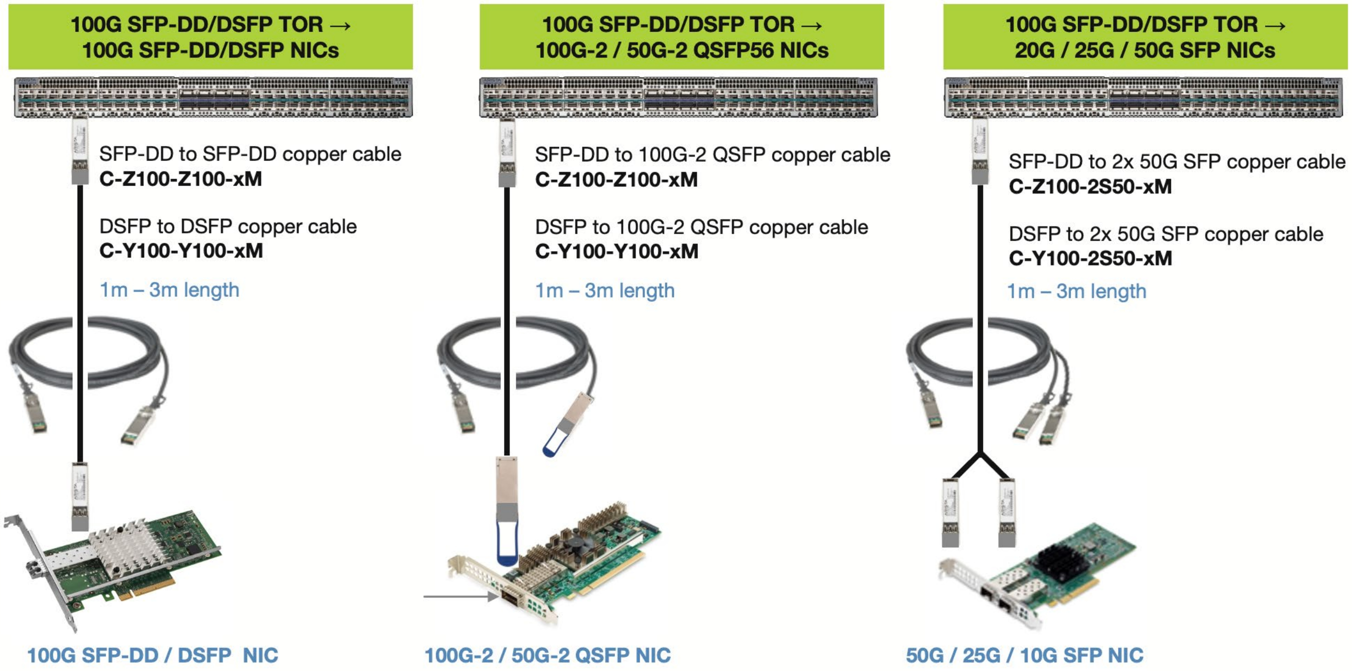


Figure 10: SFP-DD/DSFP cable selection for connecting to common network adapters

Introduction to 200G QSFPs (or QSFP56)

200G QSFPs (sometimes referred to as QSFP56), are compliant to the same physical form-factor as 100G QSFPs (QSFP28), but with the 4-lane electrical interface running at 50G PAM-4 per lane, enabling a total bandwidth of 200G. Arista's 200G connectivity solutions include optics for single mode fiber (QSFP-200G-FR4), multimode fiber (QSFP-200G-SR4), Active Optical Cables (AOCs) and copper Direct Attach Cables (DACs). All Arista 200G QSFP transceivers and cables are 100G/200G dual-rate capable.

