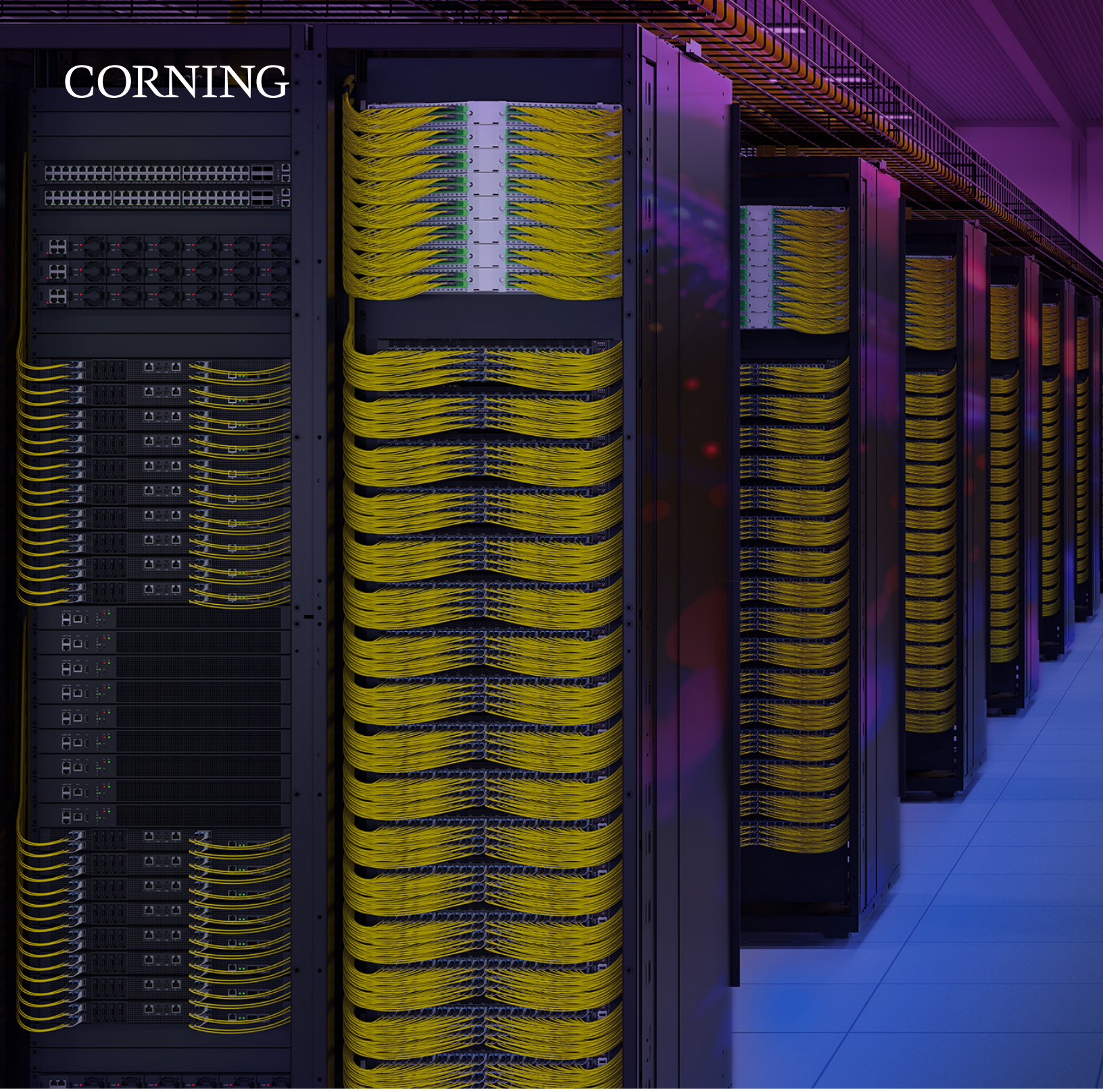


NVIDIA NVL72 GB200/GB300 Systems:

InfiniBand and Ethernet Connectivity Solutions.

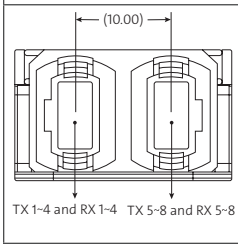
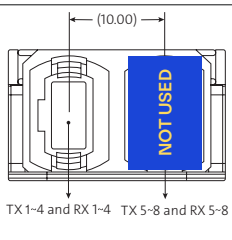
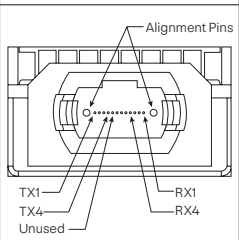
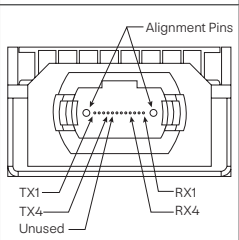
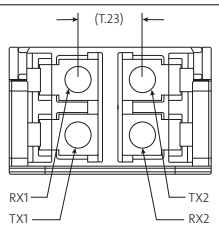
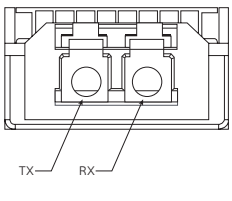
Table of Contents

1. Understanding the Transceiver Types, Port Breakouts and Cabling Scenarios

This cabling guide will explore the various fiber optic connectivity options available for working with 200G, 400G, 2x 400G, 800G, and 2x 800G transceivers. It will also cover diverse cabling solutions used within the same rack or row and across the data center, utilizing InfiniBand 400G NDR Quantum-2, InfiniBand 800G XDR Quantum-3, and Ethernet 400G Spectrum-4 switch capabilities.

With the introduction of 400G NDR and 800G XDR, NVIDIA optical transceiver modules feature a twin (dual) MPO-8/12 transceiver interface that utilizes 2x 8-fiber connections. Corning's EDGE8® solution is designed to support both Single-mode and Multimode optical interfaces based on the use of 2, 4, 8 and 16 fibers at the transceiver.

The following is a partial list of NVIDIA transceivers by connector type:

Twin MPO-8/12 APC Interface (OSFP)		MPO-8/12 APC Interface (OSFP)	MPO-8/12 APC Interface (QSFP112)	Twin LC Duplex UPC Interface (OSFP)	LC Duplex UPC Interface (QSFP-DD)
NDR (100G Optical Lanes)					
Single-mode [a]					
800G-2xDR4 OSFP [c] 2x 8-fiber transceiver MMS4X00-NM MMS4X00-NS MMS4X00-NS-FLT [d] MMS4X00-NS-T	400G-DR4 OSFP 8-fiber transceiver Uses 1 out of 2 ports MMS4X00-NS400 MMS4X00-NS400-T 200G-DR4 OSFP [e] 4/8-fiber transceiver Uses 1 out of 2 ports MMS4X00-NS400 MMS4X00-NS400-T		400G-DR4 QSFP112 8-fiber transceiver MMS1X00-NS400 MMS1X00-NS400-T 200G-DR4 QSFP112 [e] 4/8-fiber transceiver MMS1X00-NS400 MMS1X00-NS400-T	800G-2xFR4 OSFP 2x 2-fiber transceiver MMS4X50-NM MMS4X50-NM-T	400G-FR4 QSFP-DD 2-fiber transceiver MMS1V50-WM MMS1V50-WM-T
Multimode [b]					
800G-2xSR4 OSFP [c] 2x 8-fiber transceiver MMA4Z00-NS MMA4Z00-NS-FLT [d] MMA4Z00-NS-T	400G-SR4 OSFP 8-fiber transceiver Uses 1 out of 2 ports MMA4Z00-NS400 MMA4Z00-NS400-T 200G-SR4 OSFP [e] 4/8-fiber transceiver Uses 1 out of 2 ports MMA4Z00-NS400 MMA4Z00-NS400-T		400G-SR4 QSFP112 8-fiber transceiver MMA1Z00-NS400 MMA1Z00-NS400-T 200G-SR4 QSFP112 [e] 4/8-fiber transceiver MMA1Z00-NS400 MMA1Z00-NS400-T		
XDR (200G Optical Lanes)					
Single-mode [a]					
1600G-2xDR4 OSFP 2x 8-fiber transceiver MMS4A00		800G-DR4 OSFP 8-fiber transceiver MS4A20-XM800 400G-DR4 OSFP [g] 4/8-fiber transceiver MS4A20-XM800			
					

a) MPO-8/12 APC Single-mode optics are denoted by a yellow-colored pull tab and yellow-colored optical fiber. Green plastic shell on the MPO-8/12 APC connector denotes Angled Polish Connector and is not compatible with Ultra-flat Polished Connectors (UPC) used with slower line rate transceivers.

b) Multimode optics are denoted by a tan-colored pull tab and aqua-colored optical fiber. Green plastic shell on the MPO-8/12 APC connector denotes Angled Polish Connector and is not compatible with aqua colored shell for Ultra-flat Polished Connectors (UPC) for HDR.

c) Please note that in some documentation, 800G transceivers may appear as 800G-DR8 instead of 800G-2x DR4 and 800G-SR8 instead of 800G-2x SR4. However, the connectivity footprint is represented by 2x 8-fiber MPO-8/12 APC. Please follow NVIDIA's part number as the main reference.

d) The card I/Os is routed internally to four 800G Twin-port OSFP.

e) A 400G transceiver version will be able to support 200G utilizing a splitter cable (Y-Harness), activating two of the four lanes (4 out of 8-fibers) in the 400G transceiver creating a 200G device. This configuration is represented as "4/8-fiber" in the table above.

f) Transceiver part numbers ending with "-T" refer to Ethernet versions.

g) The Twin port OSFP, using a splitter cable (Y-Harness) that activates 4 out of 8 fibers within a single MPO connector, creates four 2x 200G-PAM4 links, achieving 400G (XDR400).

For more information on NVIDIA components and design, please review the Annex 3 with the references to NVIDIA Overview Whitepapers.

Table 1. Transceiver List.

1.1. Scenario 1 – 1600G, 800G and 400G – Server to Switch or Switch to Switch Applications | MPO-8/12 APC to MPO-8/12 APC using Point-to-Point Cabling

Application: InfiniBand Quantum-2 / InfiniBand Quantum-3 / Ethernet Spectrum-4 to InfiniBand Quantum-2 / InfiniBand Quantum-3 / Ethernet Spectrum-4; b) ConnectX-8 / ConnectX-7 / Bluefield-3; c) Cedar-7 links

Scenario 1 is primarily used for Point-to-Point cabling applications. This type of cabling can be used to connect a server to a switch or to connect different switches, such as Leaf to Spine or Spine to Core. However, it is not recommended to use Point-to-Point cabling if those switches are physically located in different areas within the data center.

Please refer to section 1.7., Table 8, to review the list of transceivers applicable for Use cases A to C in Scenarios 1 and 2. To learn more about Scalable Units and how to build a GPU Cluster utilizing NVL72 Systems, please refer to section 2 in this document.

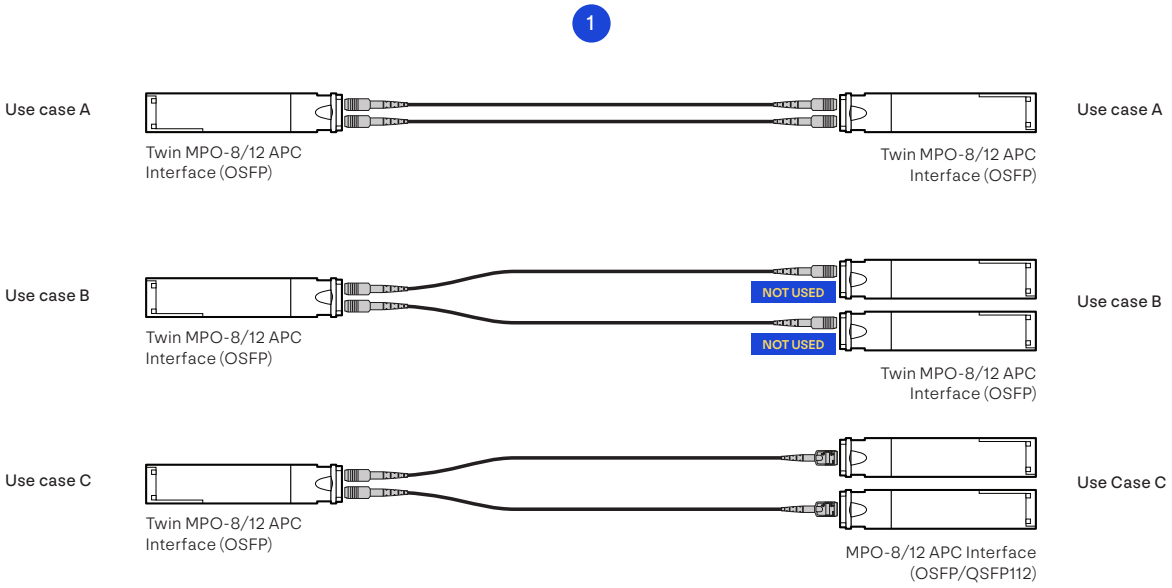


Figure 1. Use cases for Point-to-Point cabling utilizing 1600G, 800G and 400G transceivers working with MPO-8/12 APC.

Item	Reference Part Number	OS2 Part Number (Americas)	OS2 Part Number (EMEA and APJ)	OM4 Part Number (Americas)	OM4 Part Number (EMEA and APJ)	Description
1	Corning Component [a]	JE8E808GE8-NBxxxF	JE8E808GEZ-NBxxxM	JE9E908QE8-NBxxxF	JE9E908QEZ-NBxxxM	EDGE8®, 8F Jumper, MTP® APC (non-pinned) to MTP APC (non-pinned), Type-B polarity, xxxF (feet) or xxxM (meters)
	or					
	NVIDIA Component [b]	MFP7E30-Nxxx		MFP7E10-Nxxx		NVIDIA MPO-12/APC-to-MPO12/APC (8 fibers), Straight Crossover Fibers, Fiber Cable Product, Type-B polarity, xxx indicate length in meters
Instead of traditional jumpers, the following alternative is available for Item 1						
1	Corning Component [c]	GEDEDE4GPN-033-UxxxF	GEDEDE4GEZ-033-UxxxM	GEFEFE4QPN-033-UxxxF	GEFEFE4QEZ-033-UxxxM	18x 8F CORE-Trunk, MPO APC (non-pinned) to MPO APC (non-pinned), 86-in (2.2-m) legs, Type-B polarity, xxxF (feet) or xxxM (meters)

Notes:

a) Corning cables in the Americas use Plenum cable, while EMEA/APJ uses LSZH™/CPR rated cable. Single jumper lengths are available from 1 to 300 meters. Lengths in meters are also available for the Americas.

b) NVIDIA cables utilize a dual rated OFNR/LSZH™ jacket. SMF cable lengths available in 3, 5, 7, 10, 15, 20, 30, 50, 100 and 150 meters. MMF cable lengths available in 3, 5, 7, 10, 15, 20, 25, 30, 35, 40 and 50 meters.

c) CORE-Trunks lengths are available from 5 to 300 meters (furcation-to-furcation). CORE-Trunks are available in 16, 18, 32, 36 legs, and other custom leg quantities with straight or staggered legs. Lengths in meters are also available for the Americas.

d) All Corning and NVIDIA cables support InfiniBand, Ethernet and NVLink protocols.

e) Please review Corning's polarity drawings in Annex 2.

Table 2. Scenario 1 – 1600G, 800G and 400G using Point-to-Point Cabling with MPO-8/12 APC. Part Number Scheme.

1.2. Scenario 2 – 1600G, 800G and 400G – Switch to Switch Applications MPO-8/12 APC to MPO-8/12 APC using Structured Cabling Across DC with Trunk

Application: InfiniBand Quantum-2 / InfiniBand Quantum-3 / Ethernet Spectrum-4 to InfiniBand Quantum-2 / InfiniBand Quantum-3 / Ethernet Spectrum-4; b) ConnectX-8 / ConnectX-7 / Bluefield-3; c) Cedar-7 links

In Scenario 2, the implementation of Structured Cabling utilizes an optical fiber trunk as a backbone. This application is mostly used to connect different switches, such as Leaf to Spine, and can also be implemented to connect Spine to Core. It is the recommended option to follow when two different active devices are physically located in different areas within the data center.

Please refer to section 1.7., Table 8, to review the list of transceivers applicable for Use cases A to C in Scenarios 1 and 2. To learn more about Scalable Units and how to build a GPU Cluster utilizing NVL72 Systems, please refer to section 2 in this document.

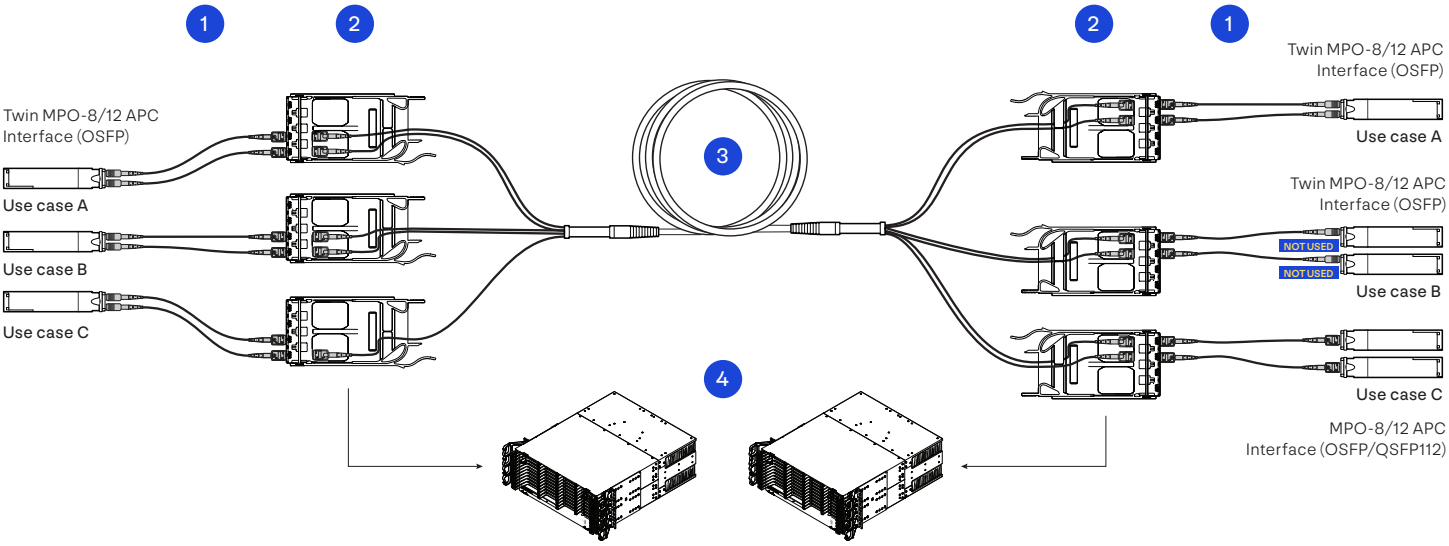


Figure 2. Use cases for Structured Cabling utilizing 1600G, 800G and 400G transceivers working with MPO-8/12 APC.

Item	Reference Part Number	OS2 Part Number (Americas)	OS2 Part Number (EMEA and APJ)	OM4 Part Number (Americas)	OM4 Part Number (EMEA and APJ)	Description
1	Corning Component[a]	JE8E808GE8-NBxxxF[b]	JE8E808GEZ-NBxxxM[b]	JE9E908QE8-NBxxxF[b]	JE9E908QEZ-NBxxxM[b]	EDGE8®, MTP® APC (non-pinned) to MTP APC (non-pinned) 8F Jumper, TIA-568 Type-B polarity, xxxF (feet) or xxxM (meters)
2		EDGE8-CP32-V1	EDGE8-CP32-V1	EDGE8-CP32-V3	EDGE8-CP32-V3	EDGE8 32F MTP Adapter Panel, (4-port)
3		GE7E7E4GPNDDU xxxF[c]	GE7E7E4GLZDDU xxxM[c]	GE2E2E4QPNDDU xxxF[c]	GE2E2E4QLZDDU xxxM[c]	EDGE8, MTP Trunk , 144F, MTP APC (pinned) to MTP APC (pinned), 33-in (840-mm) legs, Type-B polarity, pulling grip on first end only, xxxF (feet) or xxxM (meters)
4		EDGE8-xxU	EDGE8-xxU	EDGE8-xxU	EDGE8-xxU	Please refer to Annex 1 to choose the best option for your application
Instead of traditional jumpers, the following alternatives are available for Item 1						
1	Corning Component[a]	G-BND18-E8E8G-PN000-xxxF[d]	–	G-BND18-E9E9T-PN000-xxxF[d]	–	18x 8F Bundled Jumper, MPO APC (non-pinned) to MPO APC (non-pinned), 86-in (2.2-m) legs, Type-B polarity, xxxF (feet) or xxxM (meters) Grip on one end
		–	GEDEDE4GEZ-033-UxxxM[e]	–	GEFEFE4QEZ-033-UxxxM[e]	18x 8F CORE-Trunk, MPO APC (non-pinned) to MPO APC (non-pinned), 86-in (2.2-m) legs, Type-B polarity, xxxF (feet) or xxxM (meters)

Notes:
a) Corning cables in the Americas use Plenum cable, while EMEA/APJ uses LSZH™/CPR rated cable.
b) Single jumper lengths are available from 1 to 300 meters. Lengths in meters are also available for the Americas.
c) EDGE8® Trunks are available in fiber counts of 8 to 288 fibers.
d) Bundled Jumpers lengths are available from 5 to 30 meters (furcation-to-furcation) and are available only in OS2 and OM3 fiber. Bundled Jumpers use a meshed sleeve. Bundled Jumpers are available in 16, 18, 20, 32, 36 legs, and other custom leg quantities with straight or staggered legs. Lengths in meters are also available for the Americas.
e) CORE-Trunks lengths are available from 5 to 300 meters (furcation-to-furcation). CORE-Trunks are available in 16, 18, 32, 36 legs, and other custom leg quantities with straight or staggered legs.
f) All Corning cables support InfiniBand, Ethernet and NVLink protocols.
g) Please review Corning's polarity drawings in Annex 2.

Table 3. Scenario 2 – 1600G, 800G and 400G using Structured Cabling with MPO-8/12 APC. Part Number Scheme.

1.3. Scenario 3 – 1600G, 800G, 400G and 200G - Server to Switch Applications MPO-8/12 APC to MPO-8/12 APC using Point-to-Point Cabling

Application: InfiniBand Quantum-2 / InfiniBand Quantum-3 / Ethernet Spectrum-4 to InfiniBand Quantum-2 / InfiniBand Quantum-3 / Ethernet Spectrum-4; b) ConnectX-8 / ConnectX-7 / Bluefield-3

Scenario 3 is used for Point-to-Point cabling applications, where a 400G transceiver version will be able to support 200G by utilizing a splitter cable (Y-Harness) that activates two of the four lanes (4 out of 8 fibers) in the 400G transceiver, effectively creating a 200G device. Point-to-Point cabling is recommended only if the active devices are within the same rack or row. However, if the active devices are physically located in different areas within the data center, then Structured Cabling is the recommended option.

Please refer to section 1.7., Table 9, to review the list of transceivers applicable for Use cases A and B in Scenarios 3 and 4. To learn more about Scalable Units and how to build a GPU Cluster utilizing NVL72 Systems, please refer to section 2 in this document.

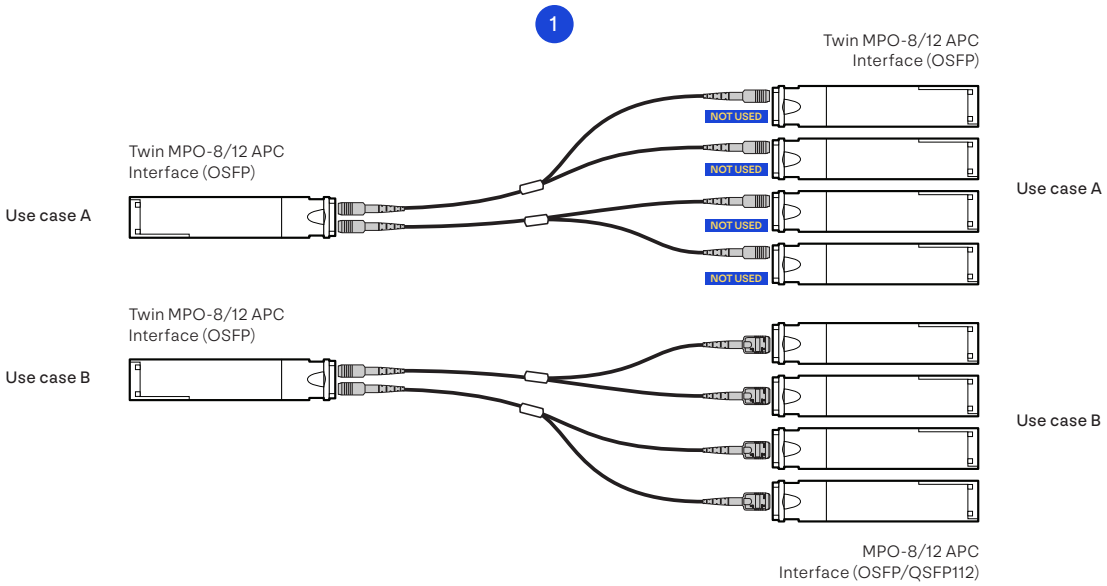


Figure 3. Use cases for Point-to-Point cabling utilizing 1600G, 800G, 400G and 200G transceivers working with MPO-8/12 APC.

Item	Reference Part Number	OS2 Part Number (Americas)	OS2 Part Number (EMEA and APJ)	OM4 Part Number (Americas)	OM4 Part Number (EMEA and APJ)	Description
1	Corning Component [a]	HE8E808GPH-LBxxxF	HE8E808GLZ-LBxxxM	HE9E908QPH-LBxxxF	HE9E908QLZ-LBxxxM	EDGE8®, 8F Y-Harness, MTP® APC (non-pinned) to two 4F (non-pinned) MTP APC, 36-in (910 mm) breakout leg length, Type-B polarity, xxxF (feet) or xxxM (meters)
	or					
	NVIDIA Component [b]	MFP7E40-Nxxx		MFP7E20-Nxxx		NVIDIA MPO-12/APC-to-MPO12/APC (8 fibers), 4-channel-to-two 2-channel splitter fiber cable, Type-B polarity, xxx indicate length in meters

Notes:

a) Corning cables in the Americas use Plenum cable, while EMEA/APJ uses LSZH™/CPR rated cable.

b) Y-Harness lengths available from 1 to 60 meters. Lengths in meters are also available for the Americas.

c) NVIDIA splitter cables utilize a dual rated OFNR/LSZH jacket. SMF and MMF splitter cable lengths available in 3, 5, 7, 10, 15, 20, 30 and 50 meters.

d) Both Corning and NVIDIA cables support InfiniBand, Ethernet and NVLink protocols.

e) A 400G transceiver version will be able to support 200G utilizing a splitter cable (Y-Harness), activating two of the four lanes (4 out of 8-fibers) in the 400G transceiver creating a 200G device.

f) Please review Corning's polarity drawings in Annex 2.

Table 4. Scenario 3 – 1600G, 800G, 400G and 200G using Point-to-Point Cabling with MPO-8/12 APC. Part Number Scheme.