200G QSFP optics and cables for Switch-to-NIC connectivity

Arista's supports a variety of 200G QSFP optics and cables that can be used to provide connectivity from 200G QSFP based Top-of-Rack (TOR) switches to:

- NICs with 200G QSFP56 or 100G QSFP28 ports
- NICs with 100G-2 QSFP56 or 50G-2 QSFP28 ports
- NICs with 10G / 25G / 50G SFP ports

Each of these connectivity options are illustrated below:

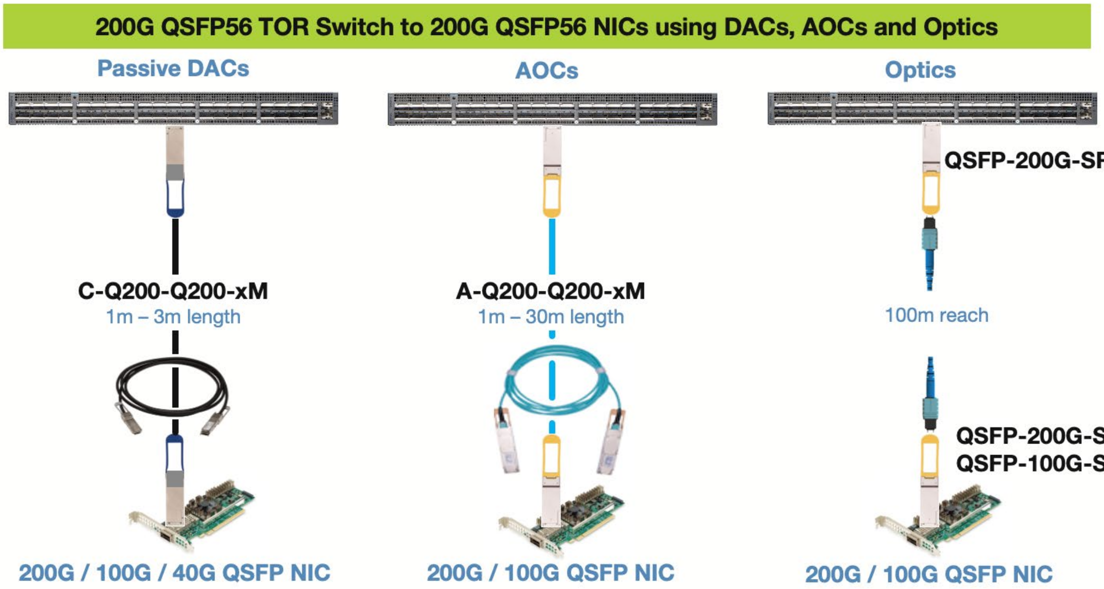


Figure 11: 200G QSFP cable selection for connecting to common network adapters

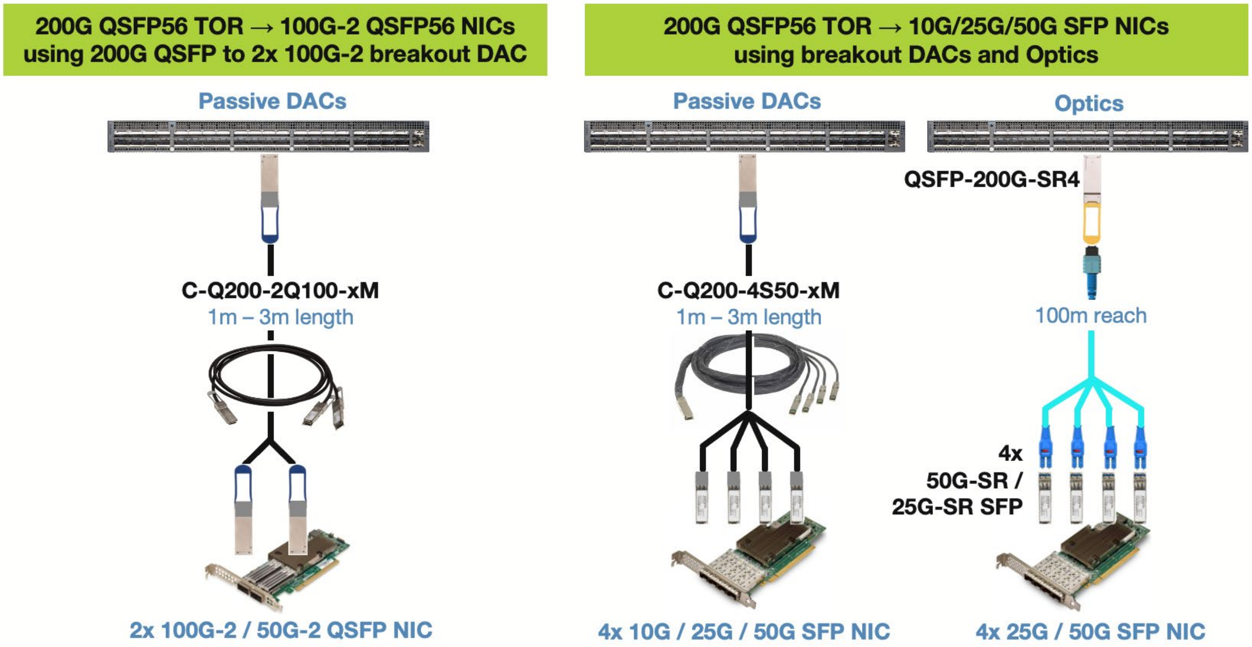

200G QSFP56 TOR → 100G-2 QSFP56 NICs
using 200G QSFP to 2x 100G-2 breakout DAC
200G QSFP56 TOR → 10G/25G/50G SFP NICs
using breakout DACs and Optics
Passive DACs
Passive DACs
Optics
QSFP-200G-SR4
C-Q200-2Q100-xM
1m – 3m length
C-Q200-4S50-xM
1m – 3m length
100m reach
4x
50G-SR /
25G-SR SFP
2x 100G-2 / 50G-2 QSFP NIC
4x 10G / 25G / 50G SFP NIC
4x 25G / 50G SFP NIC

Figure 12: 200G QSFP breakout cable selection for connecting to common network adapters

200G QSFP optics for leaf and spine connectivity

Arista offers the following 200G optical modules for leaf and spine connectivity:

- The **QSFP-200G-FR4** for 200G/100G links over 2km of duplex single mode fiber (SMF). When operated at 100G, the Arista QSFP-200G-FR4 optically interops with 100G-CWDM4 optics, simplifying 100G to 200G upgrades for duplex SMF links.

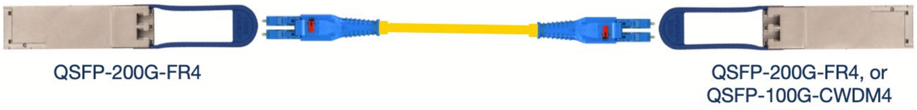


Figure 13: Overview of QSFP-200G-FR4

- The **QSFP-200G-SR4** for 200G/100G links over 100m of parallel multi-mode fiber (MMF). When operated at 100G, the Arista QSFP-200G-FR4 optically interops with 100G-SR4 optics, simplifying 100G to 200G upgrades for parallel MMF links.

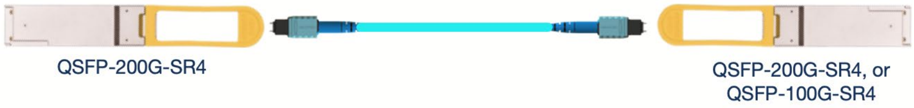


Figure 14: Overview of QSFP-200G-SR4

Introduction to 800G OSFP and QSFP-DD

Arista's 800G optical modules and cables are available in the same form-factors that are widely deployed in today's 400G networks, OSFP and QSFP-DD. 800G modules increase the speed of each lane from 50Gbps to 100Gbps, providing double the bandwidth at a lower power and cost per bit when compared to existing 400G modules.

Arista's 800G optical modules can be configured as two distinct 400G interfaces or broken out to eight 100G interfaces. The choice of OSFP or QSFP-DD form-factors, and the easy breakout to already deployed 400GbE and 100GbE optics, provides investment protection for existing 400GbE networks while enabling a seamless upgrade to 800G.

800G OSFP and QSFP-DD Connectivity Options

Arista's 800G connectivity solutions include:

- 800G-2XDR4 / PLR4 modules for 2x 400G links over 2km / 10km of parallel SMF. Arista's 800G-2XDR4 / PLR4 optics use 2x MPO-12 connectors enabling 2 distinct 400G links without the need for breakout cables. The image below shows an example of an OSFP-800G-2XDR4 connected to 2x OSFP-400G-XDR4 modules.