1.4. Scenario 4 – 1600G, 800G, 400G and 200G – Server to Switch Applications MPO-8/12 APC to MPO-8/12 APC using Structured Cabling Across DC with Trunk

Application: InfiniBand Quantum-2 / InfiniBand Quantum-3 / Ethernet Spectrum-4 to InfiniBand Quantum-2 / InfiniBand Quantum-3 / Ethernet Spectrum-4; b) ConnectX-8 / ConnectX-7 / Bluefield-3

Scenario 4 is used with Structured Cabling components, where a 400G transceiver version will be able to support 200G by utilizing a splitter cable (Y-Harness) that activates two of the four lanes (4 out of 8 fibers) in the 400G transceiver, effectively creating a 200G device. In this scenario, as the active devices are physically located in different areas within the data center. Please refer to section 1.7., Table 9, to review the list of transceivers applicable for Use cases A and B in Scenarios 3 and 4.

To learn more about Scalable Units and how to build a GPU Cluster utilizing NVL72 Systems, please refer to section 2 in this document.

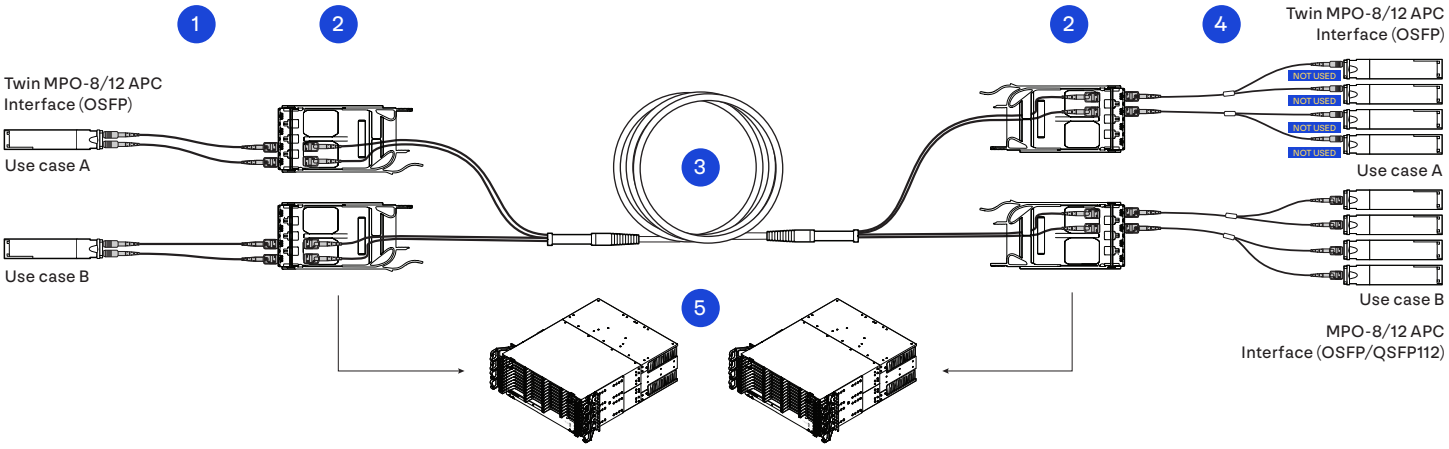


Figure 4. Use cases for Structured Cabling utilizing 1600G, 800G, 400G and 200G transceivers working with MPO-8/12 APC.

Item	Reference Part Number	OS2 Part Number (Americas)	OS2 Part Number (EMEA and APJ)	OM4 Part Number (Americas)	OM4 Part Number (EMEA and APJ)	Description
1	Corning Component[a]	JE8E808GE8-NBxxxF[b]	JE8E808GEZ-NBxxxM[b]	JE9E908QE8-NBxxxF[b]	JE9E908QEZ-NBxxxM[b]	EDGE8®, MTP® APC (non-pinned) to MTP APC (non-pinned) 8F Jumper, TIA-568 Type-B polarity, xxxF (feet) or xxxM (meters)
2		EDGE8-CP32-V1	EDGE8-CP32-V1	EDGE8-CP32-V3	EDGE8-CP32-V3	EDGE8 32-F MTP Adapter Panel, (4-port)
3		GE7E7E4GPNDDU xxxF[c]	GE7E7E4GLZDDU xxxM[c]	GE2E2E4QPNDDU xxxF[c]	GE2E2E4QLZDDU xxxM[c]	EDGE8, MTP Trunk , 144F, MTP APC (pinned) to MTP APC (pinned), 33-in (840-mm) legs, Type-B polarity, pulling grip on first end only, xxxF (feet) or xxxM (meters)
4		HE8E808GPH-LBxxxF[d]	HE8E808GLZ-LBxxxM[d]	HE9E908QPH-LBxxxF[d]	HE9E908QLZ-LBxxxM[d]	EDGE8, 8-F Y-Harness, MTP APC (non-pinned) to two 4F (non-pinned) MTP APC, 36-in (910-mm) breakout leg length, Type-B polarity, xxxF (feet) or xxxM (meters)
5		EDGE8-xxU	EDGE8-xxU	EDGE8-xxU	EDGE8-xxU	Please refer to Annex 1 to choose the best option for your application
Instead of traditional jumpers, the following alternatives are available for Item 1						
1	Corning Component[a]	G-BND18-E8E8G-PN000-xxxF[e]	–	G-BND18-E9E9T-PN000-xxxF[e]	–	18x 8F Bundled Jumper, MPO APC (non-pinned) to MPO APC (non-pinned), 86-in (2.2-m) legs, Type-B polarity, xxxF (feet) or xxxM (meters) Grip on one end
		–	GEDEDE4GEZ-033-UxxxM[f]	–	GEFEFE4QEZ-033-UxxxM[f]	18x 8F CORE-Trunk, MPO APC (non-pinned) to MPO APC (non-pinned), 86-in (2.2-m) legs, Type-B polarity, xxxF (feet) or xxxM (meters)

Notes:
a) Corning cables in the Americas use Plenum cable, while EMEA/APJ uses LSZH™/CPR rated cable.
b) Single jumper lengths are available from 1 to 300 meters. Lengths in meters are also available for the Americas.
c) EDGE8® Trunks are available in fiber counts of 8 to 288 fibers.
d) Y-Harness lengths available from 1 to 60 meters. Lengths in meters are also available for the Americas.
e) Bundled Jumpers lengths are available from 5 to 30 meters (furcation-to-furcation) and are available only in OS2 and OM3 fiber. Bundled Jumpers use a meshed sleeve. Bundled Jumpers are available in 16, 18, 20, 32, 36 legs, and other custom leg quantities with straight or staggered legs. Lengths in meters are also available for the Americas.
f) CORE-Trunks lengths are available from 5 to 300 meters (furcation-to-furcation). CORE-Trunks are available in 16, 18, 32, 36 legs, and other custom leg quantities with straight or staggered legs.
g) All Corning cables support InfiniBand, Ethernet and NVLink protocols.
h) A 400G transceiver version will be able to support 200G utilizing a splitter cable (Y-Harness), activating two of the four lanes (4 out of 8-fibers) in the 400G transceiver creating a 200G device.
i) Please review Corning's polarity drawings in Annex 2.

Table 5. Scenario 4 – 1600G, 800G, 400G and 200G using Structured Cabling with MPO-8/12 APC. Part Number Scheme.

1.5. Scenario 5 – 800G and 400G – Switch to Switch Applications LC-Duplex to LC-Duplex using Point-to-Point Cabling

Application: InfiniBand Quantum-2 / Ethernet Spectrum-4 to InfiniBand Quantum-2 / Ethernet Spectrum-4

Scenario 5 is primarily used for Point-to-Point cabling applications, which involves connecting active devices located within the same rack or row. However, it is not recommended to use Point-to-Point cabling if those devices are physically located in different areas within the data center.

The main application for the Twin LC-Duplex transceiver is to link Quantum-2 or Spectrum-4 air-cooled switches together. A bright green marking on transceiver body indicates 2 km maximum reach.

Please refer to section 1.7., Table 10, to review the list of transceivers applicable for Use cases A and B in Scenarios 5 and 6.

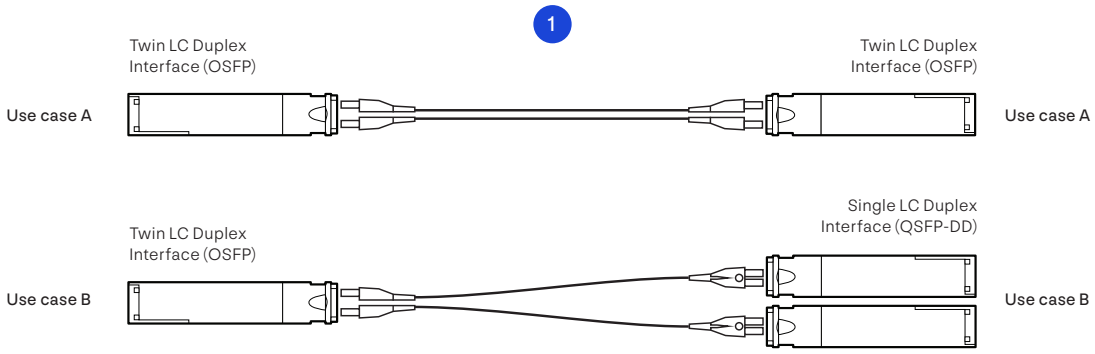


Figure 5. Use cases for Point-to-Point cabling utilizing 800G and 400G transceivers working with LC Duplex.

Item	Reference Part Number	OS2 Part Number (Americas)	OS2 Part Number (EMEA and APJ)	Description
1	Corning [(a)]	PUPU02GD116xxxF	PUPU02GD816xxxM	EDGE™ LC UPC Uniboot to LC UPC Uniboot Duplex Jumper, xxxF or xxxM

Notes:

a) Corning cables in the Americas use Plenum cable, while EMEA/APJ uses LSZH™/CPR rated cable. Jumper lengths are available from 1 to 300 meters. Lengths in meters are also available for the Americas.

b) Corning cables support InfiniBand, Ethernet and NVLink protocols.

c) Please review Corning's polarity drawings in Annex 2.

Table 6. Scenario 5 – 800G and 400G using Point-to-Point Cabling with LC Duplex. Part Number Scheme.

1.6. Scenario 6 – 800G and 400G – Switch to Switch Applications LC-Duplex UPC to LC-Duplex UPC using Structured Cabling Across DC with Trunk

Application: InfiniBand Quantum-2 / Ethernet Spectrum-4 to InfiniBand Quantum-2 / Ethernet Spectrum-4

In Scenario 6, the implementation of Structured Cabling utilizes an optical fiber trunk with MPO-8/12 connectivity as a backbone. This application is used to connect different switches, such as Leaf to Spine. It is the recommended option to follow when two different active devices are physically located in different areas within the data center.

The main application for the Twin LC-Duplex transceiver is to link Quantum-2 or Spectrum-4 air-cooled switches together. A bright green marking on transceiver body indicates 2 km maximum reach.

Please refer to section 1.7., Table 10, to review the list of transceivers applicable for Use cases A and B in Scenarios 5 and 6.

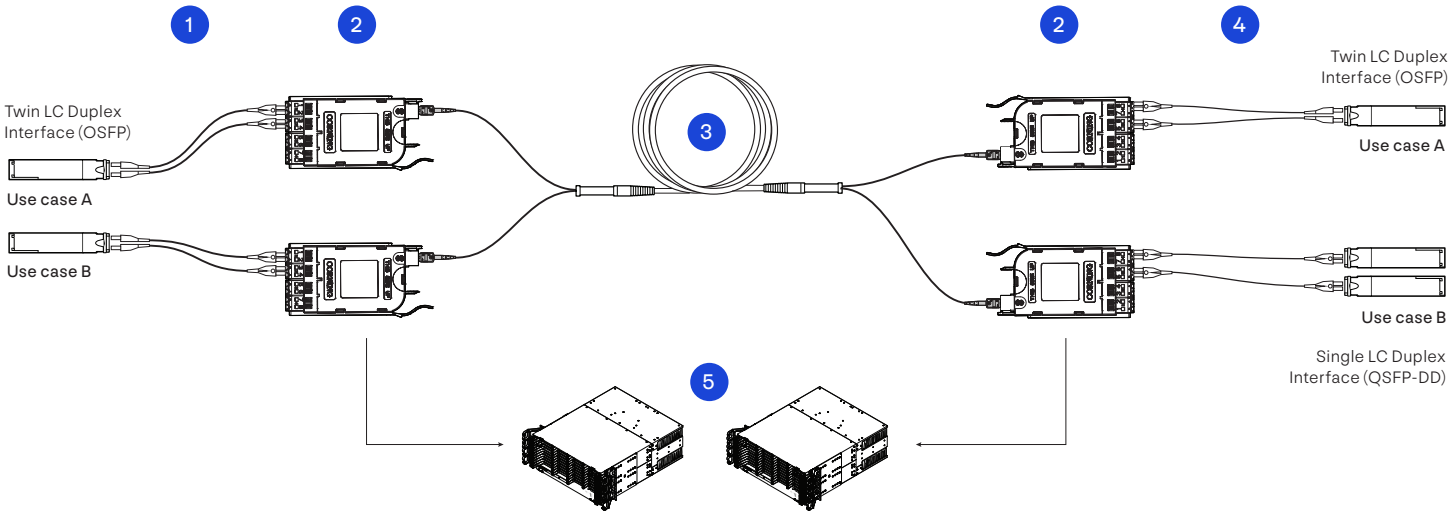


Figure 6. Use cases for Structured Cabling utilizing 800G and 400G transceivers working with LC Duplex.

Item	Reference Part Number	OS2 Part Number (Americas)	OS2 Part Number (EMEA and APJ)	Description
1	Corning [a]	PUPU02GD116xxxF	PUPU02GD816xxxM	EDGE™ LC UPC Uniboot to LC UPC Uniboot Duplex Jumper, xxxF or xxxM
2		ECM8-UM08-04-E8G-ULL	ECM8-UM08-04-E8G-ULL	EDGE8® Ultra-Low-Loss Module, LC duplex to MTP® (non-pinned), 8F, universal polarity
3		GE7E7E4GPNDDUxxxF[b]	GE7E7E4GLZDDUxxxM[b]	EDGE8, MTP Trunk , 144F, MTP APC (pinned) to MTP APC (pinned), 33-in (840 mm) legs, Type-B polarity, pulling grip on first end only, xxxF (feet) or xxxM (meters)
4		EDGE8-xxU	EDGE8-xxU	Please refer to Annex 1 to choose the best option for your application

Notes:

a) Corning cables in the Americas use Plenum cable, while EMEA/APJ uses LSZH™/CPR rated cable. Jumper lengths are available from 1 to 300 meters. Lengths in meters are also available for the Americas.

b) EDGE8® Trunks are available in fiber counts of 8 to 288 fibers.

c) Corning cables support InfiniBand, Ethernet and NVLink protocols.

d) Please review Corning's polarity drawings in Annex 2.

Table 7. Scenario 6 – 800G and 400G using Structured Cabling with LC Duplex connectivity at the active device. Part Number Scheme.

1.7. Transceiver Options and Port Breakout Connections per Scenario

To read the following tables, please follow the examples:

Example 1: In Scenario 2, Use Case B, using Table 8, an 800G-2x DR4, MMS4X00-NS with 2x MPO-8/12 APC connections on the left side of the drawing can connect to two OSFP transceivers, MMS4X00-NS400, resulting in two individual 400G connections on the right side.

Example 2: In Scenario 4, Use Case B, using Table 9, an 800G-2x DR4, MMS4X00-NS with 2x MPO-8/12 APC connections on the left side of the drawing can connect to four 200G-DR4, QSFP112, MMS1X00-NS400 utilizing Y-Harnesses, resulting in four individual 200G connections on the right side.

Use case	Near End Optic (Left)			Far End Optic (Right)			Reach	Fiber Type
	Speed	NVIDIA Port	Footprint - Fiber/ Transceiver	Speed	NVIDIA Port	Footprint - Fiber/ Transceiver		
A	1600G-2x DR4	MMS4A00	OSFP - 2x 8F	1600G-2x DR4	MMS4A00	OSFP - 2x 8F	500 m	Single-mode
	800G-2x DR4	MMS4X00-NM	OSFP - 2x 8F	800G-2x DR4	MMS4X00-NM	OSFP - 2x 8F	500 m	Single-mode
	800G-2x DR4	MMS4X00-NM-T	OSFP - 2x 8F	800G-2x DR4	MMS4X00-NM-T	OSFP - 2x 8F	500 m	Single-mode
	800G-2x DR4	MMS4X00-NS	OSFP - 2x 8F	800G-2x DR4	MMS4X00-NS	OSFP - 2x 8F	100 m	Single-mode
	800G-2x DR4	MMS4X00-NS-FLT	OSFP - 2x 8F	800G-2x DR4	MMS4X00-NS	OSFP - 2x 8F	100 m	Single-mode
	800G-2x DR4	MMS4X00-NS-T	OSFP - 2x 8F	800G-2x DR4	MMS4X00-NS-T	OSFP - 2x 8F	100 m	Single-mode
B	800G-2x DR4	MMS4X00-NS	OSFP - 2x 8F	400G-DR4	MMS4X00-NS400	OSFP - 8F	100 m	Single-mode
	800G-2x DR4	MMS4X00-NS-T	OSFP - 2x 8F	400G-DR4	MMS4X00-NS400-T	OSFP - 8F	100 m	Single-mode
C	1600G-2x DR4	MMS4A00	OSFP - 2x 8F	800G-2x DR4	MS4A20-XM800	OSFP - 8F	500m	Single-mode
	800G-2x DR4	MMS4X00-NS	OSFP - 2x 8F	400G-DR4	MMS1X00-NS400	QSFP112 - 8F	100 m	Single-mode
	800G-2x DR4	MMS4X00-NS-T	OSFP - 2x 8F	400G-DR4	MMS1X00-NS400-T	QSFP112 - 8F	100 m	Single-mode
A	800G-2x SR4	MMA4Z00-NS	OSFP - 2x 8F	800G-2x SR4	MMA4Z00-NS	OSFP - 2x 8F	50 m	Multimode
	800G-2x SR4	MMA4Z00-NS-FLT	OSFP - 2x 8F	800G-2x SR4	MMA4Z00-NS	OSFP - 2x 8F	50 m	Multimode
	800G-2x SR4	MMA4Z00-NS-T	OSFP - 2x 8F	800G-2x SR4	MMA4Z00-NS-T	OSFP - 2x 8F	50 m	Multimode
B	800G-2x SR4	MMA4Z00-NS	OSFP - 2x 8F	400G-SR4	MMA4Z00-NS400	OSFP - 8F	50 m	Multimode
	800G-2x SR4	MMA4Z00-NS-T	OSFP - 2x 8F	400G-SR4	MMA4Z00-NS400-T	OSFP - 8F	50 m	Multimode
C	800G-2x SR4	MMA4Z00-NS	OSFP - 2x 8F	400G-SR4	MMA1Z00-NS400	QSFP112 - 8F	50 m	Multimode
	800G-2x SR4	MMA4Z00-NS-T	OSFP - 2x 8F	400G-SR4	MMA1Z00-NS400-T	QSFP112 - 8F	50 m	Multimode

Table 8. Transceivers Applicable for Use Cases A to C in Scenarios 1 and 2.

Use case	Near End Optic (Left)			Far End Optic (Right)			Reach	Fiber Type
	Speed	NVIDIA Port	Footprint - Fiber/ Transceiver	Speed	NVIDIA Port	Footprint - Fiber/ Transceiver		
A	800G-2x DR4	MMS4X00-NS	OSFP - 2x 8F	200G-DR4	MMS4X00-NS400	OSFP - 4F	100 m	Single-mode
	800G-2x DR4	MMS4X00-NS-T	OSFP - 2x 8F	200G-DR4	MMS4X00-NS400-T	OSFP - 4F	100 m	Single-mode
B	1600G-2x DR4	MMS4A00	OSFP - 2x 8F	400G-DR4	MS4A20-XM800	OSFP - 4F	500 m	Single-mode
	800G-2x DR4	MMS4X00-NS	OSFP - 2x 8F	200G-DR4	MMS1X00-NS400	QSFP112 - 4F	100 m	Single-mode
	800G-2x DR4	MMS4X00-NS-T	OSFP - 2x 8F	200G-DR4	MMS1X00-NS400-T	QSFP112 - 4F	100 m	Single-mode
A	800G-2x DR4	MMA4Z00-NS	OSFP - 2x 8F	200G-SR4	MMA4Z00-NS400	OSFP - 4F	50 m	Multimode
	800G-2x DR4	MMA4Z00-NS-T	OSFP - 2x 8F	200G-SR4	MMA4Z00-NS400-T	OSFP - 4F	50 m	Multimode
B	800G-2x DR4	MMA4Z00-NS	OSFP - 2x 8F	200G-SR4	MMA1Z00-NS400	QSFP112 - 4F	50 m	Multimode
	800G-2x DR4	MMA4Z00-NS-T	OSFP - 2x 8F	200G-SR4	MMA1Z00-NS400-T	QSFP112 - 4F	50 m	Multimode

Table 9. Transceivers Applicable for Use Cases A and B in Scenarios 3 and 4.

Use case	Near End Optic (Left)			Far End Optic (Right)			Reach	Fiber Type
	Speed	NVIDIA Port	Footprint - Fiber/ Transceiver	Speed	NVIDIA Port	Footprint - Fiber/ Transceiver		
A	800G-2x FR4	MMS4X50-NM	OSFP - 2x 2F	800G-2x FR4	MMS4X50-NM	OSFP - 2x 2F	2 km	Single-mode
	800G-2x FR4	MMS4X50-NM-T	OSFP - 2x 2F	800G-2x FR4	MMS4X50-NM-T	OSFP - 2x 2F	2 km	Single-mode
B	800G-2x FR4	MMS4X50-NM	OSFP - 2x 2F	400G-FR4	MMS1V50-WM	QSFP-DD - 2F	2 km	Single-mode
	800G-2x FR4	MMS4X50-NM-T	OSFP - 2x 2F	400G-FR4	MMS1V50-WM-T	QSFP-DD - 2F	2 km	Single-mode

Table 10. Transceivers Applicable for Use Cases A and B in Scenarios 5 and 6.

2. Cabling Architecture Reference Guide for NVIDIA NVL72 Systems

In this section, we will review how to cable an NVIDIA NVL72 System. Although **this guide uses the NVIDIA NVL72 GB200 InfiniBand as a reference,** it is important to note that similar cabling components and infrastructure will be used for applications involving the **NVL72 GB200 Ethernet, NVL72 GB300 InfiniBand and Ethernet,** as well as upcoming Rubin platforms. The specific considerations for each GPU Cluster will depend on individual customer designs.

An NVL72 Cluster is formed by a certain number of Scalable Units (SUs), each containing a specific number of nodes (servers). This configuration determines the total number of GPUs in the cluster. The Scalable Unit serves as the building block for any GPU Cluster.

It is crucial to understand that the size of the cluster, the system (GB200 or GB300), as well as the protocol (InfiniBand or Ethernet), will dictate the required number of NVL72 racks, Leaf Switches, Spine Switches, and Core Switches. Based on this configuration, we can calculate the number of cables or connections needed.

In Figure 7, we can see the applicable component counts for an InfiniBand GB200 System. For example, a **cluster with 16 Scalable Units** would total 4,608 nodes (servers), which is equivalent to 18,432 GPUs. This setup would require 1,024 IB Leaf Switches, 576 IB Spine Switches, and 288 IB Core Switches. The switch applicable for **InfiniBand GB200** is the **Quantum-2 Q9700.** This cluster size will be used as a reference throughout this document.

The “Cable Counts” column provides the number of individual MPO-8/12 APC connections required to connect the system. For our 16 Scalable Unit example:

- **Node-to-Leaf Switch Connections:** To connect each node (server) to the InfiniBand Leaf Switches, 18,432 MPO-8/12 APC connections are required.
- **Leaf-to-Spine Switch Connections:** To connect the InfiniBand Leaf Switches to the InfiniBand Spine Switches, an additional 18,432 MPO-8/12 APC connections are needed.
- **Spine-to-Core Switch Connections:** To connect the InfiniBand Spine Switches to the InfiniBand Core Switches, another 18,432 MPO-8/12 APC connections are required.

In total, this 16 Scalable Unit cluster example requires **55,296 MPO-8/12 APC connections** across the cluster.

Compute Component Count				InfiniBand Switch Counts			Cable Counts		
SU	NVL 72 Racks	Node	GPU	Leaf	Spine	Core	Node-Leaf	Leaf-Spine	Spine-Core
1	16	288	1,152	64	36	N/A	1,152	1,152	N/A
2	32	576	2,304	128	72	36	2,304	2,304	2,304
4	64	1,152	4,608	256	144	72	4,608	4,608	4,608
8	128	2,304	9,216	512	288	144	9,216	9,216	9,216
16	256	4,608	18,432	1,024	576	288	18,432	18,432	18,432

Figure 7. InfiniBand GB200 Component Counts Reference Table.

2.1. Understanding the Compute Fabric Connections from the nodes (servers) in an NVL72 rack

An NVIDIA NVL72 GB200/GB300 rack is equipped with connectivity ports for an InfiniBand or Ethernet Compute Fabric, a Storage Fabric, an In-Band Management Network, and an Out-of-Band Management Network.

With the option to choose between GB200 and GB300, it is also necessary to differentiate their impact on the topology design and to understand the concepts of **single-plane and multiplane** topology.

A **plane** in networking refers to a distinct, logical layer of connectivity between GPUs and switches within the cluster. Each plane is essentially an independent network path that facilitates communication between the devices in the cluster.

With this in mind, for the Compute Fabric, **what are the key differences between GB200 and GB300 topologies?** Figure 8 provides a summarized repres-entation of what we can expect. Keep in mind that all MPO connectors described are 8-fiber MPO connectors, also referred to as MPO-8, MPO-12, or MPO-8/12. Additionally, they are angled connectors (APC), regardless of whether they are Single-mode or Multimode:

- NVL72 **GB200 InfiniBand and Ethernet:**

  - Each node or server will have 4x GPUs, with each GPU physically represented by **1x MPO-8/12** at 400G **NDR** OSFP.
  - This is equivalent to 72x MPO-8/12 compute/backend connections coming out from a single NVL72 rack.
  - The switches (**Quantum-2 QM9700 or Spectrum-4 SN5600**) for this topology will utilize a dual port MPO port (2x MPO-8/12 = 2x 400G **NDR** OSFP).
  - The topology design is based on a **single-plane** (either two-tier or three-tier architecture).

- NVL72 **GB300 InfiniBand:**

  - Each node or server will have 4x GPUs, with each GPU physically represented by **1x MPO-8/12** at 800G **XDR** OSFP.
  - This is equivalent to 72x MPO-8/12 compute/backend connections coming out from a single NVL72 rack.
  - The switches (**Quantum-3 Q3400-RA or Q3200-RA**) for this topology will utilize a dual MPO port (2x MPO-8/12 = 2x 800G **XDR** OSFP).
  - The topology design is based on a **dual-plane** (either two-tier or three-tier architecture).

- NVL72 **GB300 Ethernet:**

  - Each node or server will have 4x GPUs, with each GPU physically represented by **2x MPO-8/12** at 400G OSFP.
  - This is equivalent to 144x MPO-8/12 compute/backend connections coming out from a single NVL72 rack.
  - The switches (**Spectrum-4 SN5600**) for this topology will utilize a dual MPO port (2x MPO-8/12 = 2x 400G OSFP).
  - The topology design will be based on either a **dual-plane** (either two-tier or three-tier architecture) or **quad-plane** (two-tier architecture).
  - GB300 Ethernet introduces the need for a Shuffle Box in the quad-plane topology to reduce cabling complexity (Figure 36).

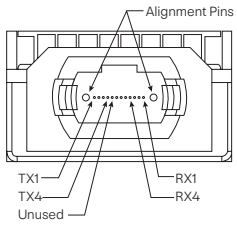
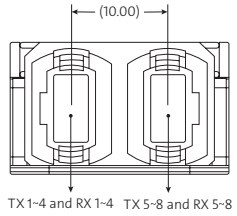
NVIDIA NVL72				
System	Features	InfiniBand	Ethernet	
GB200	• 18x Compute Nodes • 72x GB200 GPUs per NVL72 rack • Utilizes ConnectX-7 network adapters • Utilizes 400/800G Modules	• **Server:** 4x GPUs, each one with 1x MPO (400G) • **Switch:** 2x MPO (2x 400G) • **Topology Design:** Single-plane	• **Server:** 4x GPUs, each one with 1x MPO (400G) • **Switch:** 2x MPO (2x 400G) • **Topology Design:** Single-plane	
GB300	• 18x Compute Nodes • 72x GB300 GPUs per NVL72 rack • Utilizes ConnectX-8 network adapters • Utilizes 800G/1.6T Modules	• **Server:** 4x GPUs, each one with 1x MPO (1x 800G) • **Switch:** 2x MPO (2x 800G) • **Topology Design:** Dual-plane	• **Server:** 4x GPUs, each one with 2x MPO (400G) • **Switch:** 2x MPO (2x 400G) • **Topology Design:** Dual-plane or Quad-plane with Shuffle Box	

Figure 8. NVL72 (Grace-Blackwell) GB200 and GB300 Backend Topology.

2.2. Understanding the NVL72 System Switches

The choice of Leaf, Spine, and Core switches within the NVL72 Cluster depends on the data rate utilized, whether 400G NDR or 800G XDR.

Figure 9 provides a summarized view of the application and the number of ports per switch based on these data rates.

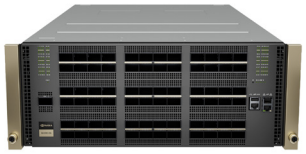
<table>
<tr><th colspan="2">NVL72 Rack</th><th colspan="2">Quantum-2 InfiniBand (NDR)</th></tr>
<tr><td>Servers
• 36 Grace CPUs and 72 Blackwell GPUs
• 18 Compute Trays with 4 GPUs/NICs per tray
• CX7 NIC (400G NDR) or CX8 NIC (800G XDR)</td><td>  </td><td>QM9700
• 1U switch with 32 OSFP (twin-MPO) ports
• Total of 64 MPO ports 400G NDR over 32 OSFPs</td><td> 
1U QM9700 (InfiniBand) = 64 ports 400G</td></tr>
<tr><th colspan="4">Quantum-3 InfiniBand (XDR)</th></tr>
<tr><td>QR3400-RA
• 4U switch with 72 OSFP (twin-MPO) ports
• Total of 144 MPO ports 800G XDR over 72 OSFPs</td><td> 
4U Q3400-RA (InfiniBand) = 144 ports 800G</td><td>QR3200-RA
• 2U switch with 2 independent 18 OSFP (twin MPO) port switches in a single enclosure (no communication between the 2 switches)
• Total of 144 MPO ports 800G XDR over 72 OSFPs</td><td> 
2U Q3200-RA (InfiniBand) = 72 ports 800G</td></tr>
<tr><th colspan="4">Spectrum-4 SN5000 Series Ethernet</th></tr>
<tr><td>SN5600
• 2U switch with 64 OSFP (twin-MPO) ports
• Total of 128 ports 400G over 64 OSFPs</td><td> 
2U SN5600 (Ethernet) = 128 Ports 400G</td><td>SN5400
• 2U switch with 64 QSFP-DD (single-MPO) ports
• Total of 64 MPO ports 400G</td><td> 
2U SN5400 (Ethernet) = 64 Ports 400G</td></tr>
</table>

Figure 9. InfiniBand and Ethernet Switch application based on the backend network data rate.

Let's analyze the applications:

- **InfiniBand 400G NDR,** commonly used with **NVL72 GB200** Systems, utilizes **Quantum-2** switches, which support 32 OSFP Twin MPO-8/12 APC ports. This means that each switch can provide up to 64 single ports of 400G NDR.

  Figure 10 illustrates an example of port mapping at the switch level, providing a general reference that applies to all switch types. This mapping example can also be extended to InfiniBand Quantum-3 and Ethernet Spectrum-4 switches. It is important to note that this is just one possible method for customers to map these ports.

  Quantum-2 switches will be deployed across the network in various roles, such as Leaf, Spine, or Core.

Orderable Part Number (OPN)	Description	
MQM9700-NS2F	64-ports 400Gb/s, 32 OSFP ports, managed, power-to-connector (P2C) airflow (forward)	
MQM9700-NS2R	64-ports 400Gb/s, 32 OSFP ports, managed, connector-to-power (C2P) airflow (reverse)	

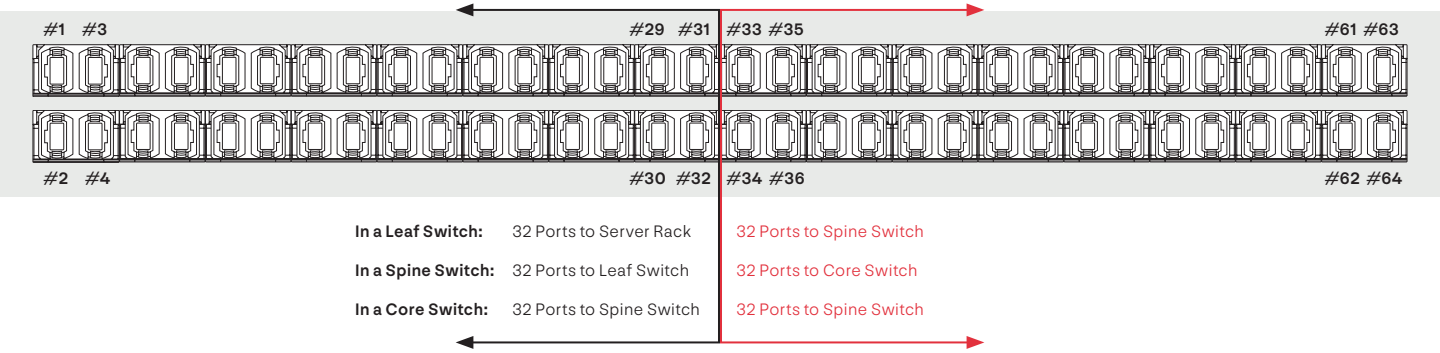


Figure 10. Example of port mapping for Quantum-2 switches.

- **InfiniBand 800G XDR,** commonly used with **NVL72 GB300** Systems, utilizes either **Quantum-3 Q3400-RA** or **Q3200-RA** switches.

  Q3400-RA can be used as Leaf, Spine or Core switches. They support 72 OSFP Twin MPO-8/12 APC ports, which means that each switch can provide up to 144 single ports of 800G XDR.

  Q3200-RA can be used as Leaf switches. A single Q3200-RA enclosure functions as two independent switches (one on the left, one on the right of the enclosure). Each side supports 18 OSFP Twin MPO-8/12 APC ports, providing up to 36 single ports of 800G XDR per side, resulting in a maximum density of up to 72 single ports of 800G XDR for the entire enclosure.

- **Ethernet 400G,** commonly used with **NVL72 GB200 and GB300** Systems, utilizes either **Spectrum-4 SN5600** or **SN5400** switches.

  SN5600 can be used as Leaf, Spine or Core switches. They support 64 OSFP Twin MPO-8/12 APC ports, which means that each switch can provide up to 128 single ports of 400G.

  SN5400 can be used as Leaf switches. They support 64 QSFP-DD single MPO-8/12 APC ports of 400G.

2.3. Understanding the concept of Scalable Unit, the building block of a GPU Cluster

The NVL72 cluster is built using building blocks called Scalable Units (SUs). For InfiniBand GB200 and GB300, each SU contains 16 NVL72 systems or racks. For Ethernet GB200, each SU contains 4 NVL72 systems. For Ethernet GB300, each SU contains 2 NVL72 systems. In all cases, the SU enables the rapid deployment of systems of various sizes.

For the purpose of this guide, we will focus on exemplifying an InfiniBand NVL72 GB200 deployment. Each InfiniBand NVL72 GB200 SU contains:

- 1,152 GPUs distributed across 16 NVL72 racks
- 64 Leaf Switches (Quantum-2)
- 36 Spine Switches (Quantum-2)

With GB300 InfiniBand and Ethernet, the number of switches will vary; however, the same products, logic, and concepts can be applied to those clusters.

Just as in DGX SuperPOD H100/H200 and B100/B200, the fabric in NVL72 systems is also Rail-optimized. The difference is that NVL72 systems use a 4-Rail-based architecture, whereas H100/H200 and B100/B200 use an 8-Rail architecture. A Rail refers to Host Channel Adapters (HCAs) with the same port number from each node being connected to the same Leaf Switch. For example, all Port 1 connections from the 18 servers in an NVL72 rack represent Rail 1; all Port 2 connections from the same 18 servers represent Rail 2; all Port 3 connections represent Rail 3; and all Port 4 connections represent Rail 4 (See Figure 13).

Each set of 18 connections from each Rail in an NVL72 rack will be connected to its respective Leaf Switch, meaning one Leaf Switch per Rail per NVL72 rack. As the scalable unit consists of 16 NVL72 racks, this means there will be 16 different Leaf Switches for Rail 1, 16 for Rail 2, 16 for Rail 3, and 16 for Rail 4.

Differing from the DGX SuperPOD architecture, where the Scalable Unit (SU) consisted only of servers and Leaf Switches, the NVL72 SU also includes a Spine switching layer. For the 16 NVL72 GB200 SU, there will be 9 Spine Switches. Each Leaf Switch in the same Rail will connect to 2 ports on each Spine Switch in the Spine Switch rack, which consolidates all Rail 1 connections. Similarly, there will be a separate rack with 9 Spine Switches for Rail 2, another rack with 9 Spine Switches for Rail 3, and yet another rack with 9 Spine Switches for Rail 4.

For NVL72 GB200, the fabric is constructed using Quantum-2 9700 InfiniBand switches with 2x 400G Twin-port OSFP transceivers.

The layout of the Scalable Unit is configurable. Figure 11 illustrates some of the possible layout examples with the relevant Compute Fabric components. It is important to note that other types of layouts can be implemented. Therefore, maintaining close communication with Corning's Engineering team during the design phase is crucial to validate the optimal cabling options for your design choice.

2-Tier Architecture, One Row SU (Scalable Cluster Ready)

2-Tier Architecture, Two Row SU (Scalable Cluster Ready)

3-Tier Architecture, with multiple SUs (2 or more in a Cluster)

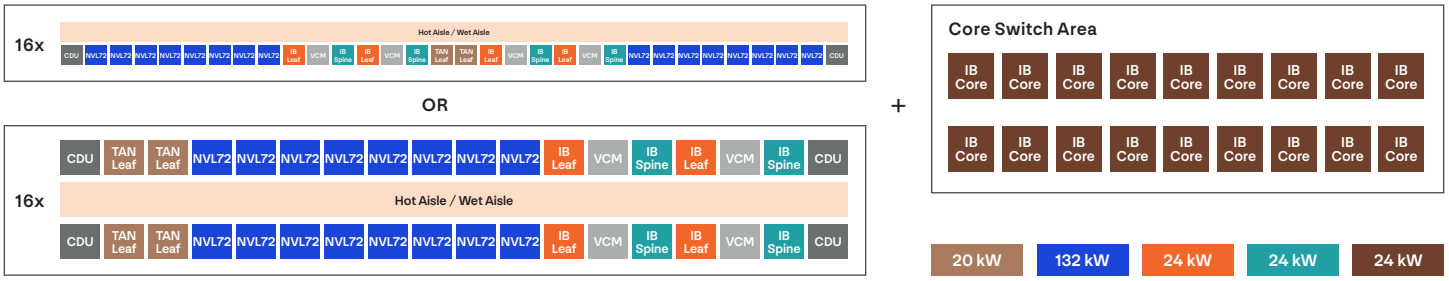


Figure 11. The Scalable Unit (SU) layout is configurable.

2.4. Implementing the Cabling Scenarios within an NVIDIA NVL72 Cluster

To facilitate the identification of the different cabling components used in building an AI/ML cluster, Corning will utilize three levels of connectivity in this guide. These levels and the number of switches are based on the example of a 16 Scalable Unit Cluster:

- Level A – Server-to-Leaf cabling
- Level B – Leaf-to-Spine cabling
- Level C – Spine-to-Core cabling

2.4.1. Level A – Server-to-Leaf cabling

A Scalable Unit can be cabled using Point-to-Point connections between the node (server) and the Leaf Switch (see Figure 1), with at least two cabling product options available (see Figure 12). In some specific custom designs, implementing Structured Cabling at the Scalable Unit level is also possible (see Figure 2).

The first cabling product option is to use **traditional, individual 8-fiber MPO jumpers** to establish connections from each NVL72 rack to each Leaf Switch, and from the Leaf Switch to the Spine Switch within the SU. Depending on your cabling supplier, individual jumpers may have larger cable diameters, which can affect cable management and routing inside and outside the racks. Point-to-Point connections from the Spine Switch to the Core Switch are also possible; however, due to the cable density and distance between the switches, a Structured Cabling solution may be preferred.

The second cabling product option is to use a mix of **CORE-Trunks with 144 fibers and 128 fibers.** A CORE-Trunk is a multifiber solution for Point-to-Point architectures that consolidates 8-fiber MPO-8/12 APC connections into a single multifiber unit with a flame-retardant jacket. Point-to-Point connections from the Spine Switch to the Core Switch are also possible using CORE-Trunks; however, due to the cable density and distance between the switches, a combination of CORE-Trunks and Structured Cabling may be preferred.

In either case, the product choice will depend on the specific requirements of the customer's design.

Figure 12 illustrates the number of components or pieces required in a Scalable Unit, depending on the choice between individual jumpers or CORE-Trunks.

Regardless of the choice, we will establish a total of 1,152 MPO-8/12 APC connections to our Leaf Switches, 1,152 connections to our Spine Switches within the SU, and an additional 1,152 connections to our Core Switches to connect the SU to the Core Switches.

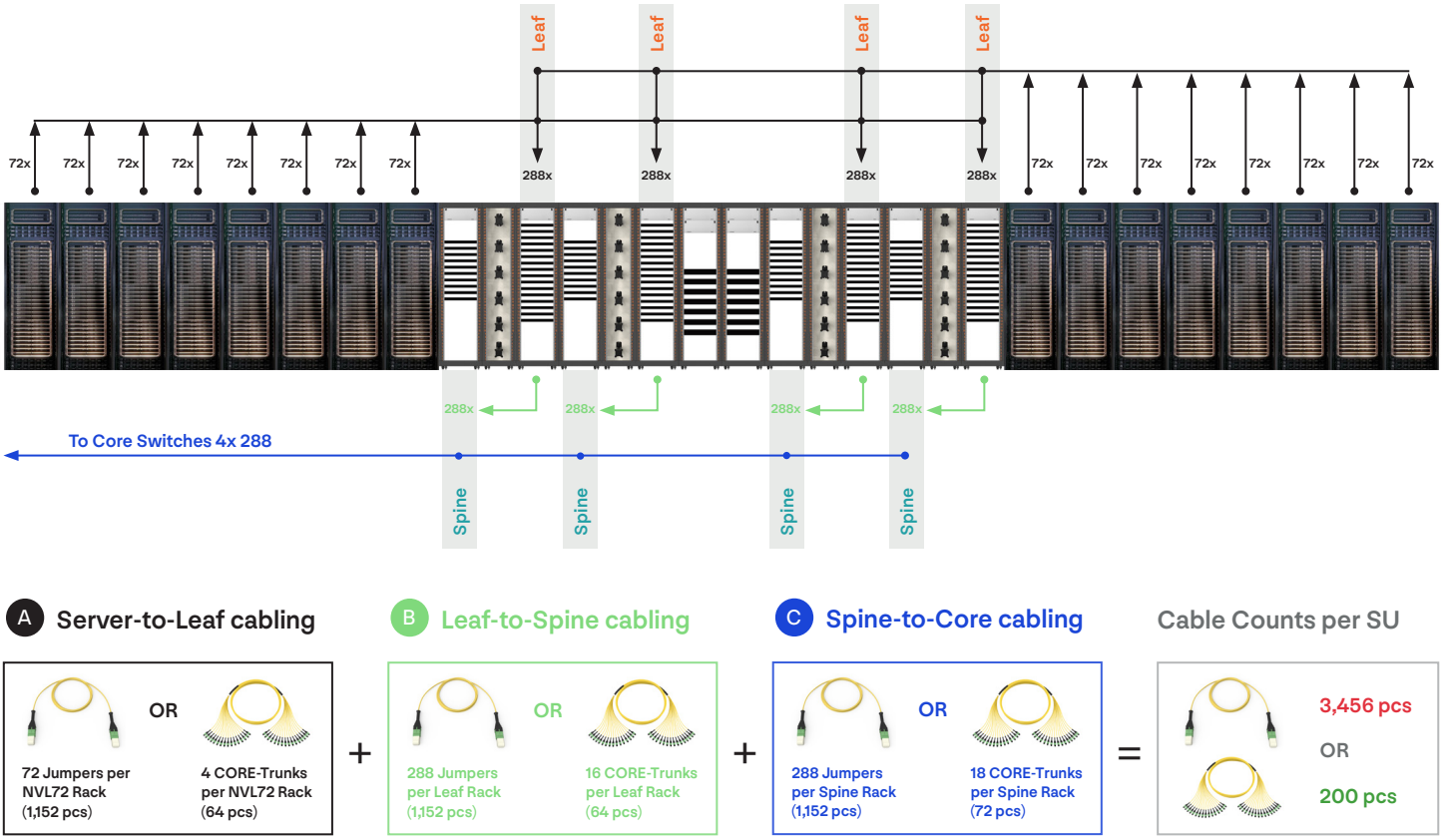


Figure 12. Number of connections applicable to a Scalable Unit and two product approaches.

Previously, we described the Compute Fabric as Rail-optimized, with a total of 4 Rails. These Rails correspond to the MPO-8/12 APC connections from each server. In the following examples, we will use color references, as indicated in Figure 13, to indicate each of these Rail connections.

For Level A, we will illustrate the connections from the NVL72 Rack to the InfiniBand Leaf Switch Racks within our Scalable Unit using CORE-Trunks, as they simplify cabling in high-density GPU clusters. If traditional individual jumpers are chosen, the rail mapping can be followed accordingly.

In Figure 13, on the left, we can visualize an NVL72 Rack with 18 servers, each tray within the NVL72 Rack depicting their respective Rails. On the right, we find the Leaf Switch Rack for Rail 1 with 16 Leaf Switches, one for each NVL72 Rack within my Scalable Unit. As we know, Quantum-2 switches allow for 32 Twin MPO-8/12 APC ports, meaning each switch can support up to 64 individual connections.

Each SU of 16x NVL72 racks requires a total of 1,152 MPO-8/12 APC 8-fiber connections (4 per Server – 72 per Rack) to the Leaf Switches.

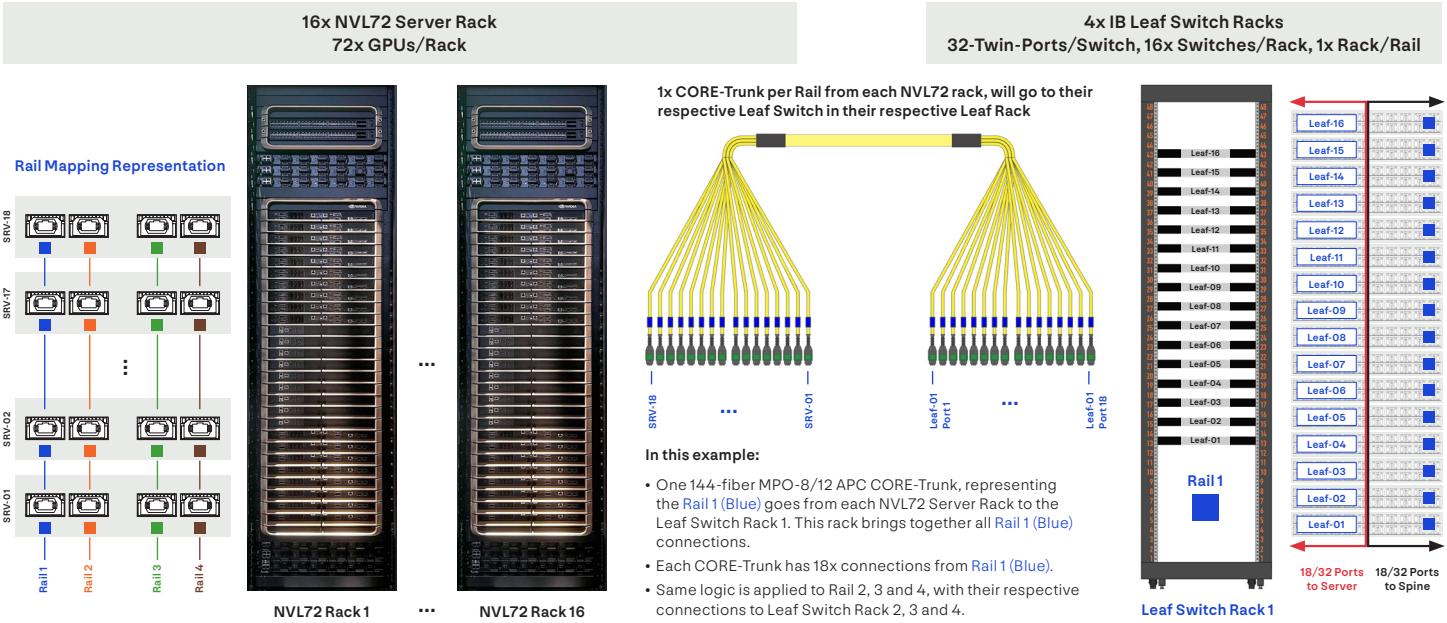
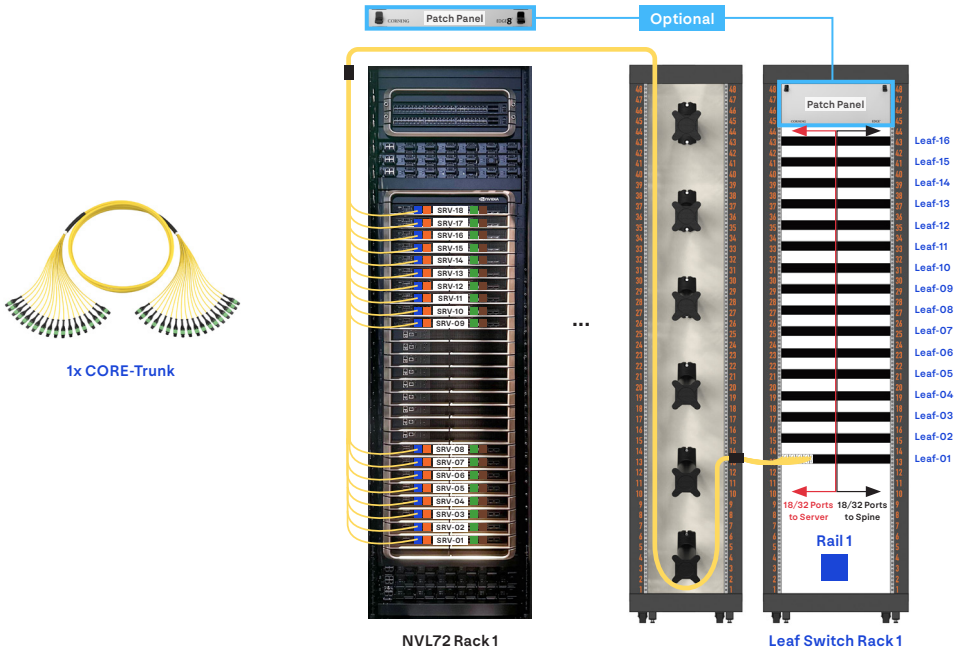


Figure 13. Level A – Server-to-Leaf cabling utilizing Point-to-Point connection with a CORE-Trunk – Example based on 16 SU Cluster.

Cabling is performed using one CORE-Trunk per Rail for each NVL72 Rack, routing all Rails of the same color from each server to their corresponding Leaf Switch. For example, Rail 1 (Blue Rail) from NVL72 Rack 1 connects to Leaf-01 Switch ports 1-18; Rail 1 (Blue Rail) from NVL72 Rack 2 connects to Leaf-02 Switch ports 1-18; and so on. In a Scalable Unit with 16 NVL72 Racks, Rail 1 completes the first 18 port connections (out of 64) on each Leaf Switch, as shown in Figures 14 and 15. Additionally, each Leaf Switch provides 18 uplink connections to the Spine Switches. This process is repeated for every Rail across all NVL72 Racks, ensuring the Scalable Unit's connections are fully mapped and completed.

Patch panels can optionally be added to the Scalable Unit when cabling NVL72 racks to Leaf Switch Racks, either if Structured Cabling is used or if improved mapping and cable management are desired. Figure 14 shows the locations where these patch panels can be implemented.



As previously explained:

- One 144-fiber MPO-8/12 APC CORE-Trunk, representing the Rail 1 (Blue) goes from each NVL72 Server Rack to the Leaf Switch Rack 1. This rack brings together all Rail 1 (Blue) connections = 16x Leaf Switches of each Rail for 16x NVL72 Racks.
- Each CORE-Trunk has 18x connections from Rail 1 (Blue).
- Same logic is applied to Rail 2, 3 and 4, with their respective connections to Leaf Switch Rack 2, 3 and 4.
- Patch Panels can be used if Structured Cabling is designed within the SU and DC layout.

Figure 14. Level A – CORE-Trunk Routing Example.

The same cabling and mapping concept is applied for Rail 2, 3, and 4. In Figure 15, for this layout, we can see 4 Leaf Switch Racks, one per Rail, each containing its respective 18 Leaf Switches. Based on the SU configuration, patch panels can be added if Structured Cabling is implemented within the SU layout.