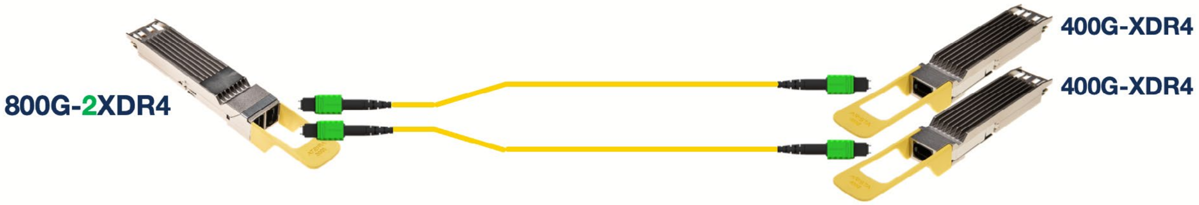


Figure 15: Overview of 800G-2XDR4 to 2 x 400G-XDR4

The 800G-2XDR4 / PLR4 modules also enable breakout to 8x 100G links, as shown below:

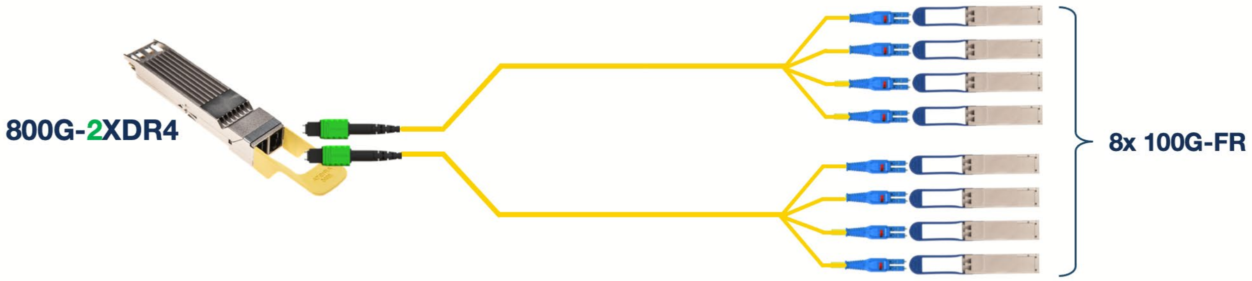


Figure 16: Overview of 800G-2XDR4 to 8 x 100G-FR

- 800G-2FR4 / LR4 for 2x 400G links over 2km / 10km of duplex SMF. Arista's 800G-2FR4 / LR4 optics use 2x duplex LC connectors enabling 2 distinct 400G links without the need for breakout cables. The image below shows an example of an OSFP-800G-2FR4 connected to 2x OSFP-400G-FR4 modules.

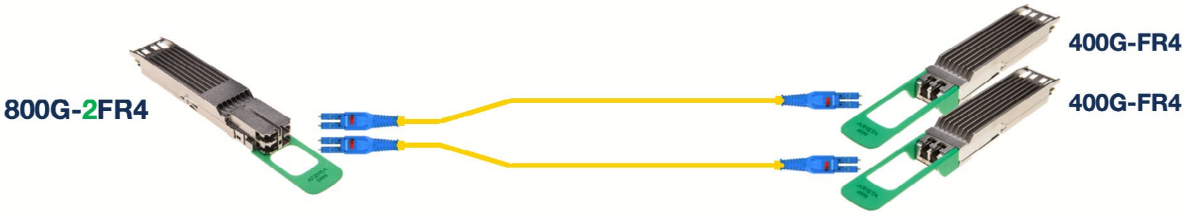


Figure 17: Overview of 800G-2FR4 to 2 x 400G-FR4

- 800G AOCs, for point to point links from 1m to 30m
- 800G DACs, for lengths up to 2m
- 800G Active Electrical Cables (AECs), for lengths up to 5m.