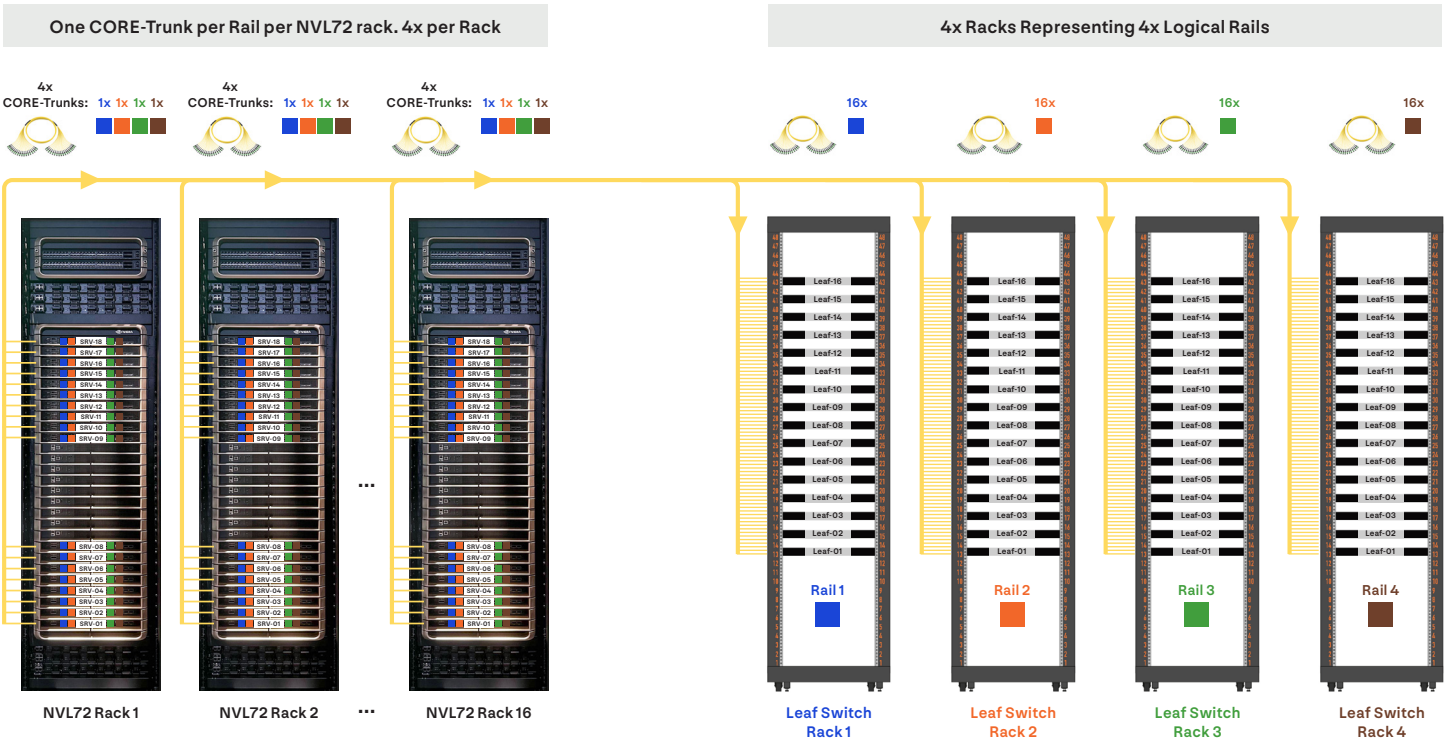


Figure 15. Level A – CORE-Trunk Routing to Leaf Switches.

2.4.2. Level B – Leaf-to-Spine cabling

Since the Leaf and Spine Switches are physically located within the same Scalable Unit, the Leaf Switches can be connected to the Spine Switches using the same product options mentioned earlier (individual jumpers or CORE-Trunks), as well as through Point-to-Point cabling or also utilizing Structured Cabling.

In Figure 16, on the left, we can visualize the Leaf Switch Rack containing all the 16 Leaf Switches of the Rail 1. On the right, we find the Spine Switch Rack with 9 Spine Switches for Rail 1 (Quantum-2, 32 Twin MPO-8/12 APC ports), one for each NVL72 Rack within the Scalable Unit.

Each Leaf Rack requires 288x connections to each Spine Rack. One SU requires a total of 1,152 MPO-8/12 APC 8-fiber connections Leaf-Spine.

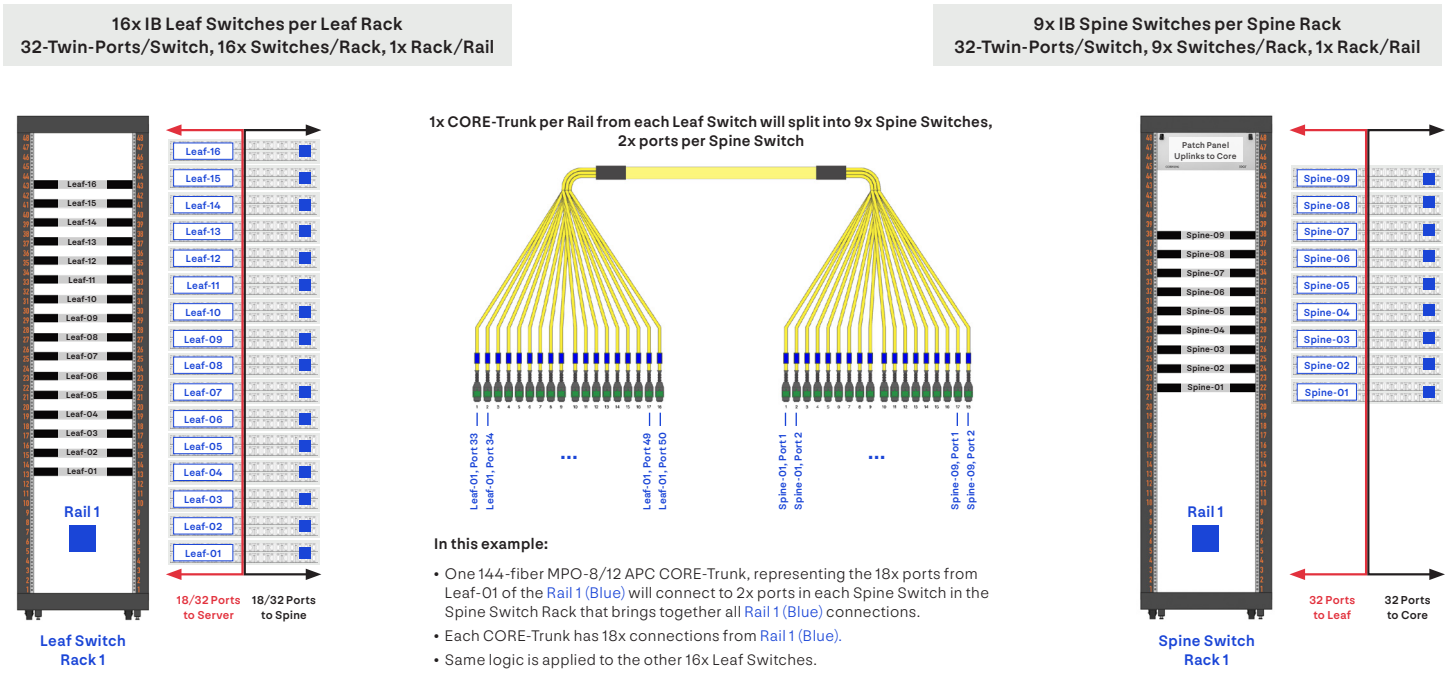
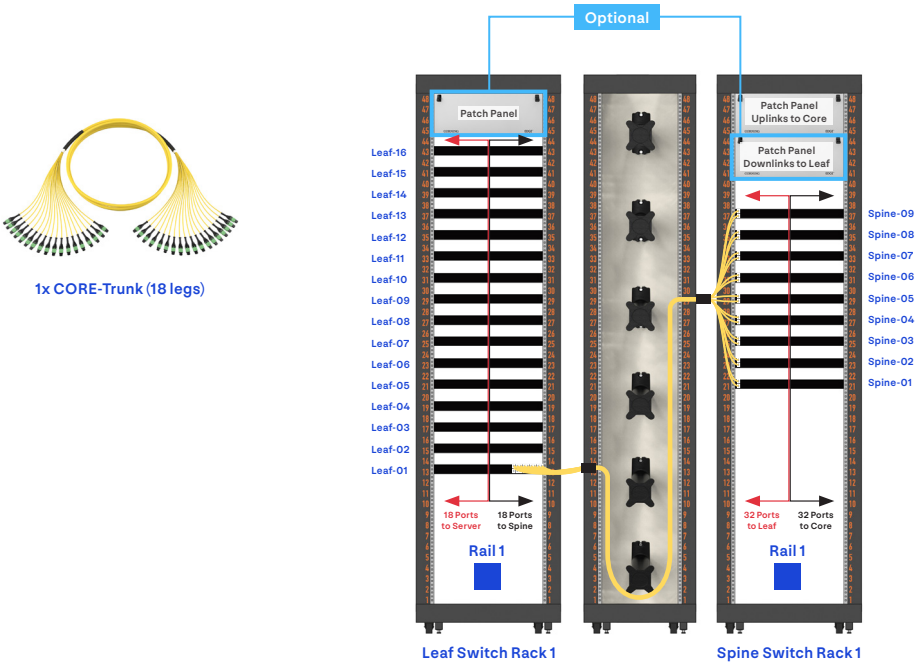


Figure 16. Level B – Leaf-to-Spine cabling utilizing Point-to-Point connection with a CORE-Trunk – Example based on 16 SU Cluster.

When using CORE-Trunks, a single 144-fiber trunk makes it possible to route the 18 ports from Leaf-01 of Rail 1 (Blue) to 2 ports on each Spine Switch within the Spine Switch Rack as described in Figure 17.

Each CORE-Trunk is designed to handle the 18 connections coming from Rail 1 (Blue), streamlining cabling and reducing complexity. The same principle is applied to the other 16 Leaf Switches within Rail 1 (Blue), with each Leaf Switch utilizing the same CORE-Trunk configuration to establish connectivity to the Spine Switch Rack.



As previously explained:

- One 144-fiber MPO-8/12 APC CORE-Trunk, representing the 18x ports from Leaf-01 of the Rail 1 (Blue) will connect to 2x ports in each Spine Switch in the Spine Switch Rack that brings together all Rail 1 (Blue) connections.
- Each CORE-Trunk has 18x connections from Rail 1 (Blue).
- Same logic is applied to the other 16x Leaf Switches.
- Patch Panels can be used if Structured Cabling is designed within the SU (connections to the Leaf Switches.) Structured Cabling is recommended when implementing connections to the Core Switch area within the DC.

Figure 17. Level B – CORE-Trunk Routing Example.

For improved cable management, Patch Panels can be used when implementing Structured Cabling within the Scalable Unit (SU). These panels provide an organized and scalable solution for Leaf Switch connections. Structured Cabling is especially recommended when routing connections to the centralized Core Switch area within the data center (DC), as it simplifies cable routing, enhances system organization, and facilitates troubleshooting.

The same cabling and mapping concept is applied for Rail 2, 3, and 4. In Figure 18, for this layout, we can see 4 Leaf Switch Racks, one per Rail; as well as the 4 Spine Switch Racks, one per Rail.

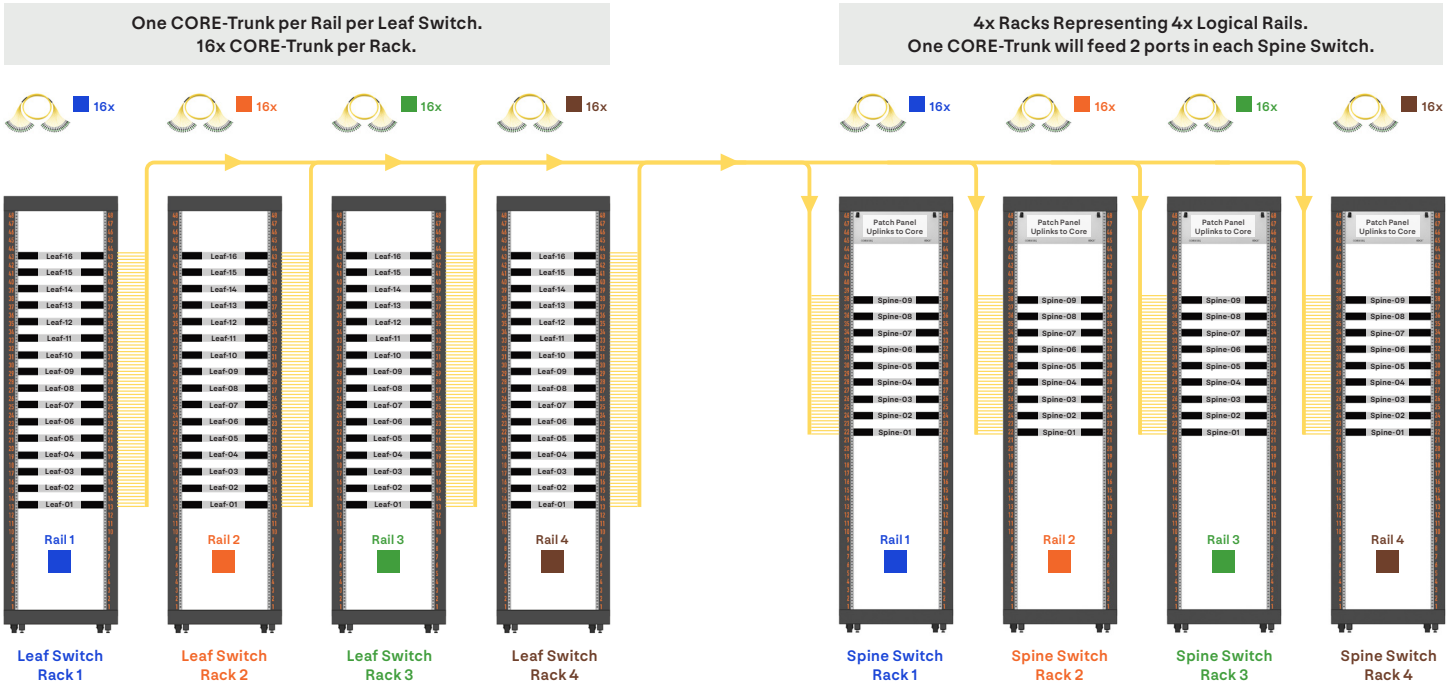


Figure 18. Level B – CORE-Trunk Routing to Spine Switches.

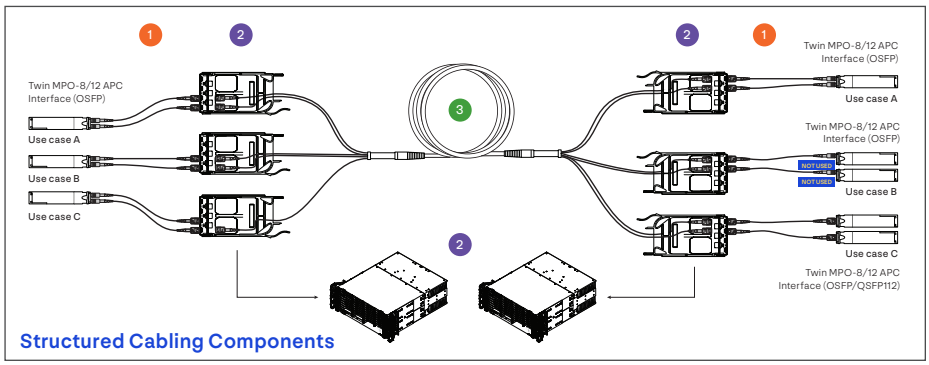
2.4.3. Level C – Spine-to-Core cabling

Spine-to-Core Switches are recommended to be cabled using Structured Cabling. In some specific custom designs, implementing Point-to-Point cabling at the Spine-to-Core connection level is also possible; this will depend on the physical location of the Core in relation to the Spine.

In Figure 19, Spine-to-Core cabling is implemented using Structured Cabling, with CORE-Trunks connecting active equipment and EDGE8® Trunks serving as the backbone. In a 16 SU Cluster, there are nine Core Groups distributed across 18 Core Racks, with each Core Rack housing 16 Core Switches, resulting in 32 Core Switches per Group.

In this example:

- For a 16x SU Cluster, there will be 9x Core Groups distributed across 18 Core Racks.
- Each Core Rack contains 16x Core Switches (32x per Group).
- Each Core Switch aggregates all Rails from all SUs.
- Each Core Switch will receive 1x connection from each Spine Switch within each SU. Note that the number of connections varies depending on the size of the cluster.
- Structured Cabling is the preferred option in this example, utilizing a Centralized Core.
- One 128-fiber MPO-8/12 APC CORE-Trunk, representing 16/32x uplink connections from a single Spine Switch, will connect to one port in each of the 16x Core Switches within a Core Rack. Each CORE-Trunk supports 16x connections.



Refer to Scenario 2, Table 3 for the component part numbers and descriptions.

Example for Use case A:

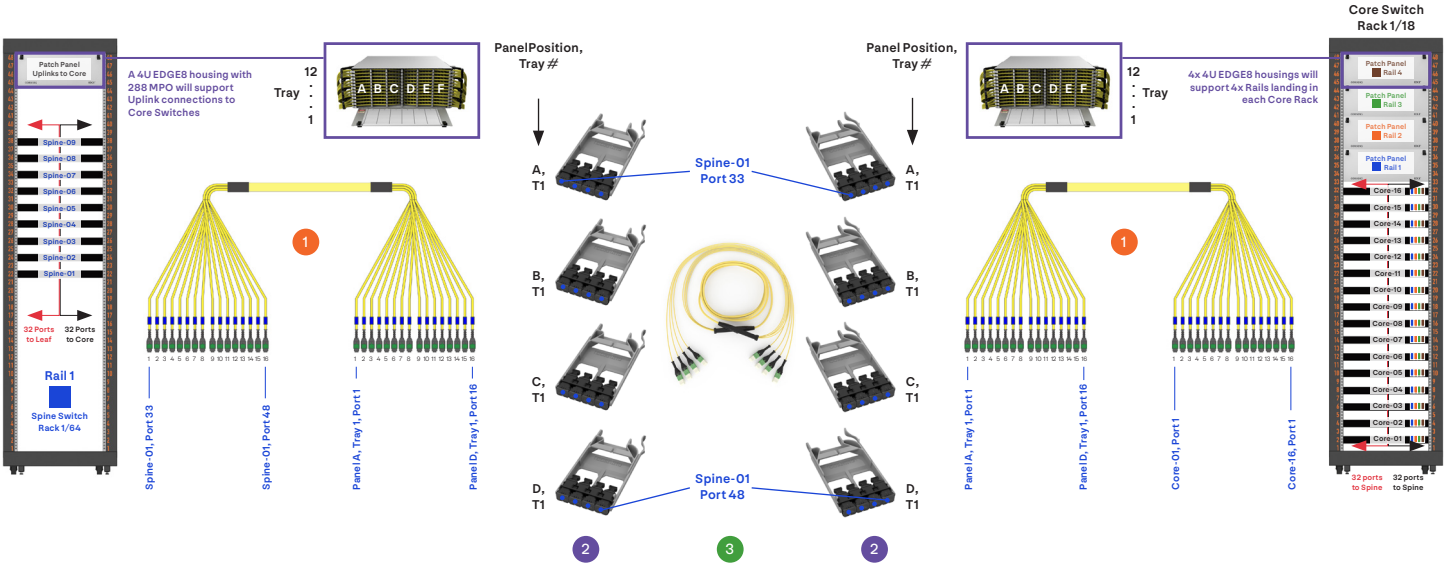
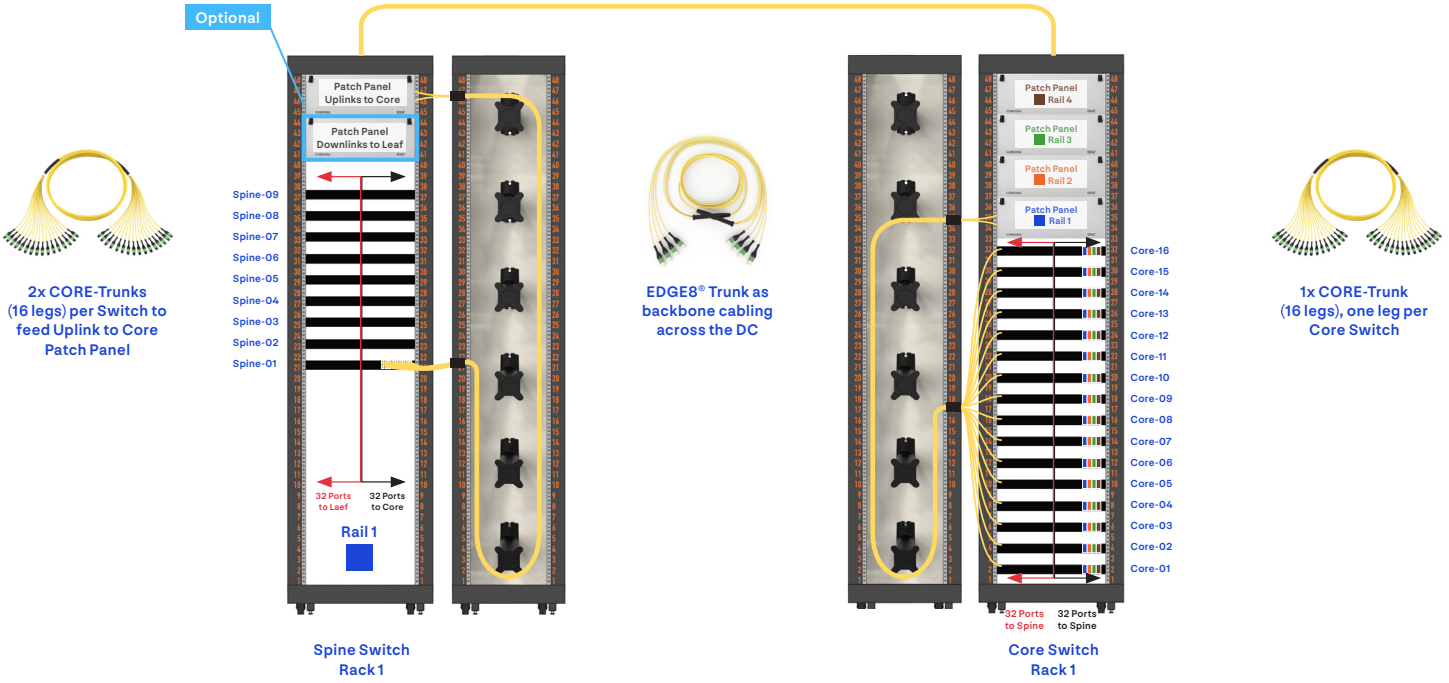


Figure 19. Level C – Spine-to-Core cabling utilizing Structured Cabling with CORE-Trunks to the active equipment and EDGE8 Trunks as a backbone – Example based on 16 SU Cluster.

Figure 20 illustrates the aggregation process, where each Core Switch consolidates connections from all Rails across all SUs and receives a single connection from each Spine Switch within each SU. The total number of connections varies based on the cluster size.

In this configuration, Structured Cabling is the preferred approach, leveraging a Centralized Core. A single 128-fiber MPO-8/12 APC CORE-Trunk facilitates 16 or 32 uplink connections from a Spine Switch and connects to one port on each of the 16 Core Switches within a Core Rack. Each CORE-Trunk is designed to support up to 16 connections.



Summary:

- Two 128-fiber MPO-8/12 APC Core-Trunks will bring connections of 32 ports from a single Spine Switch of the Rail 1 (Blue) into the Uplinks to Core Patch Panel at the top of the Spine Rack 1. A total of 18x CORE-Trunks will be required. As alternative, 288x traditional, individual jumpers can also be used for this part of the cabling.
- EDGE8 Trunks (Multifiber Assemblies Trunk) will support distributing the fibers across the DC towards multiple Patch Panels located at the Core Switch Racks.
- A total of 64x 128-fiber MPO-8/12 APC CORE-Trunks will be required to cable the 64 ports of the Core Switches located in each Core Switch Rack. As alternative, 1,024x traditional, individual jumpers can also be used for this part of the cabling.

Figure 20. Level C – CORE-Trunk and EDGE8 Trunk. Routing Example.

Figure 21 presents the overall layout of CORE-Trunks and EDGE8 Trunks for each POD, demonstrating their routing to the centralized Spine Switch Racks within the Structured Cabling framework.

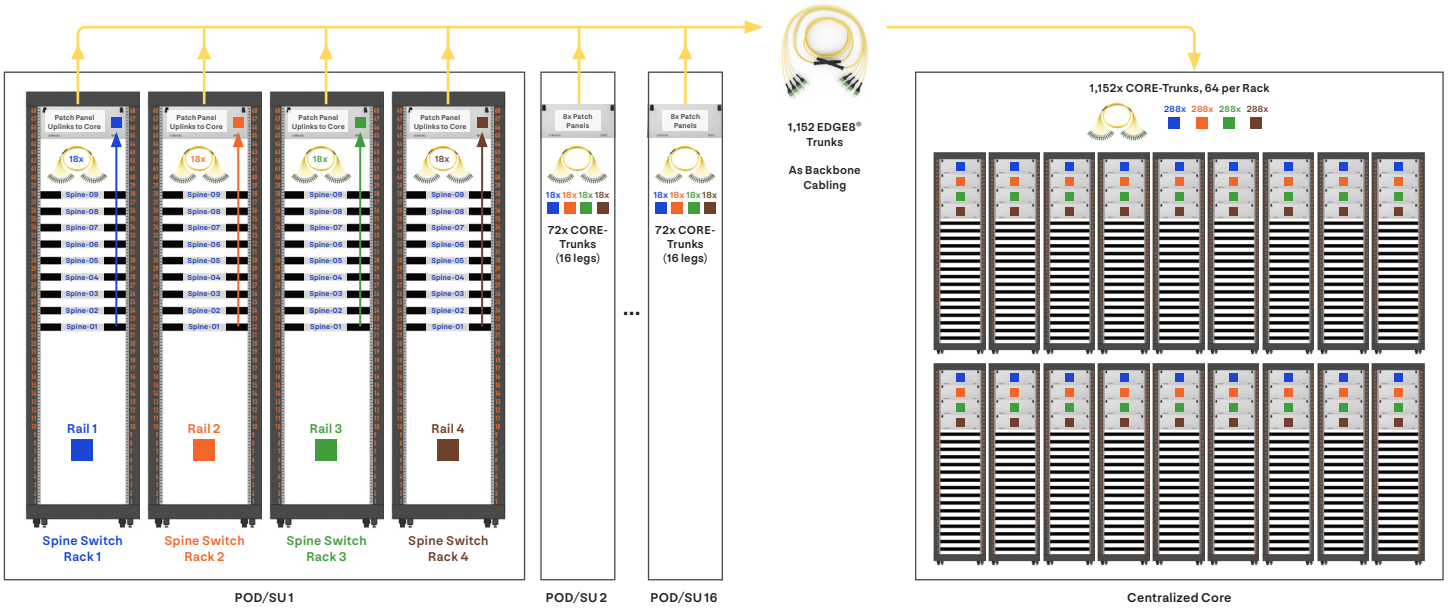


Figure 21. Level C – CORE-Trunks and EDGE8 Trunks in a Structured Cabling layout being routed to Spine Switches.

2.5. Multimode vs. Single-mode

The choice between Multimode and Single-mode fiber for the network will be determined by specific design requirements. Multimode fiber offers a maximum reach of up to 50 meters, making it primarily suitable for connections within the Scalable Unit, such as Server-to-Leaf and Leaf-to-Spine links. However, as Spine and Core Switches are often not physically located near each other, Single-mode fiber is recommended for this portion of the design due to its ability to support longer distances effectively, with a reach of up to 500 meters.

2.6. The Big Picture

Now that we have learned about the different Cluster sizes, as well as how cabling can be done between the active equipment in the Compute Fabric, let's visually summarize the components that can be used; these components will be dependent on the specific designs but would be mostly based on the different products and part numbers we have reviewed in this document. The following examples are based on Quantum-2 InfiniBand Switches but can also serve as design references for implementations using Quantum-3 InfiniBand or Spectrum-4 Ethernet switches.

2.6.1. Cabling a 1 Scalable Unit (SU) Cluster

Compute Component Count				InfiniBand Switch Counts			Cable Counts		
SU	NVL72 Racks	Node	GPU	Leaf	Spine	Core	Node-Leaf	Leaf-Spine	Spine-Core
1	16	288	1,152	64	36	N/A	1,152	1,152	N/A

As previously mentioned, the Scalable Unit (SU) serves as the foundational building block of a GPU Cluster. When referring to a 1 SU Cluster, two distinct approaches can be considered, as illustrated in Figure 22. By applying the different levels of cabling we reviewed, we can summarize the configurations as follows:

1. **Non-Scalable Cluster:** This setup (Figure 23) includes 64 Leaf Switches and 18 Spine Switches, but it lacks scalability, remaining limited to a Two-Tier design:

   - **1,152 MPO connections from Node to Leaf (Level A):** These connections can be implemented using Point-to-Point cabling with 64 CORE-Trunks of 144-fibers each, or 1,152 individual 8-fiber jumpers.

   - **1,152 MPO connections from Leaf to Spine (Level B):** These connections can also be implemented using Point-to-Point cabling with 64 CORE-Trunks of 144-fibers each, or 1,152 individual 8-fiber jumpers.

2. **Scalable Cluster:** In this configuration (Figure 24), a 1 SU consists of 64 Leaf Switches and 36 Spine Switches, with the ability to expand to 2 or more SUs by incorporating a Core Switch layer, transitioning to a Three-Tier design:

   - **1,152 MPO connections from Node to Leaf (Level A):** These connections can be implemented using Point-to-Point cabling with 64 CORE-Trunks of 144-fibers each, or 1,152 individual 8-fiber jumpers.

   - **1,152 MPO connections from Leaf to Spine (Level B):** These connections can also be implemented using Point-to-Point cabling with 64 CORE-Trunks of 144-fibers each, or 1,152 individual 8-fiber jumpers.

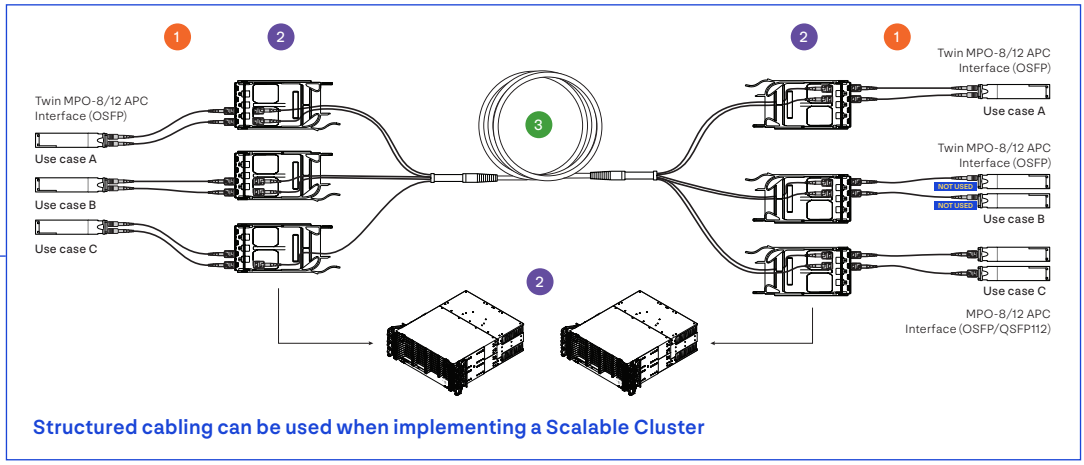
   - **Core Connections (Level C):** When a Core Switch layer is introduced, an additional 1,152 MPO connections will be required. Refer to Figures 25 to 32 for details on implementing connections to a Centralized Core Switch Area.

Cabling from GPU to Spine can utilize either 128 (144F) CORE-Trunks (64 from Server to Leaf + 64 from Leaf to Spine) per SU, instead of 2,304 (8F) individual jumpers (1,152 from Server to Leaf + 1,152 from Leaf to Spine), managing complexity within the Scalable Unit.

A **Server-to-Leaf cabling using either point-to-point jumpers or CORE-Trunks**

B **Leaf-to-Spine cabling using either point-to-point jumpers, CORE-Trunks or AOCs**

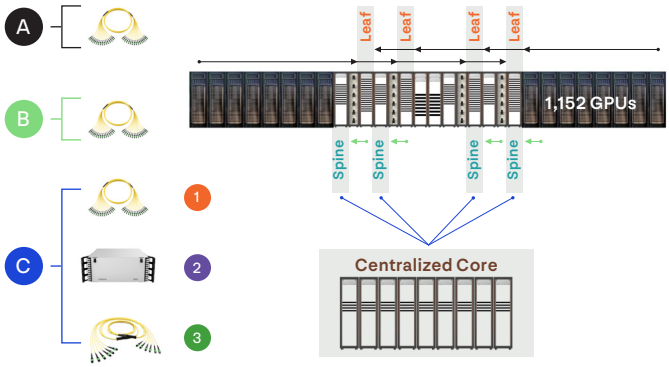
C **Spine-to-Core cabling using structured cabling**



Structured cabling can be used when implementing a Scalable Cluster

Scalable Cluster

- 1x SU
- 64 Leaf Switches
- 36 Spine Switches
- Cluster can grow to 2x or more SU by implementing a Core Switch Level
- Utilizes Structured Cabling for the Spine-Core connectivity



Non-Scalable Cluster

- 1x SU
- 64 Leaf Switches
- 18 Spine Switches
- Cluster is not scalable

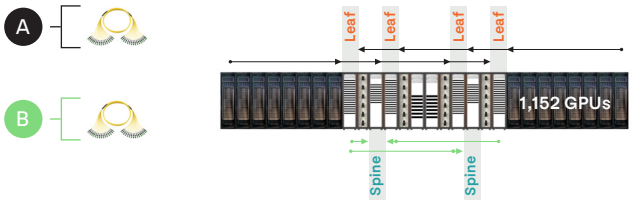


Figure 22. Cabling a 1 Scalable Unit (SU) Cluster.

In a Non-Scalable Cluster, each Leaf Switch receives 18 MPO connections from the servers, while each Spine Switch receives one MPO connection from each Leaf, resulting in a total of 64 MPO connections per Spine, as illustrated in Figure 23.

1,152 GPU Cluster utilizing a Two-Tier design, with 18 Spine Switches.

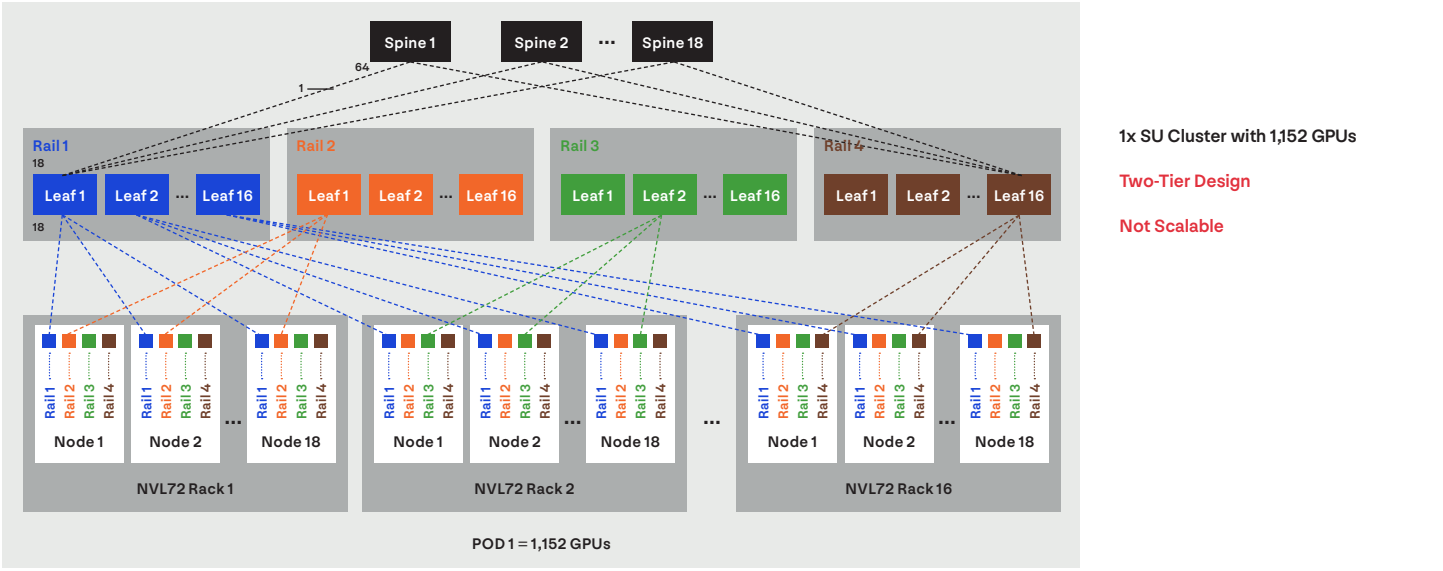


Figure 23. Compute Fabric for 1 SU (Two-Tier design, Not-Scalable) Cluster.

In a Scalable Cluster, each Leaf Switch receives 18 MPO connections from the servers, while each Spine Switch receives two MPO connections from each Leaf within the same Rail, resulting in a total of 32 MPO connections per Spine. Each Spine then forwards a certain number of connections to the Core Switches, depending on the size of the Cluster, as illustrated in Figure 24.

1,152 GPU Cluster utilizing a Two-Tier design, with 36 Spine Switches.

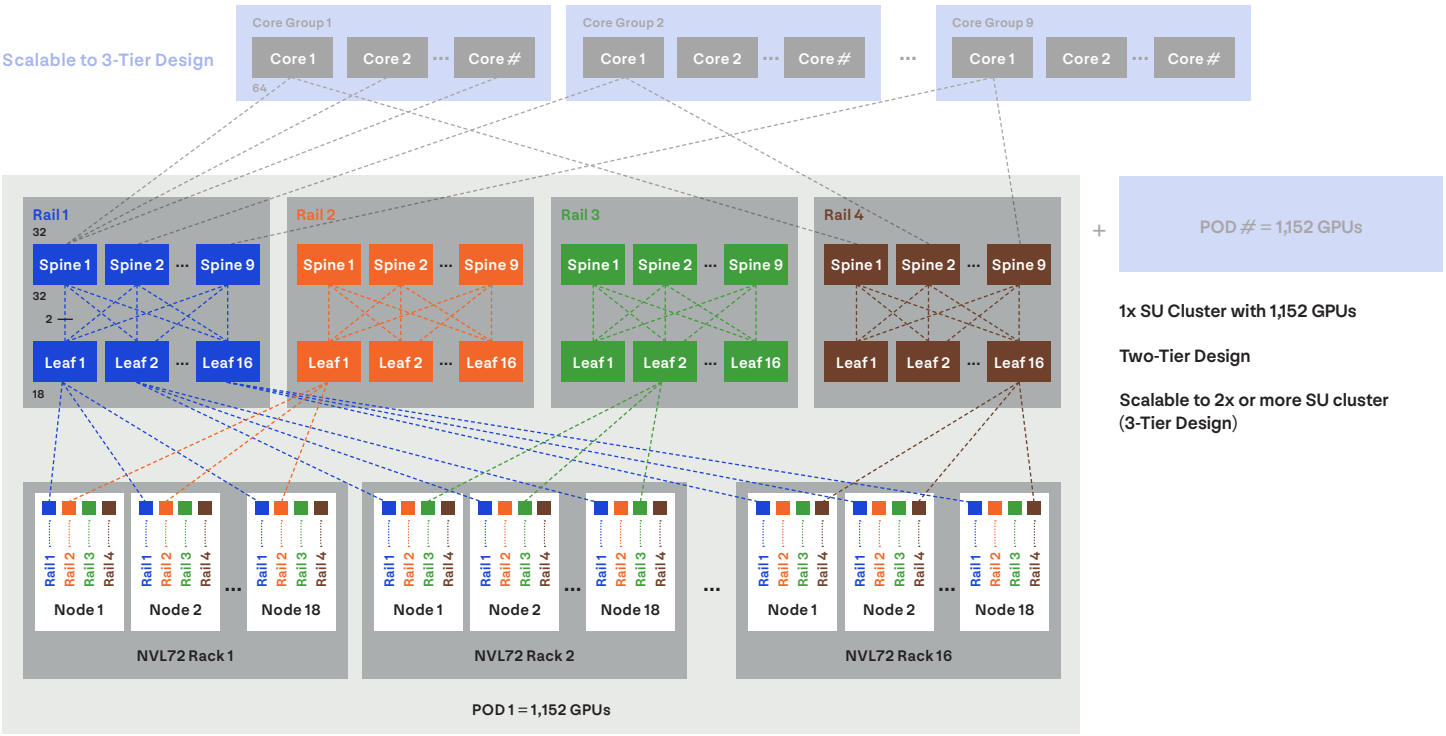


Figure 24. Compute Fabric for 1 SU (Scalable Version) Cluster.

2.6.2. Cabling a 2 Scalable Unit (SU) Cluster

Compute Component Count				InfiniBand Switch Counts			Cable Counts		
SU	NVL72 Racks	Node	GPU	Leaf	Spine	Core	Node-Leaf	Leaf-Spine	Spine-Core
2	32	576	2,304	128	72	36	2,304	2,304	2,304

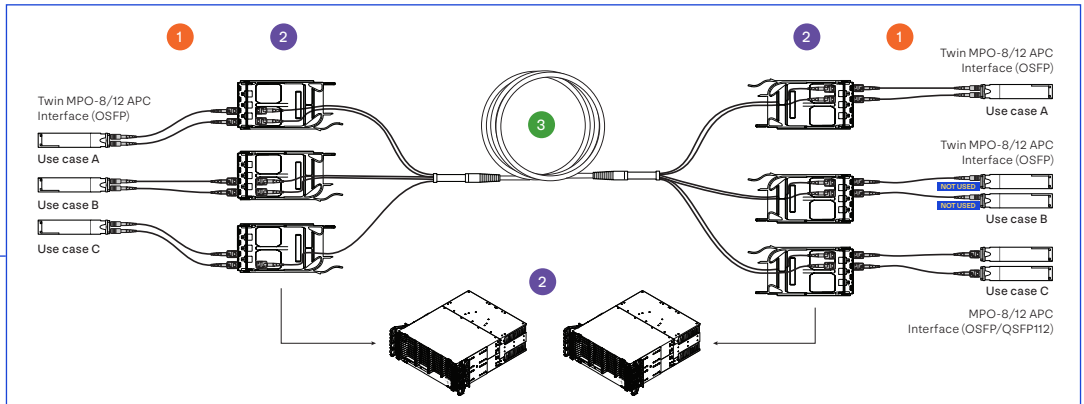
In Figure 25, cabling a 2 SU Cluster with 576 servers, 128 Leaf Switches, 72 Spine Switches and 36 Core Switches requires:

- 2,304 MPO connections from Node-to-Leaf
- 2,304 MPO connections from Leaf-to-Spine
- 2,304 MPO connections from Spine-to-Core

A **Server-to-Leaf cabling using either point-to-point jumpers or CORE-Trunks**

B **Leaf-to-Spine cabling using either point-to-point jumpers, CORE-Trunks or AOCs**

C **Spine-to-Core cabling using structured cabling**



Backbone cabling from Spine to Core will use 128 EDGE8® trunks (64 per SU, 144F), replacing 2,304 individual jumpers to simplify data center management.

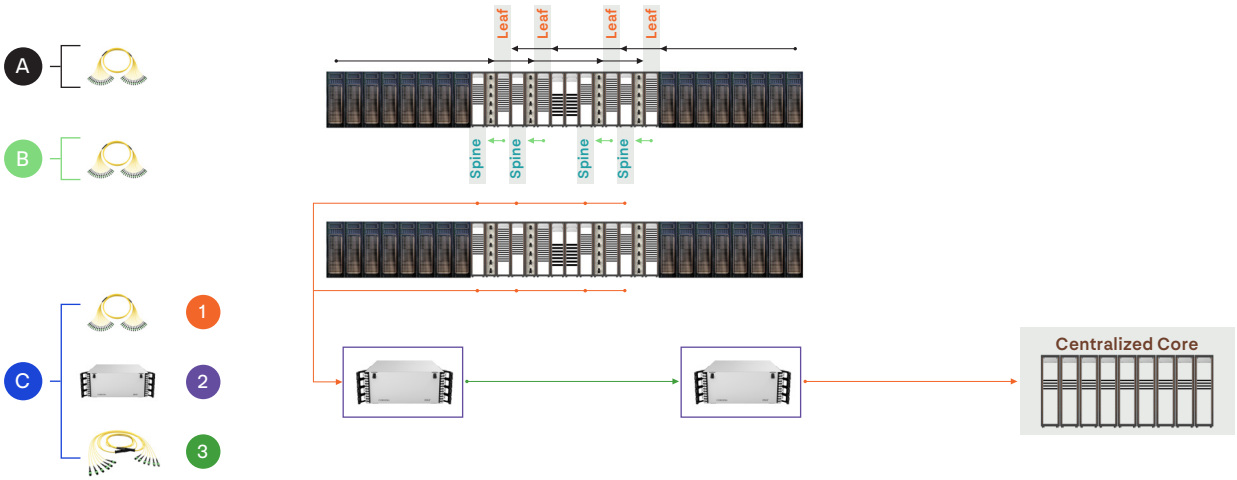


Figure 25. Cabling a 2 Scalable Unit (SU) Cluster.

By applying the different levels of cabling we reviewed, we can summarize the configurations as follows:

- Data Center Operators can utilize **Level A** within the Scalable Unit to perform Server-to-Leaf cabling using Point-to-Point cabling, either with jumpers or CORE-Trunks. This approach would require either 2,304 individual 8-fiber jumpers of varying lengths or 128 CORE-Trunks, each containing 144-fibers of varying lengths, per SU.

- At **Level B,** Leaf-to-Spine cabling can be performed using Point-to-Point cabling, either with jumpers or CORE-Trunks. This approach would require either 2,304 individual 8-fiber jumpers of varying lengths or 128 CORE-Trunks, each containing 144-fibers of varying lengths, per SU.

- At **Level C,** Spine-to-Core cabling is best performed using **Structured Cabling** as the preferred method, particularly for managing cables over longer distances (up to 500 meters). This approach is recommended as the Core Switch Area is typically centralized in a different area or data hall within the data center. **Alternatively, Point-to-Point Cabling** can be used; however, in this case, the use of CORE-Trunks (a total of 144 CORE-Trunks of 128-fibers up to 500 meters) is strongly advised to simplify cabling and minimize the complexity associated with managing individual jumpers. The specific details for each approach are as follows:

  - **(1) Connectivity from Spine Switch Rack to a Patch Panel:**
    This can be achieved using either:
    - 144 CORE-Trunks of 128-fibers; or
    - 2,304 individual 8-fiber patch cords.

  - **(2) Patch Panel at Spine Switch Racks:**
    High-density EDGE8® 4U Housings with adapter panels (see Annex 1) are required. A total of 8 housings, one for each Spine Switch Rack, will be needed.

  - **(3) Backbone Cabling:**
    EDGE8 Trunks will be used as backbone cabling between the Spine Switch Racks at the SU and the Core Switch Racks at the centralized Core area. A total of 128 EDGE8 Trunks with 144-fibers each, with lengths up to 500 meters, will be required.

  - **(2) Patch Panel at Core Switch Racks:**
    High-density EDGE8 4U Housings with adapter panels (see Annex 1) are required. A total of 8 housings will be needed at the Core Switch Rack area.

  - **(1) Connectivity from Patch Panel to Core Switch Rack:**
    This can be achieved using either:
    - 144 CORE-Trunks of 128-fibers; or
    - 2,304 individual 8-fiber jumpers.

  - **Summary for Level C:**
    - **(1) Number of 128-fiber CORE-Trunks:** 288 (144 from Spine to Patch Panel + 144 from Patch Panel to Core Switch Rack); **or Number of Individual Jumpers:** 4,608 (2,304 from Spine to Patch Panel + 2,304 from Patch Panel to Core Switch Rack).
    - **(2) Number of EDGE8 Housings with adapter panels:** 16 (8 for Spine Switch Racks + 8 for Core Switch Racks).
    - **(3) Number of 144-fiber EDGE8 Trunks:** 128 (used for backbone cabling).

In a 2 SU Cluster, each Leaf Switch receives 18 MPO connections from the Servers, while each Spine Switch receives two MPO connections from each Leaf within the same Rail. This results in a total of 32 MPO connections per Spine Switch. Each Spine Switch then forwards eight MPO connections to the Core Switches, as shown in Figure 26.

It is important to note that there are **9 Spine Switches per Rail,** which is why there are **9 Groups of Core Switches.** The distribution of connections follows this logic:

- **Core Group 1** receives all connections from Spine-01 Switch across all Rails.
- **Core Group 2** receives all connections from Spine-02 Switch across all Rails, and so on for subsequent groups.

For example, each Spine-01 Switch (see rack unit 23 in each Spine Rack in Figure 21), from each Rail forwards eight MPO connections to Core-01 Switch in Group 1. When multiplied by four Rails and two PODs, this results in a total of 64 incoming MPO connections to Core-01 Switch in Group 1.

2,304 GPU Cluster utilizing 9x Core Groups, each one with 4 Switches. In previous figure this is represented by 9 Core Racks with 4 Core Switches each.

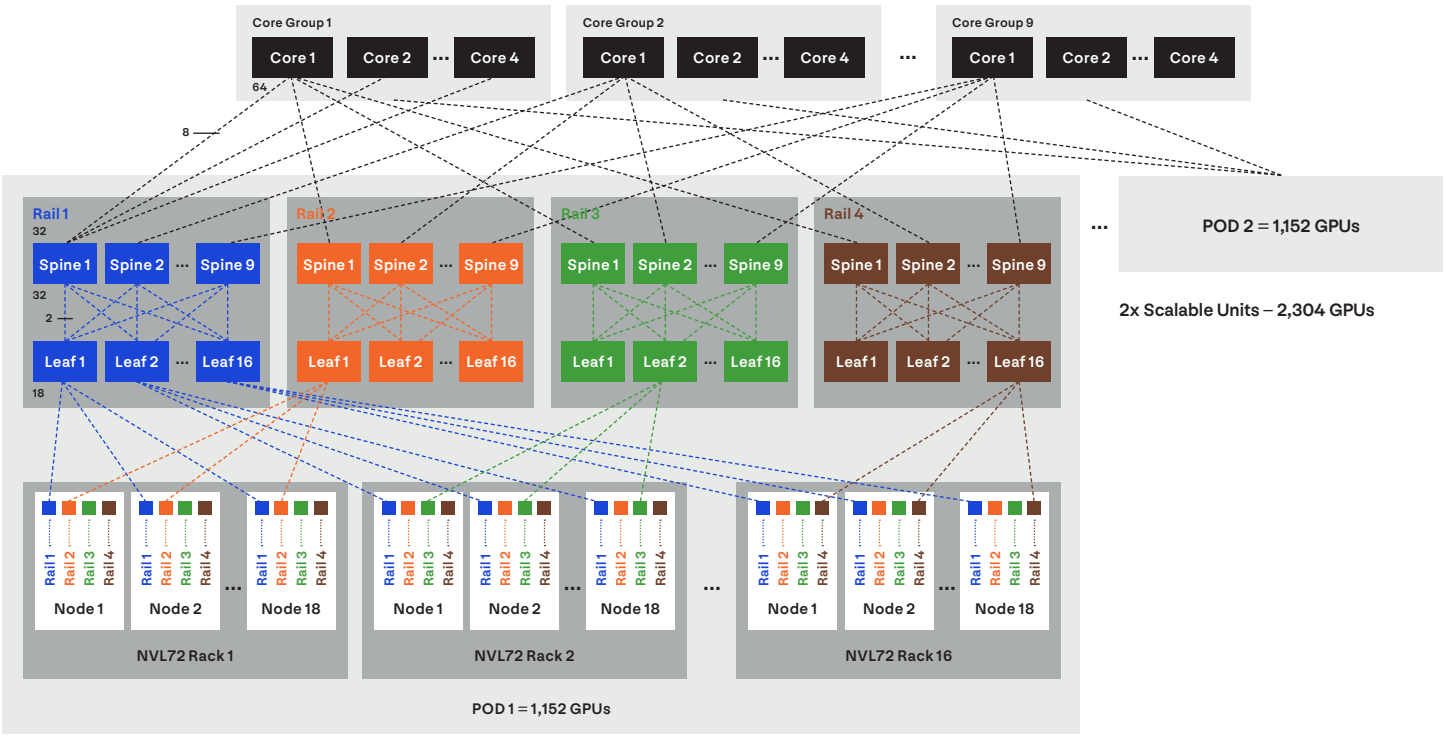


Figure 26. Compute Fabric for 2 SU Cluster.

2.6.3. Cabling a 4 Scalable Unit (SU) Cluster

Compute Component Count				InfiniBand Switch Counts			Cable Counts		
SU	NVL72 Racks	Node	GPU	Leaf	Spine	Core	Node-Leaf	Leaf-Spine	Spine-Core
4	64	1,152	4,608	256	144	72	4,608	4,608	4,608

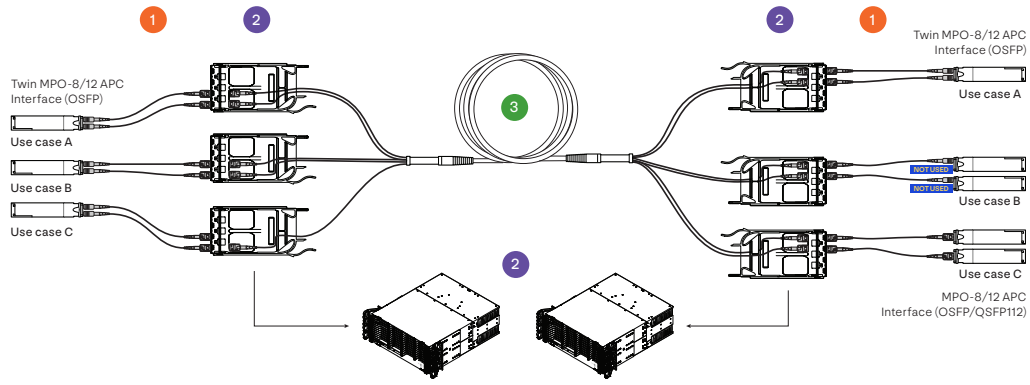
In Figure 27, cabling a 4 SU Cluster with 1,152 servers, 256 Leaf Switches, 144 Spine Switches and 72 Core Switches requires:

- 4,608 MPO connections from Node to Leaf
- 4,608 MPO connections from Leaf to Spine
- 4,608 MPO connections from Spine to Core

A **Server-to-Leaf cabling using either point-to-point jumpers or CORE-Trunks**

B **Leaf-to-Spine cabling using either point-to-point jumpers, CORE-Trunks or AOCs**

C **Spine-to-Core cabling using structured cabling**



Backbone cabling from Spine to Core will use 256 EDGE8® Trunks (64 per SU, 144F), replacing 4,608 individual jumpers to simplify DC management.