Putting all the pieces together

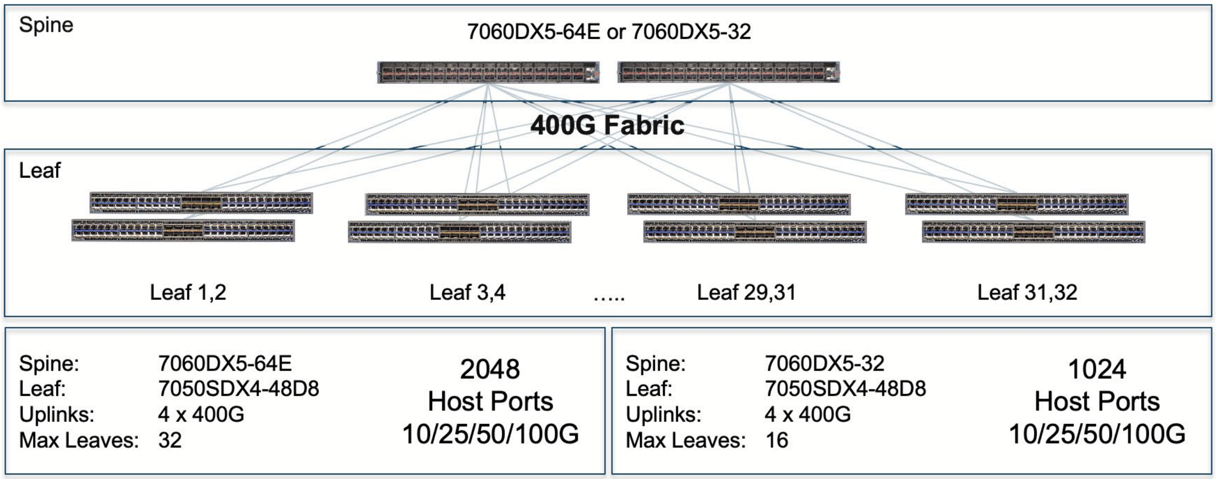
Having reviewed the sweet spot of each member of the family and the range of connectivity available, we can now put the building blocks together with some real-world examples.

The first example demonstrates a classic leaf-spine architecture network with a twist - the 7050SDX4-48D8 and 7050SPX4-48D8 provide the ability to support host connections from 10G up to 100G using the common SFP interface while also leaving plenty of options to scale up/down the contention ratio and number of supported leaf switches.

Two simple scenarios are shown with each leaf switch offering 48 SFP ports and 4 x 400G uplinks to a two-way spine. Each leaf switch also has four spare 400G ports which could be used for further host connections, increasing the leaf-spine bandwidth or adding more spine switches.

The leaves connect upstream to either 2 x 7060DX5-32 or 2 x 7060DX5-64E spine switches. With the number of spine ports available dictating the maximum number of leaf switches and therefore the maximum number of host connections - in this case 1024 or 2048 end hosts.

This very typical network topology offers many advantages including high density, deterministic performance, a simple and easily scalable configuration and active-active redundancy. It also offers considerable flexibility to increase or decrease contention ratios, the level of redundancy and the total size of the cluster over time without a complete fork-lift upgrade.



Spine: 7060DX5-64E
Leaf: 7050SDX4-48D8
Uplinks: 4 x 400G
Max Leaves: 32

2048
Host Ports
10/25/50/100G

Spine: 7060DX5-32
Leaf: 7050SDX4-48D8
Uplinks: 4 x 400G
Max Leaves: 16

1024
Host Ports
10/25/50/100G

Figure 18: Example of a classic leaf-spine network based on the 7050X4 and 7060X5 series

Next, we can go one step further and consider the example of a heterogeneous network that leverages the choices provided in the 7050X4 and 7060X5 portfolio to support multiple different applications within a single network, again in a classic Leaf-Spine architecture.

The goal for this architecture is to provide a single leaf-spine topology that supports existing and emerging applications efficiently while offering the same elastic scale and performance choices of the homogenous topology described in the previous example.