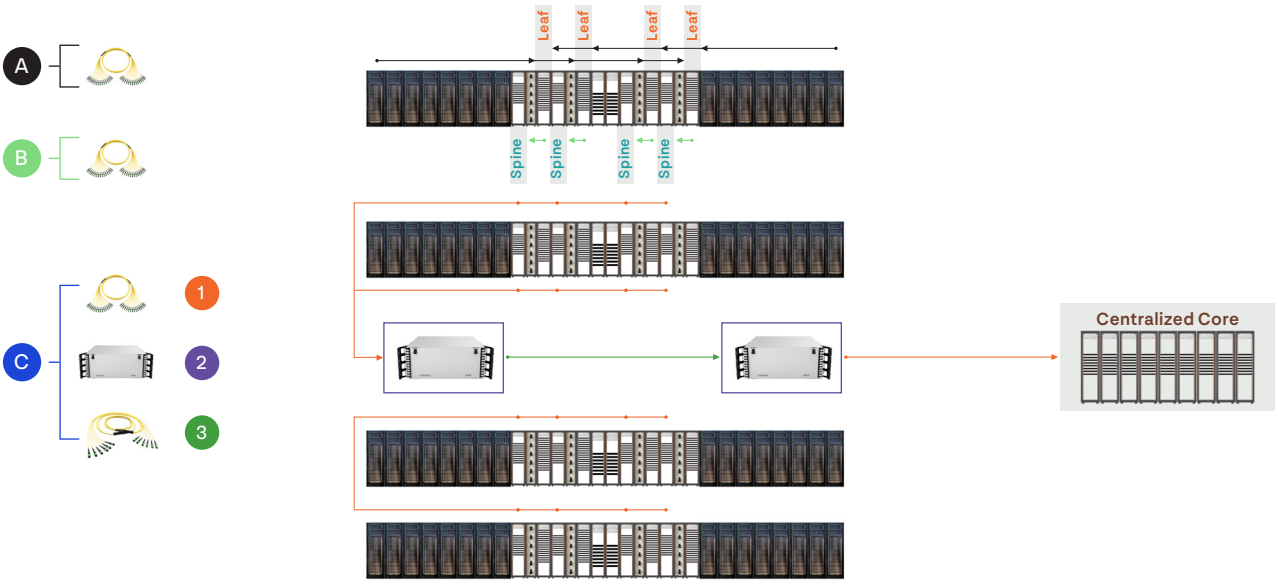


Figure 27. Cabling a 4 Scalable Unit (SU) Cluster.

By applying the different levels of cabling we reviewed, we can summarize the configurations as follows:

- Data Center Operators can utilize **Level A** within the Scalable Unit to perform Server-to-Leaf cabling using Point-to-Point cabling, either with jumpers or CORE-Trunks. This approach would require either 4,608 individual 8-fiber jumpers of varying lengths or 256 CORE-Trunks, each containing 144-fibers of varying lengths, per SU.

- At **Level B,** Leaf-to-Spine cabling can be performed using Point-to-Point cabling, either with jumpers or CORE-Trunks. This approach would require either 4,608 individual 8-fiber jumpers of varying lengths or 256 CORE-Trunks, each containing 144-fibers of varying lengths, per SU.

- At **Level C,** Spine-to-Core cabling is best performed using **Structured Cabling** as the preferred method, particularly for managing cables over longer distances (up to 500 meters). This approach is recommended as the Core Switch Area is typically centralized in a different area or data hall within the data center. **Alternatively, Point-to-Point Cabling** can be used; however, in this case, the use of CORE-Trunks (a total of 288 CORE-Trunks of 128-fibers up to 500 meters) is strongly advised to simplify cabling and minimize the complexity associated with managing individual jumpers. The specific details for each approach are as follows:

  - **(1) Connectivity from Spine Switch Rack to a Patch Panel:**
    This can be achieved using either:
    - 288 CORE-Trunks of 128-fibers; or
    - 4,608 individual 8-fiber jumpers.

  - **(2) Patch Panel at Spine Switch Racks:**
    High-density EDGE8® 4U Housings with adapter panels (see Annex 1) are required. A total of 16 housings, one for each Spine Switch Rack, will be needed.

  - **(3) Backbone Cabling:**
    EDGE8 Trunks will be used as backbone cabling between the Spine Switch Racks at the SU and the Core Switch Racks at the centralized Core area. A total of 256 EDGE8 Trunks with 144-fibers, each with lengths up to 500 meters, will be required.

  - **(2) Patch Panel at Core Switch Racks:**
    High-density EDGE8 4U Housings with adapter panels (see Annex 1) are required. A total of 16 housings will be needed at the Core Switch Rack area.

  - **(1) Connectivity from Patch Panel to Core Switch Rack:**
    This can be achieved using either:
    - 288 CORE-Trunks of 128-fibers; or
    - 4,608 individual 8-fiber jumpers.

  - **Summary for Level C:**
    - **(1) Number of 128-fiber CORE-Trunks:** 576 (288 from Spine to Patch Panel + 288 from Patch Panel to Core Switch Rack); **or Number of Individual Jumpers:** 9,216 (4,608 from Spine to Patch Panel + 4,608 from Patch Panel to Core Switch Rack).
    - **(2) Number of EDGE8 Housings with adapter panels:** 32 (16 for Spine Switch Racks + 16 for Core Switch Racks).
    - **(3) Number of 144-fiber EDGE8 Trunks:** 256 (used for backbone cabling).

In a 4 SU Cluster, each Leaf Switch receives 18 MPO connections from the Servers, while each Spine Switch receives 2 MPO connections from each Leaf within the same Rail. This results in a total of 32 MPO connections per Spine Switch. Each Spine Switch then forwards eight MPO connections to the Core Switches, as shown in Figure 28.

Since there are **9 Spine Switches per Rail,** this results in **9 Groups of Core Switches.** The connections are distributed according to the following logic:

- **Core Group 1** receives all connections from Spine-01 Switch across all Rails.
- **Core Group 2** receives all connections from Spine-02 Switch across all Rails, and so on for subsequent groups.

For example, each Spine-01 Switch (see rack unit 23 in each Spine Rack in Figure 21), from each Rail forwards four MPO connections to Core-01 Switch in Group 1. When multiplied by four Rails and four PODs, this results in a total of 64 incoming MPO connections to Core-01 Switch in Group 1.

4,608 GPU Cluster utilizing 9x Core Groups, each one with 8 Switches. In previous figure this is represented by 9 Core Racks with 8 Core Switches each.

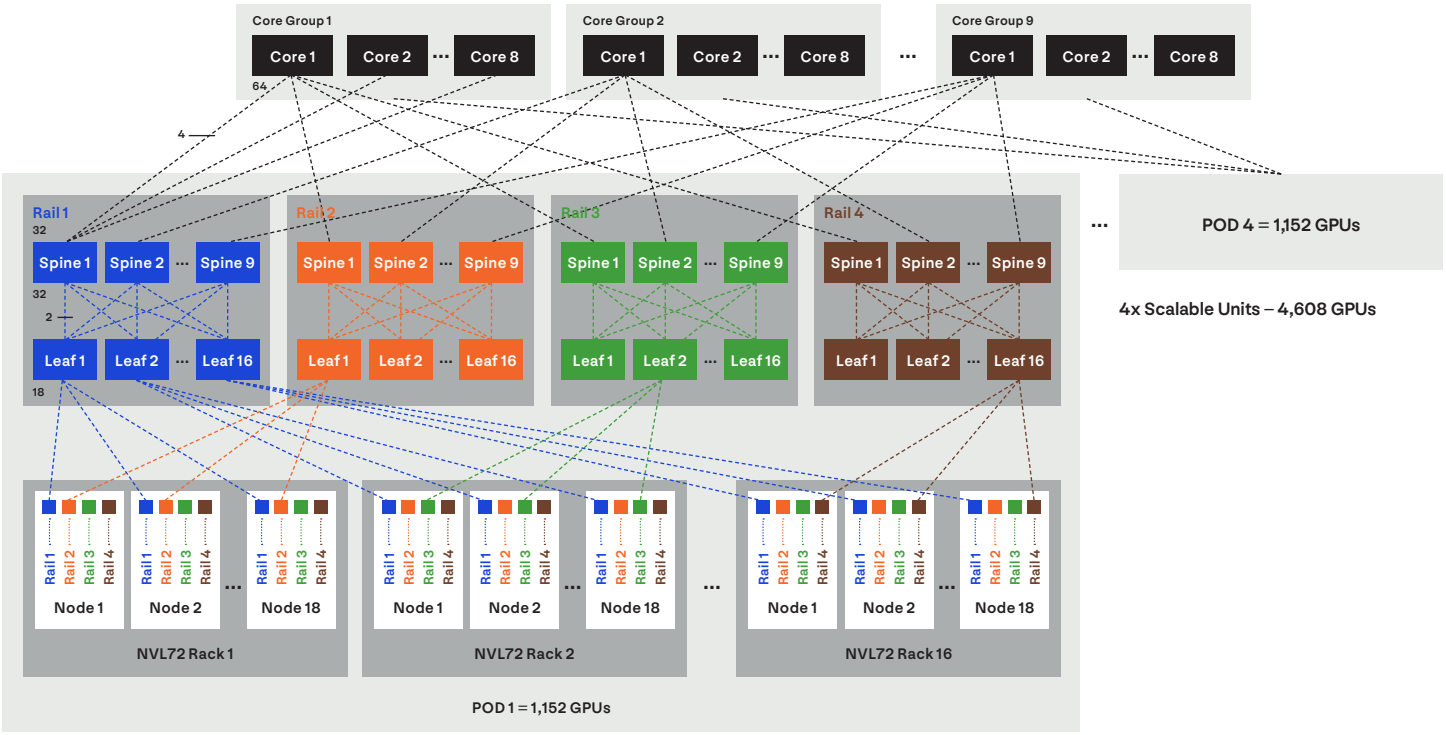


Figure 28. Compute Fabric for 4 SU Cluster.

2.6.4. Cabling an 8 Scalable Unit (SU) Cluster

Compute Component Count				InfiniBand Switch Counts			Cable Counts		
SU	NVL72 Racks	Node	GPU	Leaf	Spine	Core	Node-Leaf	Leaf-Spine	Spine-Core
8	128	2,304	9,216	512	288	144	9,216	9,216	9,216

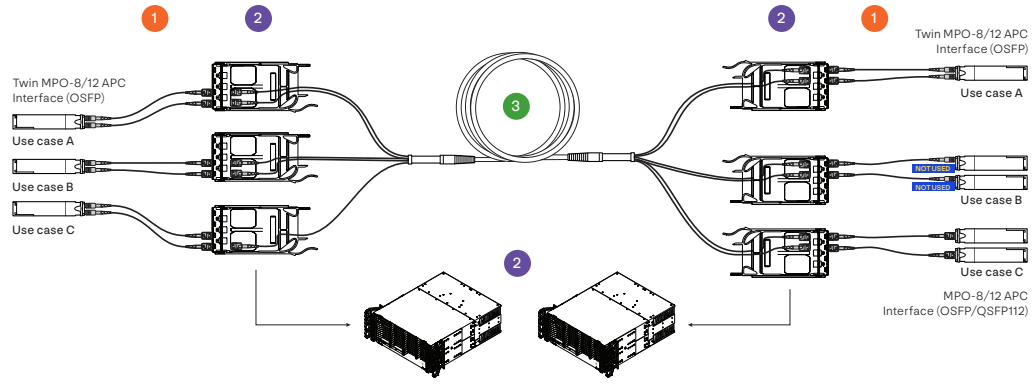
In Figure 29, cabling an 8 SU Cluster with 2,304 servers, 512 Leaf Switches, 288 Spine Switches and 144 Core Switches requires:

- 9,216 MPO connections from Node-to-Leaf
- 9,216 MPO connections from Leaf-to-Spine
- 9,216 MPO connections from Spine-to-Core

A **Server-to-Leaf cabling using either point-to-point jumpers or CORE-Trunks**

B **Leaf-to-Spine cabling using either point-to-point jumpers, CORE-Trunks or AOCs**

C **Spine-to-Core cabling using structured cabling**



Backbone cabling from Spine to Core will use 512 EDGE® Trunks (64 per SU, 144F), replacing 9,216 individual jumpers to simplify DC management.

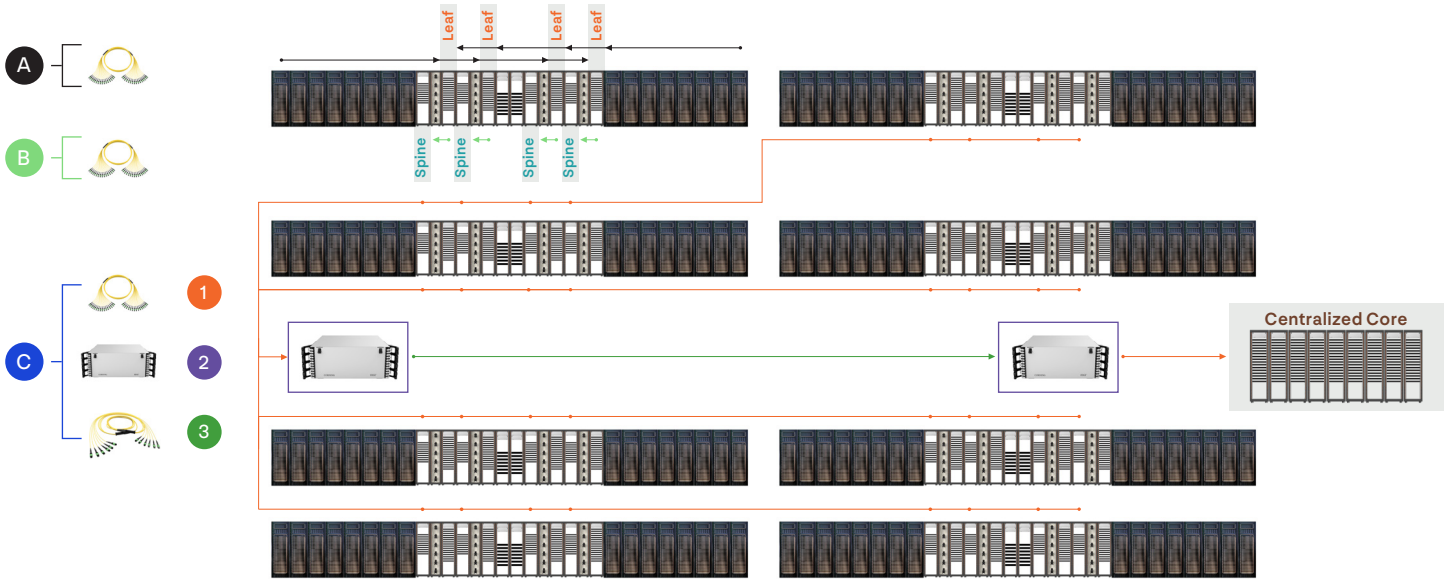


Figure 29. Cabling an 8 Scalable Unit (SU) Cluster.

By applying the different levels of cabling we reviewed, we can summarize the configurations as follows:

- Data Center Operators can utilize **Level A** within the Scalable Unit to perform Server-to-Leaf cabling using Point-to-Point cabling, either with jumpers or CORE-Trunks. This approach would require either 9,216 individual 8-fiber jumpers of varying lengths or 512 CORE-Trunks, each containing 144-fibers of varying lengths, per SU.

- At **Level B,** Leaf-to-Spine cabling can be performed using Point-to-Point cabling, either with jumpers or CORE-Trunks. This approach would require either 9,216 individual 8-fiber jumpers of varying lengths or 512 CORE-Trunks, each containing 144-fibers of varying lengths, per SU.

- At **Level C,** Spine-to-Core cabling is best performed using **Structured Cabling** as the preferred method, particularly for managing cables over longer distances (up to 500 meters). This approach is recommended as the Core Switch Area is typically centralized in a different area or data hall within the data center. **Alternatively, Point-to-Point Cabling** can be used; however, in this case, the use of CORE-Trunks (a total of 576 CORE-Trunks of 128-fibers up to 500 meters) is strongly advised to simplify cabling and minimize the complexity associated with managing individual jumpers. The specific details for each approach are as follows:

  - **(1) Connectivity from Spine Switch Rack to a Patch Panel:**
    This can be achieved using either:
    - 576 CORE-Trunks of 128-fibers; or
    - 9,216 individual 8-fiber jumpers.

  - **(2) Patch Panel at Spine Switch Racks:**
    High-density EDGE8® 4U Housings with adapter panels (see Annex 1) are required. A total of 32 housings, one for each Spine Switch Rack, will be needed.

  - **(3) Backbone Cabling:**
    EDGE8 Trunks will be used as backbone cabling between the Spine Switch Racks at the SU and the Core Switch Racks at the centralized Core area. A total of 512 EDGE8 Trunks with 144-fibers, each with lengths up to 500 meters, will be required.

  - **(2) Patch Panel at Core Switch Racks:**
    High-density EDGE8 4U Housings with adapter panels (see Annex 1) are required. A total of 32 housings will be needed at the Core Switch Rack area.

  - **(1) Connectivity from Patch Panel to Core Switch Rack:**
    This can be achieved using either:
    - 576 CORE-Trunks of 128-fibers; or
    - 9,216 individual 8-fiber jumpers.

  - **Summary for Level C:**
    - **(1) Number of 128-fiber CORE-Trunks:** 1,152 (576 from Spine to Patch Panel + 576 from Patch Panel to Core Switch Rack); **or Number of Individual Jumpers:** 18,432 (9,216 from Spine to Patch Panel + 9,216 from Patch Panel to Core Switch Rack).
    - **(2) Number of EDGE8 Housings with adapter panels:** 64 (32 for Spine Switch Racks + 32 for Core Switch Racks).
    - **(3) Number of 144-fiber EDGE8 Trunks:** 512 (used for backbone cabling).

In an 8 SU Cluster, each Leaf Switch receives 18 MPO connections from the Servers, while each Spine Switch receives 2 MPO connections from each Leaf within the same Rail. This results in a total of 32 MPO connections per Spine Switch. Each Spine Switch then forwards eight MPO connections to the Core Switches, as shown in Figure 30.

Since there are **9 Spine Switches per Rail,** this results in **9 Groups of Core Switches.** The connections are distributed according to the following logic:

- **Core Group 1** receives all connections from Spine-01 Switch across all Rails.
- **Core Group 2** receives all connections from Spine-02 Switch across all Rails, and so on for subsequent groups.

For example, each Spine-01 Switch (see rack unit 23 in each Spine Rack in Figure 21), from each Rail forwards two MPO connections to Core-01 Switch in Group 1. When multiplied by four Rails and eight PODs, this results in a total of 64 incoming MPO connections to Core-01 Switch in Group 1.

9,216 GPU Cluster utilizing 9x Core Groups, each one with 16 Switches. In previous figure this is represented by 9 Core Racks with 16 Core Switches each.

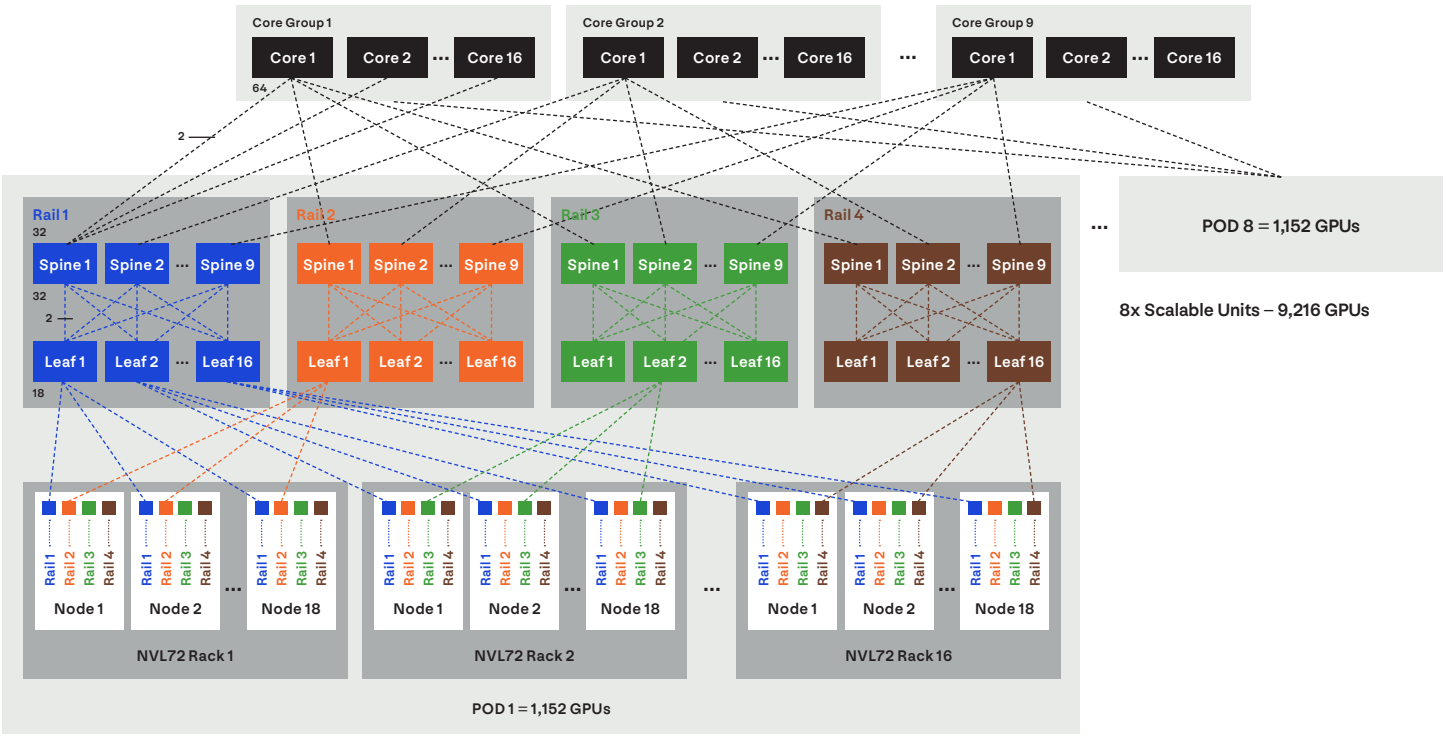


Figure 30. Compute Fabric for 8 SU Cluster.

2.6.5. Cabling a 16 Scalable Unit (SU) Cluster

Compute Component Count				InfiniBand Switch Counts			Cable Counts		
SU	NVL72 Racks	Node	GPU	Leaf	Spine	Core	Node-Leaf	Leaf-Spine	Spine-Core
16	256	4,608	18,432	1,024	576	288	18,432	18,432	18,432

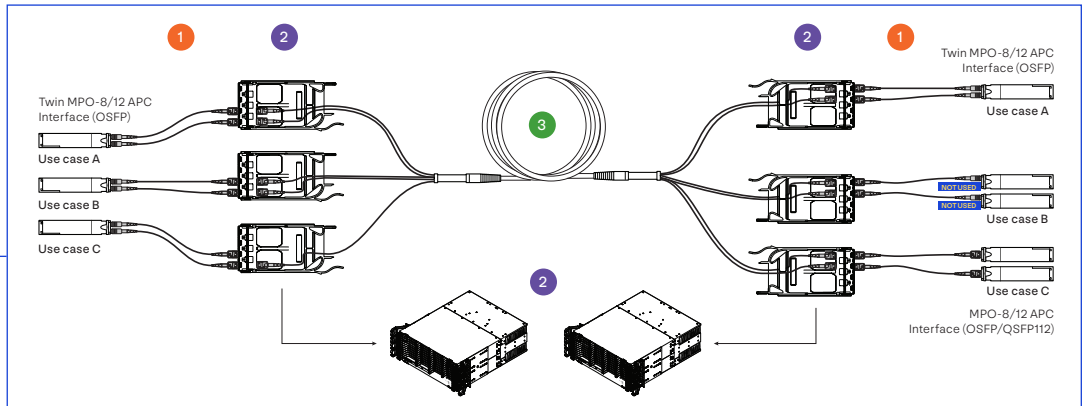
In Figure 31, cabling a 16 SU Cluster with 4,608 servers, 1,024 Leaf Switches, 576 Spine Switches and 288 Core Switches requires:

- 18,432 MPO connections from Node-to-Leaf
- 18,432 MPO connections from Leaf-to-Spine
- 18,432 MPO connections from Spine-to-Core

A **Server-to-Leaf cabling using either point-to-point jumpers or CORE-Trunks**

B **Leaf-to-Spine cabling using either point-to-point jumpers, CORE-Trunks or AOCs**

C **Spine-to-Core cabling using structured cabling**



Backbone cabling from Spine to Core will use 1,024 EDGE8 Trunks (64 per SU, 144F), replacing 18,432 individual jumpers to simplify DC management.

Figure 29. Cabling an 8 Scalable Unit (SU) Cluster.

By applying the different levels of cabling we reviewed, we can summarize the configurations as follows:

- Data Center Operators can utilize **Level A** within the Scalable Unit to perform Server-to-Leaf cabling using Point-to-Point cabling, either with jumpers or CORE-Trunks. This approach would require either 18,432 individual 8-fiber jumpers of varying lengths or 1,024 CORE-Trunks, each containing 144-fibers of varying lengths, per SU.

- At **Level B,** Leaf-to-Spine cabling can be performed using Point-to-Point cabling, either with jumpers or CORE-Trunks. This approach would require either 18,432 individual 8-fiber jumpers of varying lengths or 1,024 CORE-Trunks, each containing 144-fibers of varying lengths, per SU.

- At **Level C,** Spine-to-Core cabling is best performed using **Structured Cabling** as the preferred method, particularly for managing cables over longer distances (up to 500 meters). This approach is recommended as the Core Switch Area is typically centralized in a different area or data hall within the data center. **Alternatively, Point-to-Point Cabling** can be used; however, in this case, the use of CORE-Trunks (a total of 1,152 CORE-Trunks of 128-fibers up to 500 meters) is strongly advised to simplify cabling and minimize the complexity associated with managing individual jumpers. The specific details for each approach are as follows:

  - **(1) Connectivity from Spine Switch Rack to a Patch Panel:**
    This can be achieved using either:
    - 1,152 CORE-Trunks of 128-fibers; or
    - 18,432 individual 8-fiber jumpers.

  - **(2) Patch Panel at Spine Switch Racks:**
    High-density EDGE8® 4U Housings with adapter panels (see Annex 1) are required. A total of 64 housings, one for each Spine Switch Rack, will be needed.

  - **(3) Backbone Cabling:**
    EDGE8 Trunks will be used as backbone cabling between the Spine Switch Racks at the SU and the Core Switch Racks at the centralized Core area. A total of 1,024 EDGE8 Trunks with 144-fibers, each with lengths up to 500 meters, will be required.

  - **(2) Patch Panel at Core Switch Racks:**
    High-density EDGE8 4U Housings with adapter panels (see Annex 1) are required. A total of 64 housings will be needed at the Core Switch Rack area.

  - **(1) Connectivity from Patch Panel to Core Switch Rack:**
    This can be achieved using either:
    - 1,152 CORE-Trunks of 128-fibers; or
    - 18,432 individual 8-fiber jumpers.

  - **Summary for Level C:**
    - **(1) Number of 128-fiber CORE-Trunks:** 2,304 (1,152 from Spine to Patch Panel + 1,152 from Patch Panel to Core Switch Rack); **or Number of Individual Jumpers:** 36,864 (18,432 from Spine to Patch Panel + 18,432 from Patch Panel to Core Switch Rack).
    - **(2) Number of EDGE8 Housings with adapter panels:** 128 (64 for Spine Switch Racks + 64 for Core Switch Racks).
    - **(3) Number of 144-fiber EDGE8 Trunks:** 1,024 (used for backbone cabling).