Leaf Switches

This network will continue to support an existing fleet of 10G/25G SFP and 100G QSFP28 based hosts while being able to onboard 50G/100G SFP and 100G/200G/400G QSFP based devices as required; the following switches are ideal choices for each role:

Figure 19: Arista 7050X4 and 7060X5 family for Leaf deployment

Table 2: Typical use cases and corresponding 7050X4 and 7060X5 leaf switches			
Use Case	Connectivity Required Today	Future Requirement	Device
General Compute	10/25G SFP	Support co-existence of 50/100G SFP	7050SPX4-48D8 or 7050SDX4-48D8
Existing Storage / HPC	40/100G QSFP28	Aggregate and integrate to new topology / Encrypt	7050CX4M-48D8
Emerging Storage / HPC	40/100G QSFP28	100G/200G QSFP56	7050CX4-24D8
High End AI/ML	50/100G QSFP28	100/200G QSFP56 and 400G	7060DX5-64E or 7060PX5-64E or 7060DX5-32 or 7050CX4-24D4 or

Spine Switches

While Arista's extensive portfolio offers choices for spine platforms ranging from the 7050X3 with 32 ports of 100G, up to the 7816R3 with 576 ports of 400G, inevitably, the choice of spine platform will depend on the number of leaf switches, the number of uplinks and bandwidth required as well as other performance, scale and elasticity goals.

The family of 64 x 400G and 32 x 400G 7060X5 systems provide a level of capacity that suits many enterprise requirements while providing flexible choices for connectivity, size and modularity. Support for dense 100G-400G is closely matched with the uplink connectivity provided by both the current and new generation leaf switches.

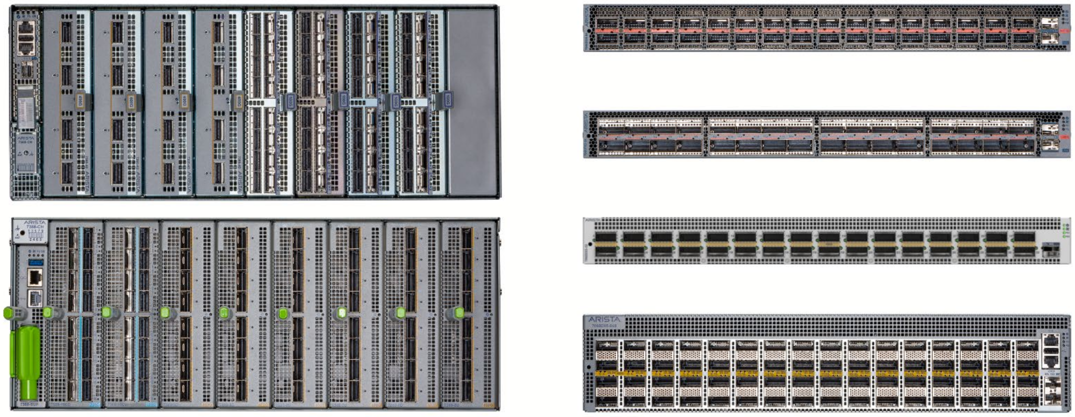


Figure 21: Arista candidates for an ideal Spine deployment - 7358X4, 7388X5, 7060X5 and 7050X4 series

Table 3: Comparison of different X4 and X5 series products for Spine deployment

Switch	Footprint	Max 100G	Max 200G	Max 400G
7060DX5-64E 7060PX5-64E	1RU OSFP800 or QSFP-DD800	256	128	64
7060DX5-64S	2RU QSFP-DD	256	128	64
7388X5	4RU Compact Modular	256	128	64
7358X4	4RU Compact Modular	128	64	32
7060DX5-32	1RU QSFP-DD	128	64	32
7050DX4-32S 7050PX4-32S	1RU QSFP-DD or OSFP	128	64	32

Solution

Putting the building blocks together, we can see that building a solution capable of supporting multiple generations of hosts and multiple types of applications is made easy by the availability of multiple form factors for both leaf and spine.