In an 8 SU Cluster, each Leaf Switch receives 18 MPO connections from the Servers, while each Spine Switch receives 2 MPO connections from each Leaf within the same Rail. This results in a total of 32 MPO connections per Spine Switch. Each Spine Switch then forwards eight MPO connections to the Core Switches, as shown in Figure 30. MPO connections to Core-01 Switch in Group 1.

Since there are **9 Spine Switches per Rail,** this results in **9 Groups of Core Switches.** The connections are distributed according to the following logic:

- **Core Group 1** receives all connections from Spine-01 Switch across all Rails.
- **Core Group 2** receives all connections from Spine-02 Switch across all Rails, and so on for subsequent groups.

For example, each Spine-01 Switch (see rack unit 23 in each Spine Rack in Figure 21), from each Rail forwards two MPO connections to Core-01 Switch in Group 1. When multiplied by four Rails and eight PODs, this results in a total of 64 incoming MPO connections to Core-01 Switch in Group 1.

18,432 GPU Cluster utilizing 9x Core Groups, each one with 32 Switches. In previous figure this is represented by 18 Core Racks with 16 Core Switches each.

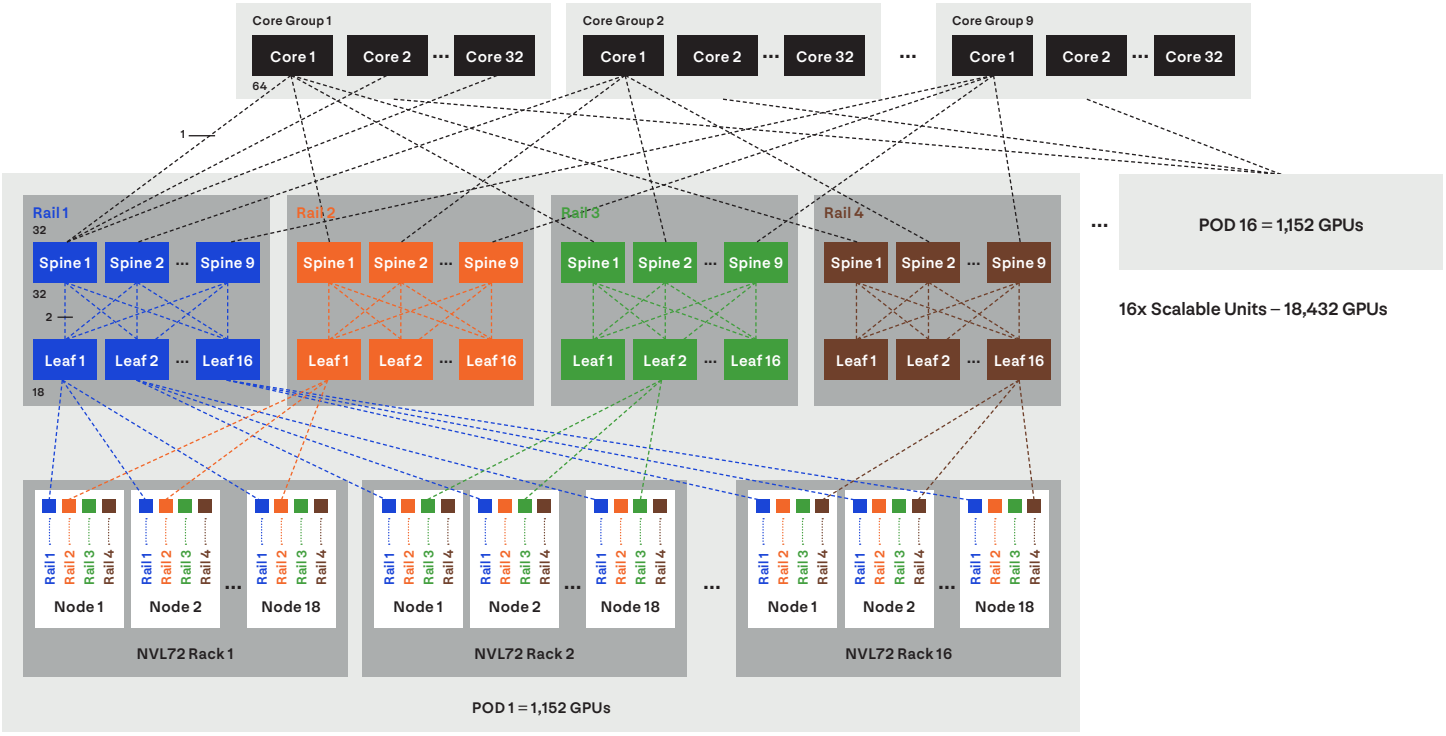


Figure 32. Compute Fabric for 16 SU Cluster.

2.7. NVL72 GB300 Clusters

So far, we have reviewed and understood the architecture and deployment of GB200 clusters. Now, let's explore the key aspects of GB300 clusters, focusing on how they differ and build upon the GB200 systems.

The primary difference lies in the GPU connectivity and data rates:

- GB200 clusters use NVIDIA Grace-Blackwell GPUs with 400G NDR connectivity, optimized for single-plane topologies.
- GB300 clusters feature the latest NVIDIA Grace-Blackwell GPUs with 800G XDR connectivity, enabling dual-plane and quad-plane topologies for higher scalability and bandwidth.

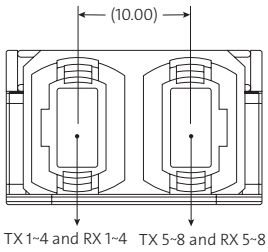
This section will explore, at a high level, the key features, network designs, and modular Scalable Units (SUs) that enable the deployment of GB300 clusters. Details on Ethernet and InfiniBand implementations are covered in subsections 2.7.1. and 2.7.2., respectively.

2.7.1. NVL72 GB300 Ethernet

The GB300 Ethernet Compute Fabric is designed to provide high-performance connectivity for GPUs in NVIDIA NVL72 systems, utilizing Spectrum-4 Ethernet switches for scalability, low latency, and non-blocking communication. Here's what needs to be considered to deploy this architecture:

- **Modular Scalable Units (SUs)**
  - Each Scalable Unit (SU) consists of **two GB300 NVL72 racks,** with each rack containing 72 GPUs (see Figure 33).
  - SUs are designed to enable rapid deployment and scale seamlessly for larger configurations.

Ethernet Optical Interface

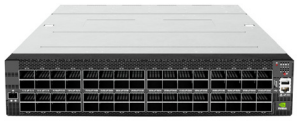


- **Server:** 4x GPUs, each with 2x MPO (2x400G)
- **Switch:** 2x MPO (2x400G)
- **Topology design:** Dual-plane or Quad-plane with Shuffle box

Spectrum-4 SN5000 Series Ethernet

SN5600

- 2U switch with 64 OSFP (twin-MPO) ports
- Total of 128 ports 400G over 64 OSFPs



2U SN5600 (Ethernet) = 128 Ports 400G

SN5400

- 2U switch with 64 QSFP-DD (single-MPO) ports
- Total of 64 ports 400G

2U SN5400 (Ethernet) = 64 Ports 400G

Compute Component Count				Ethernet Switch Counts		Link Count		Total Individual Cables
SU	NVL72 Racks	Node	GPU	Leaf	Spine	Node-Leaf	Leaf-Spine	
1	2	36	144	8	4	288	288	576
2	4	72	288	16	4	576	576	1,152
3	6	108	432	24	8	864	864	1,728
4	8	144	576	32	12	1,152	1,152	2,304
5	10	180	720	40	12	1,440	1,440	2,880
6	12	216	864	48	18	1,728	1,728	3,456
7	14	252	1,008	56	18	2,016	2,016	4,032
8	16	288	1,152	64	18	2,304	2,304	4,608
9	18	324	1,298	72	24	2,592	2,592	5,184
10	20	360	1,440	80	24	2,880	2,880	5,760
11	22	396	1,584	88	36	3,168	3,168	6,336
12	24	432	1,728	96	36	3,456	3,456	6,912
13	26	468	1,872	104	36	3,744	3,744	7,488
14	28	504	2,016	112	36	4,032	4,032	8,064
15	30	540	2,160	120	36	4,320	4,320	8,640
16	32	576	2,304	128	36	4,608	4,608	9,216
17	34	612	2,448	136	72	4,896	4,896	9,792
18	36	648	2,592	144	72	5,184	5,184	10,368
19	38	684	2,736	152	72	5,472	5,472	10,944
20	40	720	2,880	160	72	5,760	5,760	11,520
21	42	756	3,024	168	72	6,048	6,048	12,096
22	44	792	3,168	176	72	6,336	6,336	12,672
23	46	828	3,312	184	72	6,624	6,624	13,248
24	48	864	3,456	192	72	6,912	6,912	13,824
25	50	900	3,600	200	72	7,200	7,200	14,400
26	52	936	3,744	208	72	7,488	7,488	14,976
27	54	972	3,888	216	72	7,776	7,776	15,552
28	56	1,008	4,032	224	72	8,064	8,064	16,128
29	58	1,044	4,176	232	72	8,352	8,352	16,704
30	60	1,080	4,320	240	72	8,640	8,640	17,280
31	62	1,116	4,464	248	72	8,928	8,928	17,856
32	64	1,152	4,608	256	72	9,216	9,216	18,432

Figure 33. Detailed GB300 Ethernet Cluster sizes and component counts with two-tier dual-plane design.

- **Dual-Plane or Quad-Plane Network Design**

  - **Dual-Plane Topology:**
    - Each GPU connects to two independent planes for load balancing and redundancy.
    - Based on the number of GPUs being deployed, a **Two-Tier (Leaf-Spine)** or **Three-Tier architecture (Leaf-Spine-SuperSpine)** can be implemented. See Figures 34 and 35 as reference.

  - **Quad-Plane Topology (alternative to Three-Tier architecture for higher scalability):**
    - Represents an alternative to the Three-Tier architecture by flattening the network to a Two-Tier design (removing the SuperSpine). However, it introduces additional planes to further increase bandwidth and reduce congestion.
    - Quad-Plane splits the network into 4 planes of 200G, doubling the switch radix (from 128 ports at 400G to 256 ports at 200G), which allows for more Scalable Units and higher GPU density, the GPU remains agnostic to the Quad-Plane setup, as it continues to see an 800G interface for communication within the cluster.
    - Shuffle Boxes are required to manage the cabling complexity in Quad-Plane designs, allowing a large number of GPUs to be deployed using a **Two-Tier (Leaf-Spine)** architecture (Figure 36). The Shuffle Box can be placed between the Server and the Leaf Switch or between the Leaf Switch and the Spine Switch.

4,608 GPU Cluster, 2x 400G, connected rails, disconnected planes.

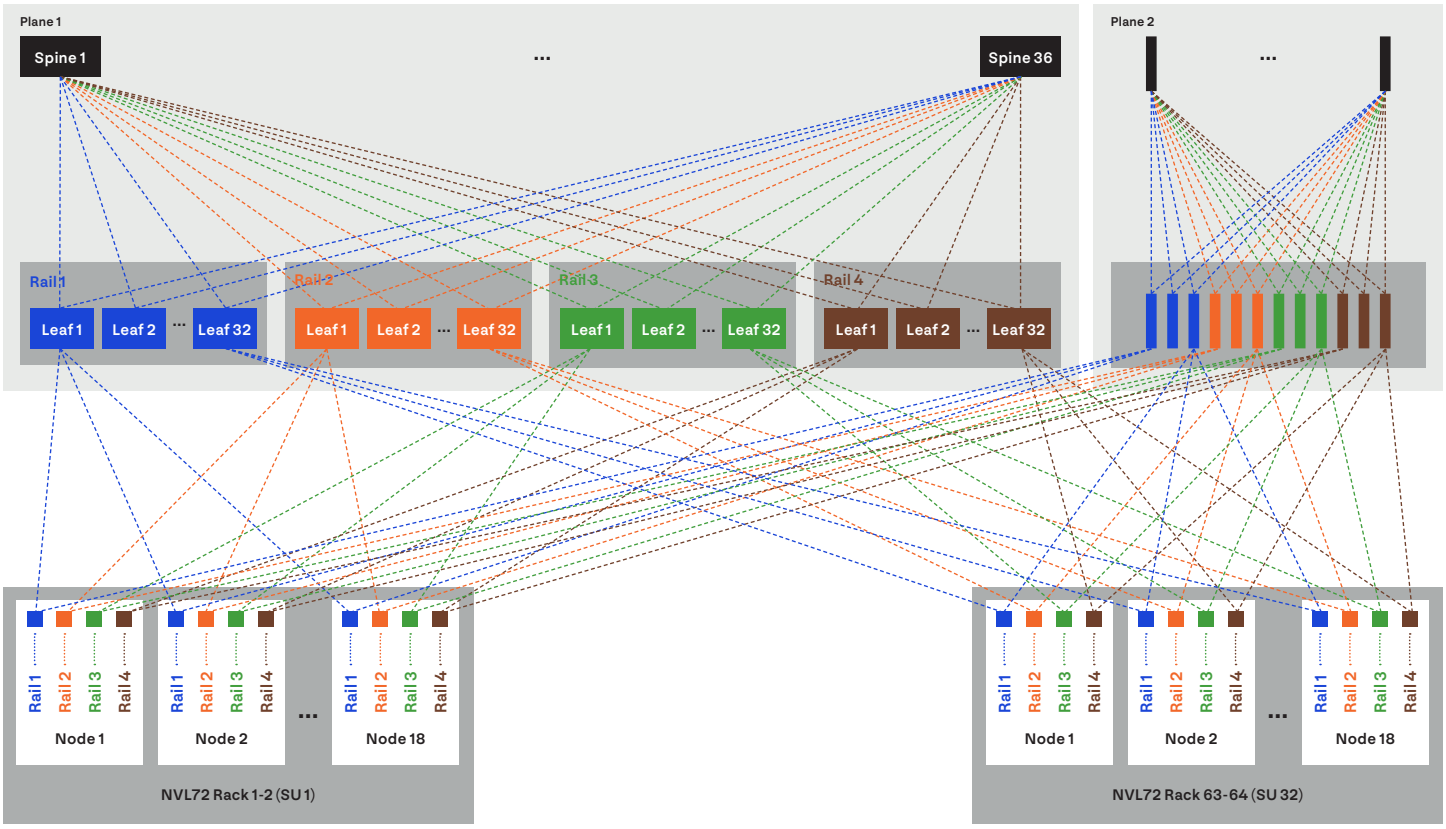


Figure 34. Example of 4,608 GPU Compute Fabric GB300 Ethernet, Dual-Plane Topology utilizing a Two-Tier architecture.

36,864 GPU Cluster, 2x 400G, connected rails, disconnected planes.

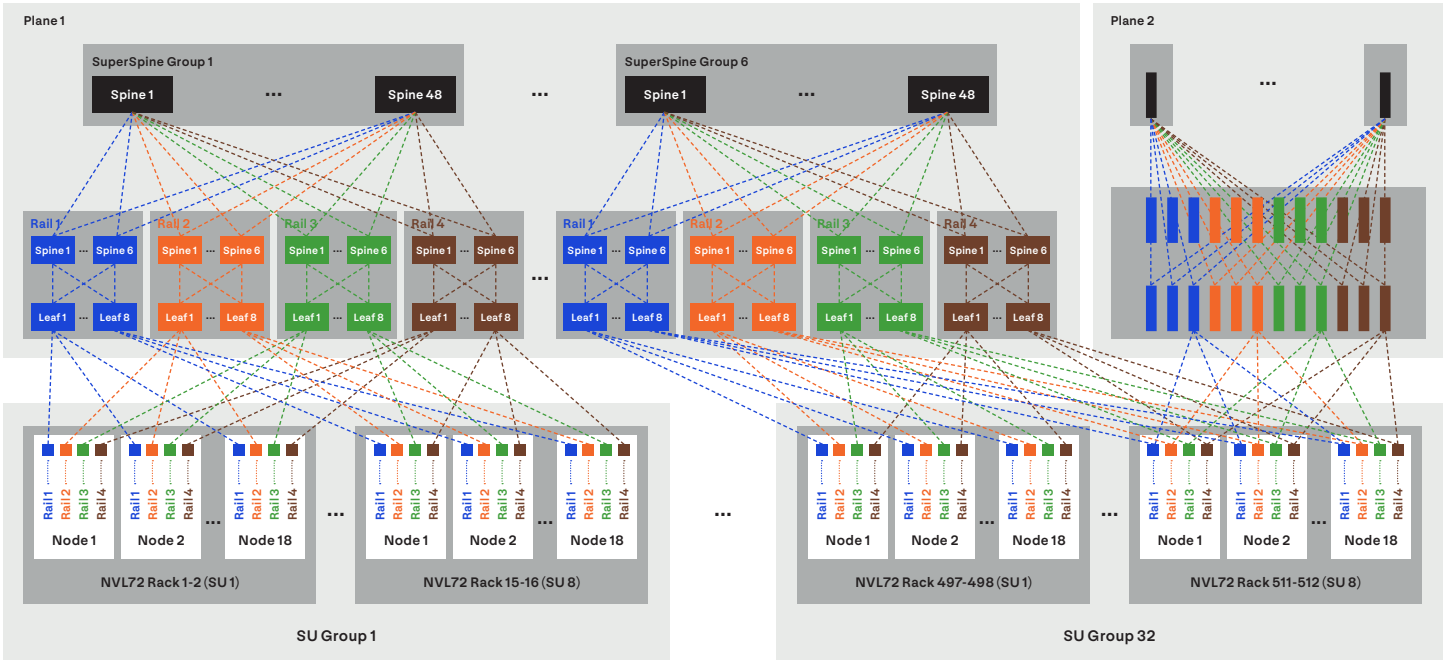


Figure 35. Example of 36,864 GPU Compute Fabric GB300 Ethernet, Dual-Plane Topology utilizing a Three-Tier architecture.

18,432 GPU Cluster, 4x 400G, connected rails, disconnected planes.

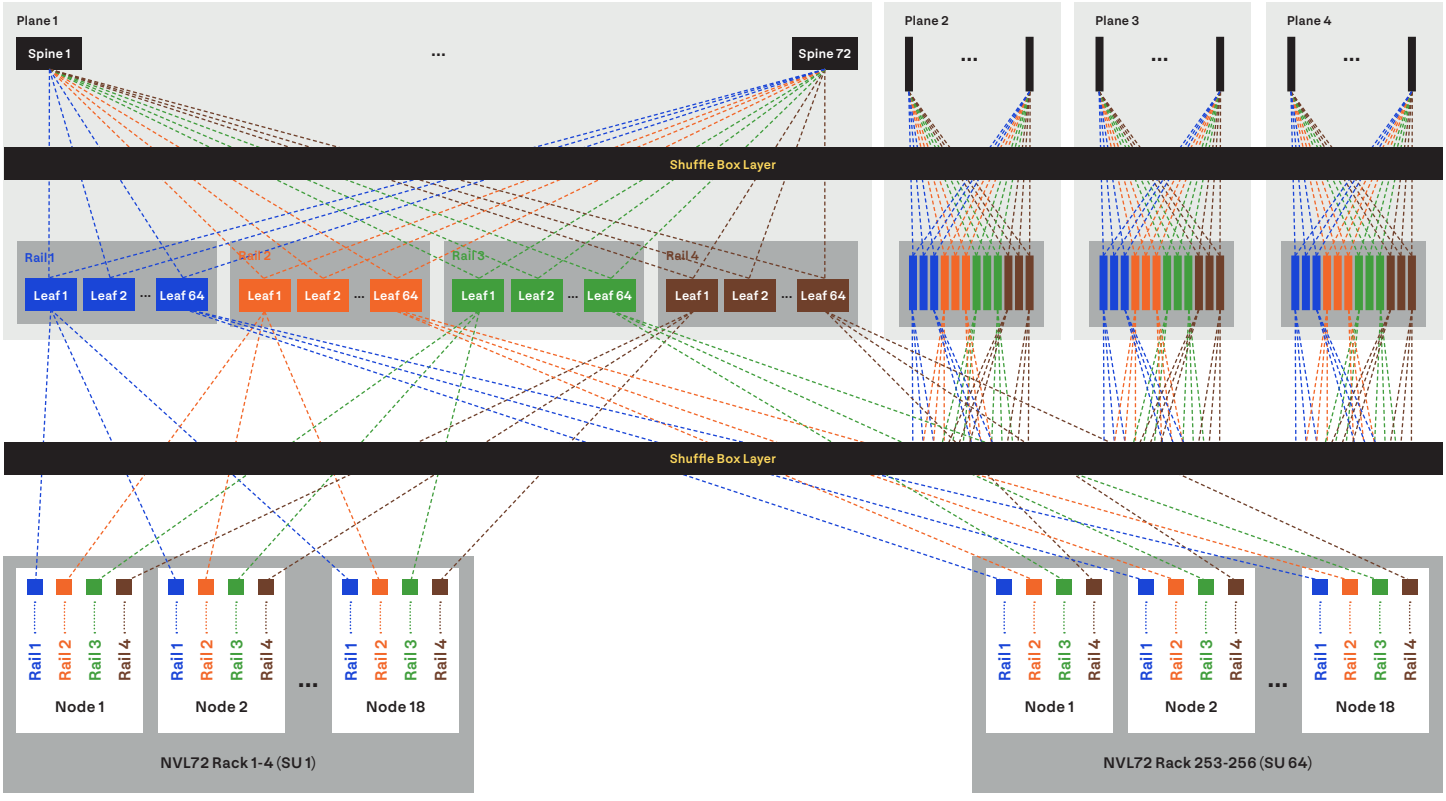


Figure 36. Example of 18,432 GPU Compute Fabric GB300 Ethernet, Quad-Plane Topology utilizing a Two-Tier architecture with Shuffle Box.

- **Rail-Optimized Connectivity**

  - The 4 Rail architecture remains as a constant for GB300 Ethernet.
  - However, differing from GB200, each node with 4x GPUs was physically represented by one single MPO-8/12, in GB300 each server will still have 4x GPUs, but each GPU physically represented **by 2x MPO-8/12** at 400G OSFP (see figure 33).
  - This is equivalent to 144x MPO-8/12 compute/backend connections coming out from a single NVL72 rack.

- **Network Components**

  - **Leaf Switches:**
    - Spectrum-X SN5600 switches support 64 OSFP Twin MPO-8/12 APC ports (128x 400G links).
    - Responsible for connecting GPUs to the Spine layer.

  - **Spine Switches:**
    - Aggregate traffic from multiple Leaf switches.
    - SN5600 switches are used for Spine layers in dual-plane or quad-plane designs.

  - **SuperSpine Layer (large deployments with Three-Tier architecture):**
    - The SuperSpine layer is specific to three-tier compute fabric designs used for large-scale deployments. It acts as the third tier in the hierarchy, connecting multiple Spine-Leaf Groups together to enable scalability across thousands of GPUs or racks.

- **Cabling Requirements**

  - 400G MPO-8/12 APC cables are used for connections between GPUs, Leaf switches, Spine switches and SuperSpine switches whenever applicable.
  - Multimode Transceivers (SR4) for short distances (up to 50 meters).
  - Single-mode Transceivers (DR4) for longer distances (up to 500 meters).
  - The cabling solution set includes Corning CORE-Trunks, traditional individual MPO jumpers for Point-to-Point cabling, and Structured Cabling solutions using EDGE8® systems. These cabling components can be applied to any GB300 topology, ensuring flexibility and compatibility across deployments.

2.7.2. NVL72 GB300 InfiniBand

The GB300 InfiniBand Compute Fabric is designed to provide high-performance connectivity for GPUs in NVIDIA NVL72 systems, utilizing Quantum-3 InfiniBand switches for scalability, low latency, and non-blocking communication. Here's what needs to be considered to deploy this architecture:

- **Modular Scalable Units (SUs)**
  - As in GB200 InfiniBand, each Scalable Unit (SU) consists of **16 GB300 NVL72 racks,** with each rack containing 72 GPUs (see Figure 37).

- **Dual-Plane Topology**
  - Each GPU connects to two independent planes for load balancing and redundancy.
  - Based on the number of GPUs being deployed, a Two-Tier (Leaf-Spine) or Three-Tier architecture (Leaf-Spine-Core) can be implemented.

- **Rail-Optimized Connectivity**
  - The 4 Rail architecture remains as a constant for GB300 InfinBand.
  - Each node or server will have 4x GPUs, with each GPU physically represented by **1x MPO-8/12** at 800G **XDR** OSFP.
  - This is equivalent to 72x MPO-8/12 compute/backend connections coming out from a single NVL72 rack.

- **Network Components**
  - **Leaf Switches:**
    - Quantum-X Q3200-RA switches are 2U switches that contain two independent 18 OSFP (twin-MPO) port switches within a single enclosure, with no communication between the two switches.
    - Each 2U enclosure provides a total of 2x 36 MPO ports (800G XDR) across 2x 18 OSFP ports.
    - Responsible for connecting GPUs to the Spine layer.

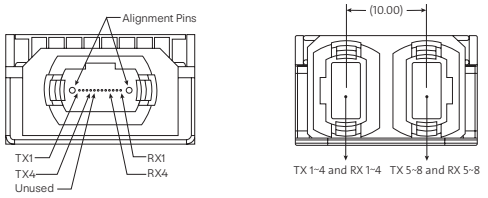
  - **Spine Switches:**
    - Aggregate traffic from multiple Leaf switches.
    - The Quantum-X Q3400-RA switch is a 4U switch equipped with 72 OSFP (twin-MPO) ports within a single enclosure.
    - It provides a total of 144 MPO ports (800G XDR) across the 72 OSFP ports.

  - **Core Switches:**
    - Q3400-RA switches are utilized as Core switches in larger-scale deployments, ensuring scalability and high-bandwidth connectivity.

- **Cabling Requirements**

  - 800G MPO-8/12 APC cables are used for connections between GPUs, Leaf switches, Spine switches and Core switches.
  - Single-mode Transceivers (DR4) used for distances up to 500 meters.
  - The cabling solution set includes Corning CORE-Trunks, traditional individual MPO jumpers for Point-to-Point cabling, and Structured Cabling solutions using EDGE8® systems. These cabling components can be applied to any GB300 topology, ensuring flexibility and compatibility across deployments.

InfiniBand Optical Interface



- **Server:** 4x GPUs, each with 1x MPO (1x 800G)
- **Switch:** 2x MPO (2x 800G)
- **Topology design:** Dual-plane

Compute Component Count				Ethernet Switch Counts			Cable Counts		
SU	NVL72 Racks	Node	GPU	Leaf (a)	Spine (b)	Core (b)	Node-Leaf	Leaf-Spine	Spine-Core
1	16	288	1,152	32	16	N/A	1,152	1,152	N/A
2	32	576	2,304	64	32	16	2,304	2,304	2,304
4	64	1,152	4,608	128	64	32	4,608	4,608	4,608
8	128	2,304	9,216	256	128	64	9,216	9,216	9,216
16	256	4,608	18,432	512	256	128	18,432	18,432	18,432

InfiniBand GB300 Component Counts Reference Table

a) Using Q3200-RA as Leaf Switches
b) Using Q3400-RA as Spine and Leaf Switches

Quantum-3 InfiniBand (XDR)

4U Q3400-RA (InfiniBand) = 144 Ports 800G

Q3400-RA
- 4U switch with 72 OSFP (twin-MPO) ports
- Total of 144 MPO ports 800G XDR over 72 OSFPs

2U Q3200-RA (InfiniBand) = 72 Ports 800G

Q3200-RA
- 2 independent 18 OSFP (twin-MPO) port switches in a single 2U enclosure
- Total of 2x 36 MPO ports 800G XDR over 2x 18 OSFPs

Figure 37. GB300 InfiniBand Cluster sizes and component counts in a dual-plane design.