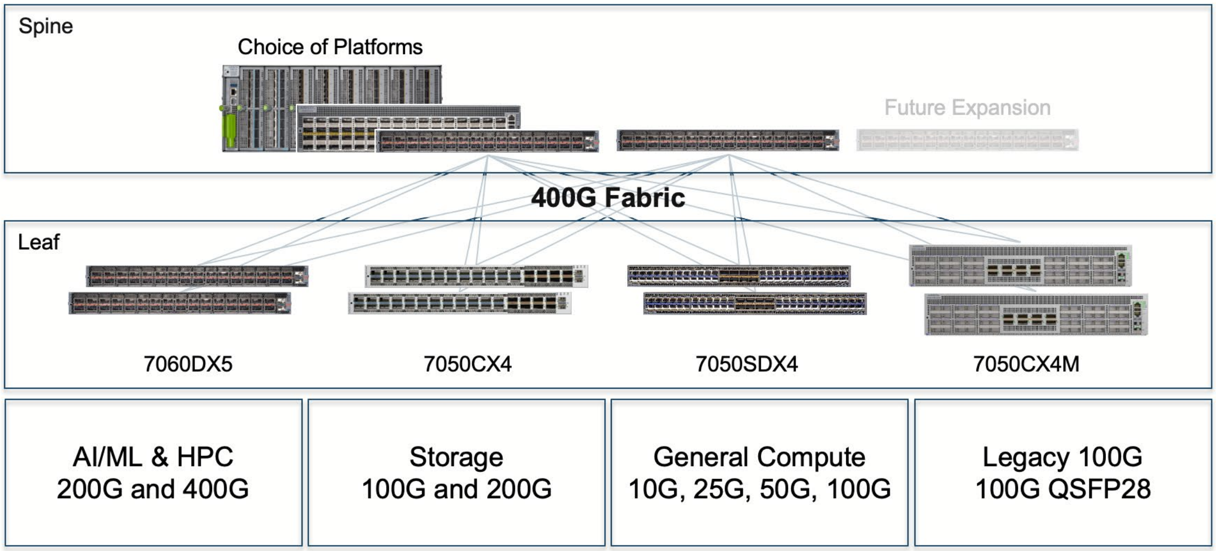


Figure 21: Blueprint for a leaf-spine architecture supporting heterogeneous current and future workloads

While the 7050X4 and 7060X5 families provide comprehensive choices which solve for many common use-cases, it is important to remember that with the unified EOS operating system and cross platform consistency, the choice of building blocks is not limited to X4 and X5 systems alone.

Any Arista platform that best fits the use case can be integrated as required. For example, the 7280R3 family, with deep buffers and full featured high scale routing, to handle demanding workloads or the 7800R3 at the spine to provide huge scalability - the choice is yours.

Summary

The emergence of PCIe Gen 4 and Gen 5 technologies in the server domain dramatically increases the bandwidth available to applications and lowers the cost of deploying high speed connectivity to the data center edge in volume.

Efficient data center networks need to be capable of onboarding new nodes without restricting performance while also retaining backwards compatibility, protecting existing investments and minimizing costs.

The Arista 7050X4 and 7060X5 families of products, along with a comprehensive portfolio of cables and transceivers, provide the broadest choice of connectivity options enabling efficient, high performance, right-sized designs that ensure existing compute fleets are not left behind while optimally supporting the needs of the next generation, maximizing value and investment protection for customers.

Santa Clara—Corporate Headquarters
5453 Great America Parkway,
Santa Clara, CA 95054

Phone: +1-408-547-5500
Fax: +1-408-538-8920
Email: info@arista.com

Ireland—International Headquarters
3130 Atlantic Avenue
Westpark Business Campus
Shannon, Co. Clare
Ireland

Vancouver—R&D Office
9200 Glenlyon Pkwy, Unit 300
Burnaby, British Columbia
Canada V5J 5J8

San Francisco—R&D and Sales Office
1390 Market Street, Suite 800
San Francisco, CA 94102

India—R&D Office
Global Tech Park, Tower A, 11th Floor
Marathahalli Outer Ring Road
Devarabeesanahalli Village, Varthur Hobli
Bangalore, India 560103

Singapore—APAC Administrative Office
9 Temasek Boulevard
#29-01, Suntec Tower Two
Singapore 038989

Nashua—R&D Office
10 Tara Boulevard
Nashua, NH 03062