2.8. Conclusion

In summary, understanding the cabling requirements for each level (A, B, and C) is essential for effectively deploying an NVL72 GB200 or GB300 GPU Cluster. Implementing CORE-Trunks or Structured Cabling wherever possible simplifies cable management and improves efficiency, particularly in larger-scale deployments.

Collaborating with Corning's engineering team during the design phase ensures that cabling strategies are tailored to meet the specific needs of the data center and align with customer requirements.

Annex 1 – High Density Housings

EDGE8® HD housings mount in 19-inch racks or cabinets and provide industry-leading ultra-high-density connectivity when combined with EDGE8 modules, panels, harnesses, trunks, and jumpers.

As each customer and project has specific needs, please add the housing that best suits your needs to the BOM:

	Part Number (Global)	Maximum Number of Modules or Panels	Maximum Fiber Density		Height
	EDGE8-01U-SP	18	LC	144 fibers	1U
			MTP®	576 fibers	
	EDGE8-02U	36	LC	288 fibers	2U
			MTP	1,152 fibers	
	EDGE8-04U	72	LC	576 fibers	4U
			MTP	2,304 fibers	

Table 11 – High Density Housings.

Annex 2 – Polarity Drawings

Polarity drawings, often referred to as fiber optic polarity diagrams, are essential when designing and implementing data center links using fiber optic cabling. They play a crucial role in ensuring proper connectivity, signal integrity, and compatibility between different network components.

This section will cover the specific polarity drawings applicable to each one of the scenarios previously described.

Scenario 1 – 1600G, 800G and 400G – Server to Switch Applications MPO-8/12 APC to MPO-8/12 APC using Point-to-Point Cabling

Use Case A

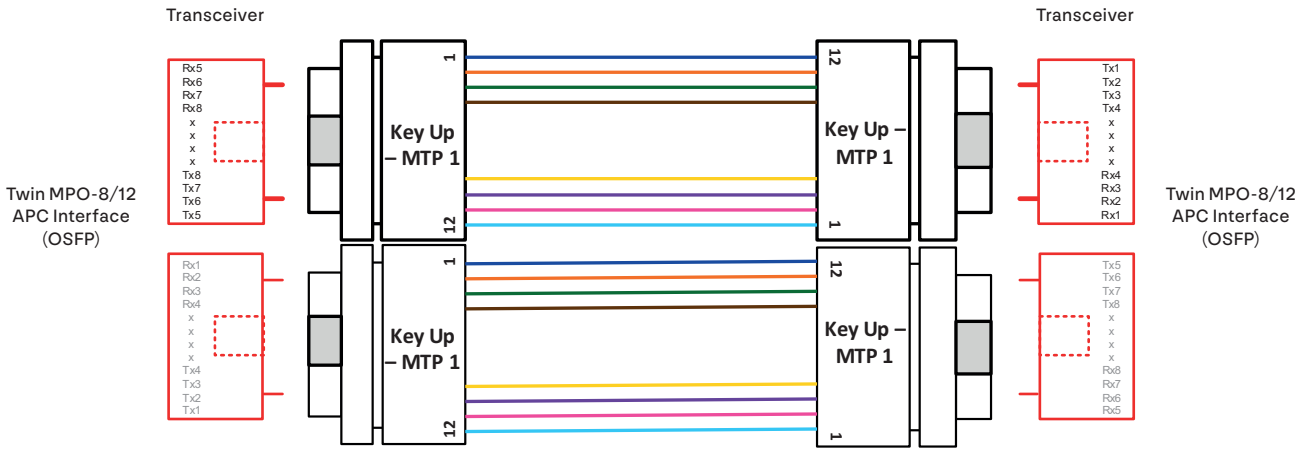


Figure 38. Scenario 1 – 1600G, 800G and 400G – Switch to Server Local. Use Case A.

Use Case B

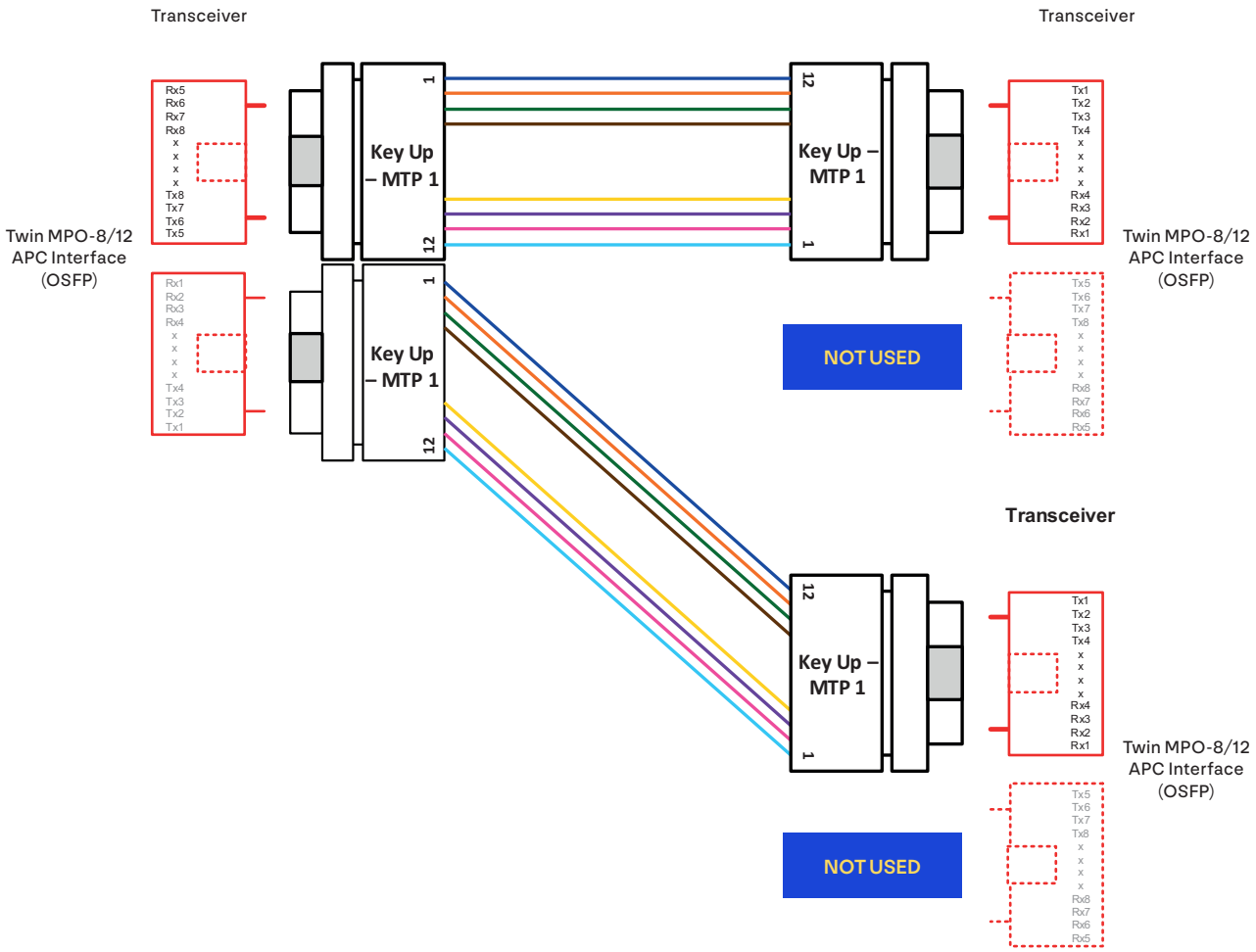


Figure 39. Scenario 1 – 1600G, 800G and 400G – Switch to Server Local. Use Case B.

Use Case C

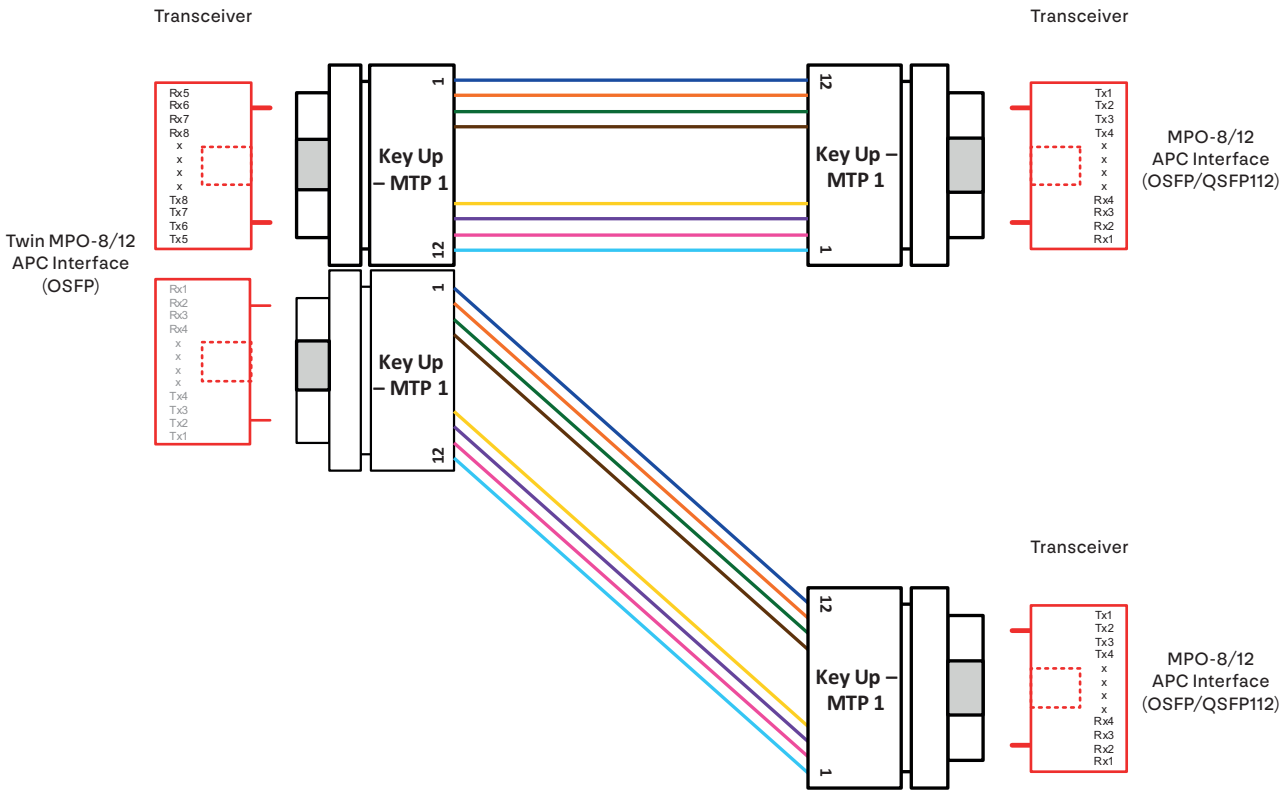


Figure 40. Scenario 1 – 1600G, 800G and 400G – Switch to Server Local. Use Case C.

Scenario 2 – 1600G, 800G and 400G - Switch to Switch Applications MPO-8/12 APC to MPO-8/12 APC using Structured Cabling Across DC with Trunk

Use Case A

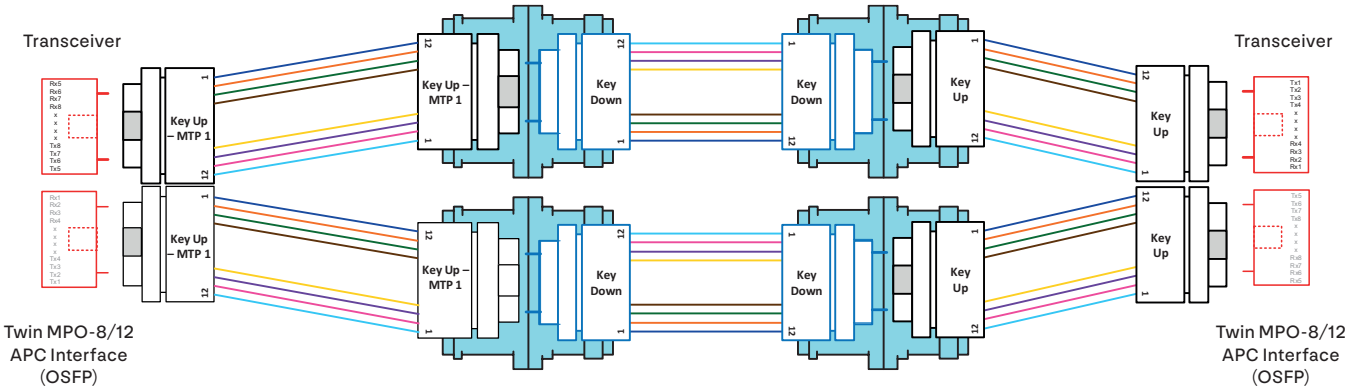


Figure 41. Scenario 2 – 1600G, 800G and 400G – Switch to Switch Across DC with Trunk. Use Case A

Use Case B

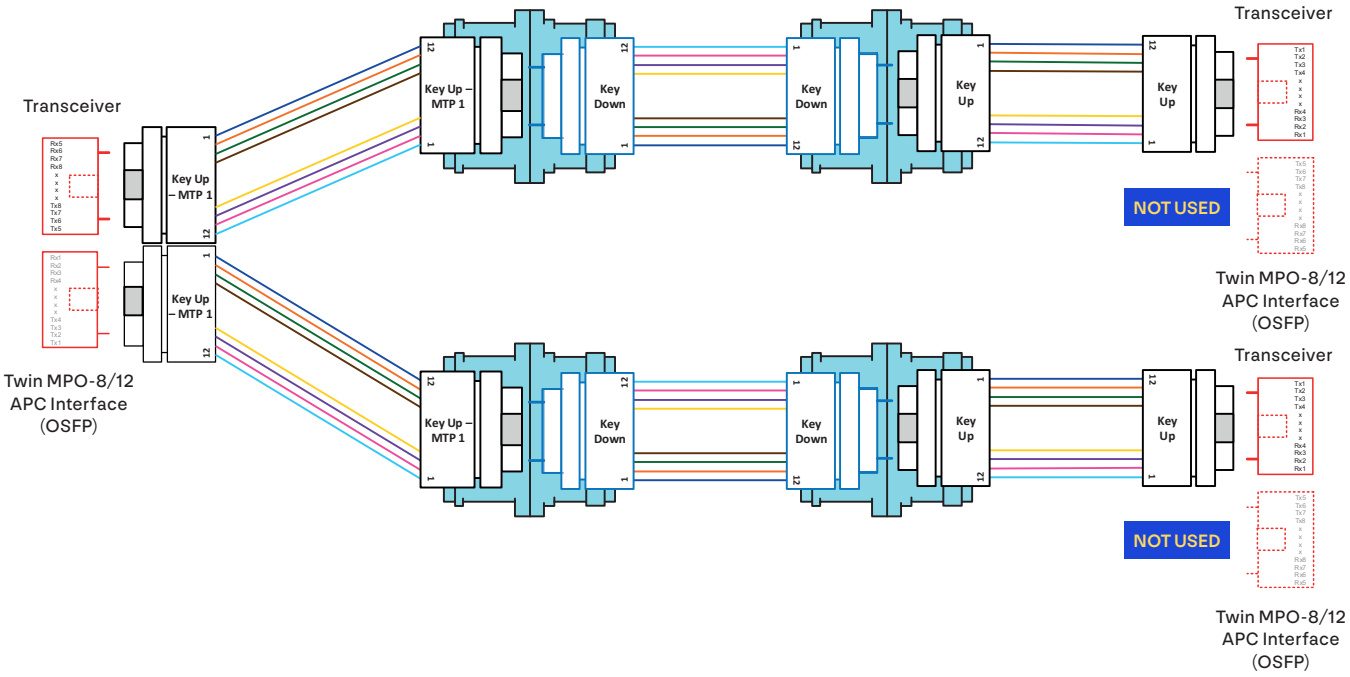


Figure 42. Scenario 2 – 1600G, 800G and 400G – Switch to Switch Across DC with Trunk. Use Case B

Use Case C

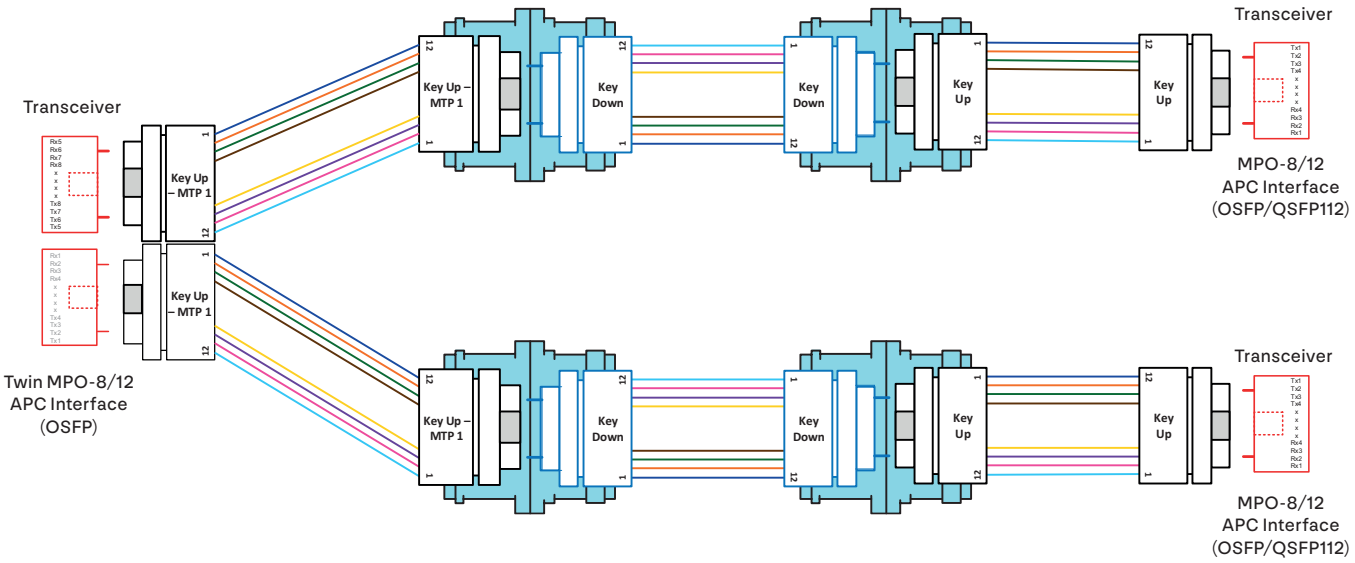


Figure 43. Scenario 2 – 1600G, 800G and 400G – Switch to Switch Across DC with Trunk. Use Case C.

Scenario 3 – 1600G, 800G, 400G and 200G - Server to Switch Applications MPO-8/12 APC to MPO-8/12 APC using Point-to-Point Cabling

Use Case A

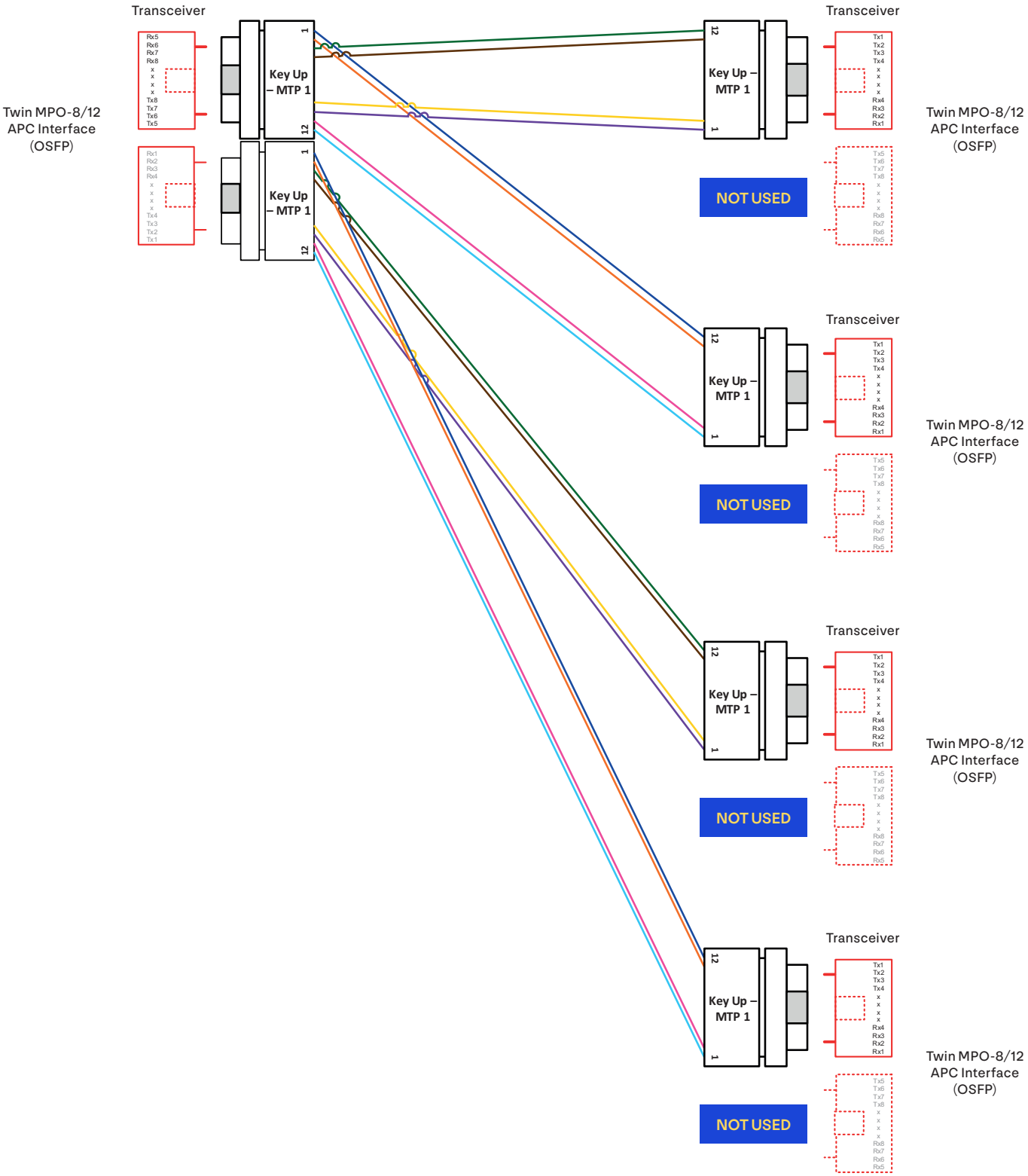


Figure 44. Scenario 3 – 1600G, 800G, 400G and 200G – Switch to Server Local. Use Case A.

Use Case B

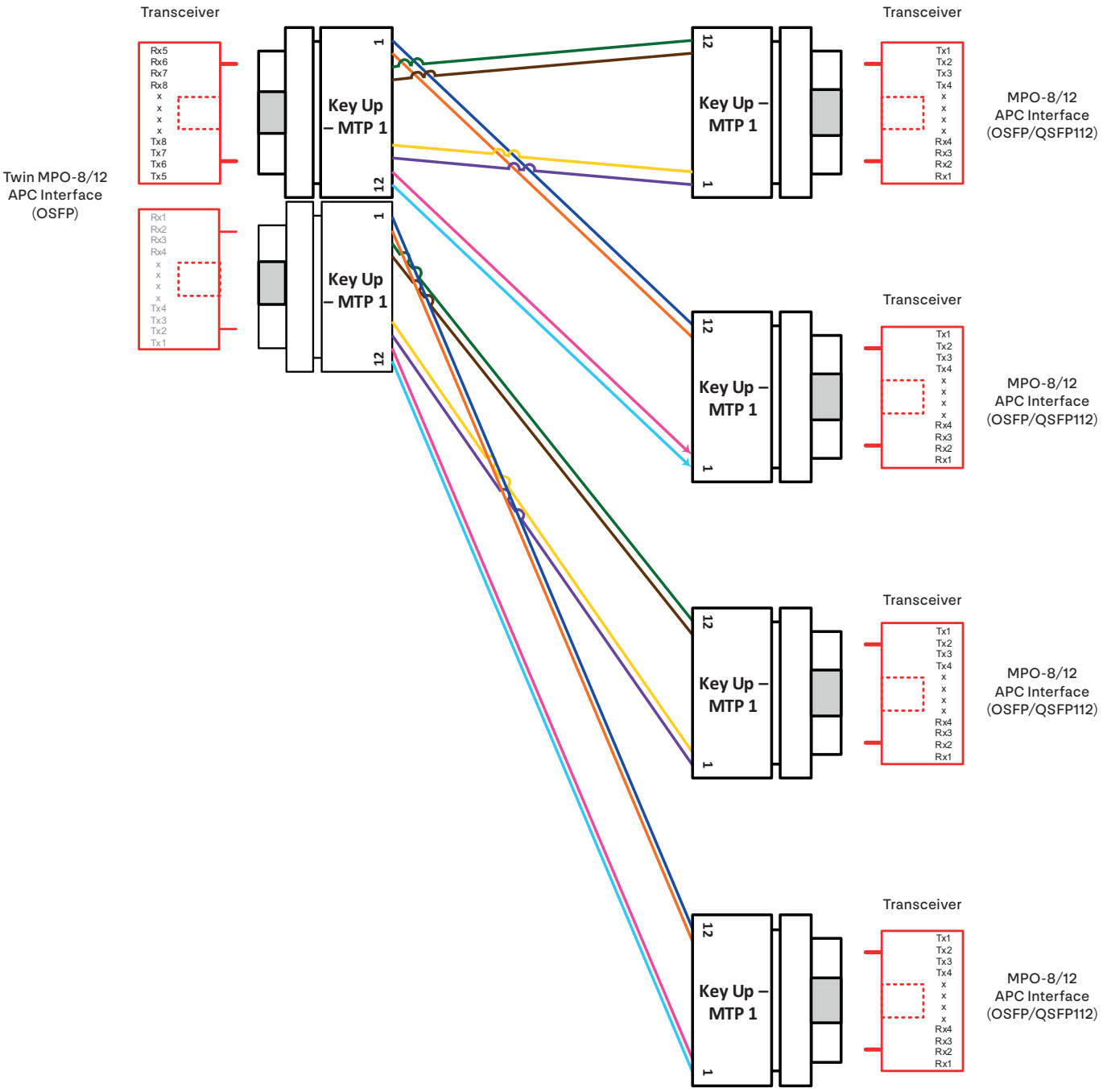


Figure 45. Scenario 3 – 1600G, 800G, 400G and 200G – Switch to Server Local. Use Case B.

Scenario 4 – 1600G, 800G, 400G and 200G - Switch to Switch Applications MPO-8/12 APC to MPO-8/12 APC using Structured Cabling Across DC with Trunk

Use Case A

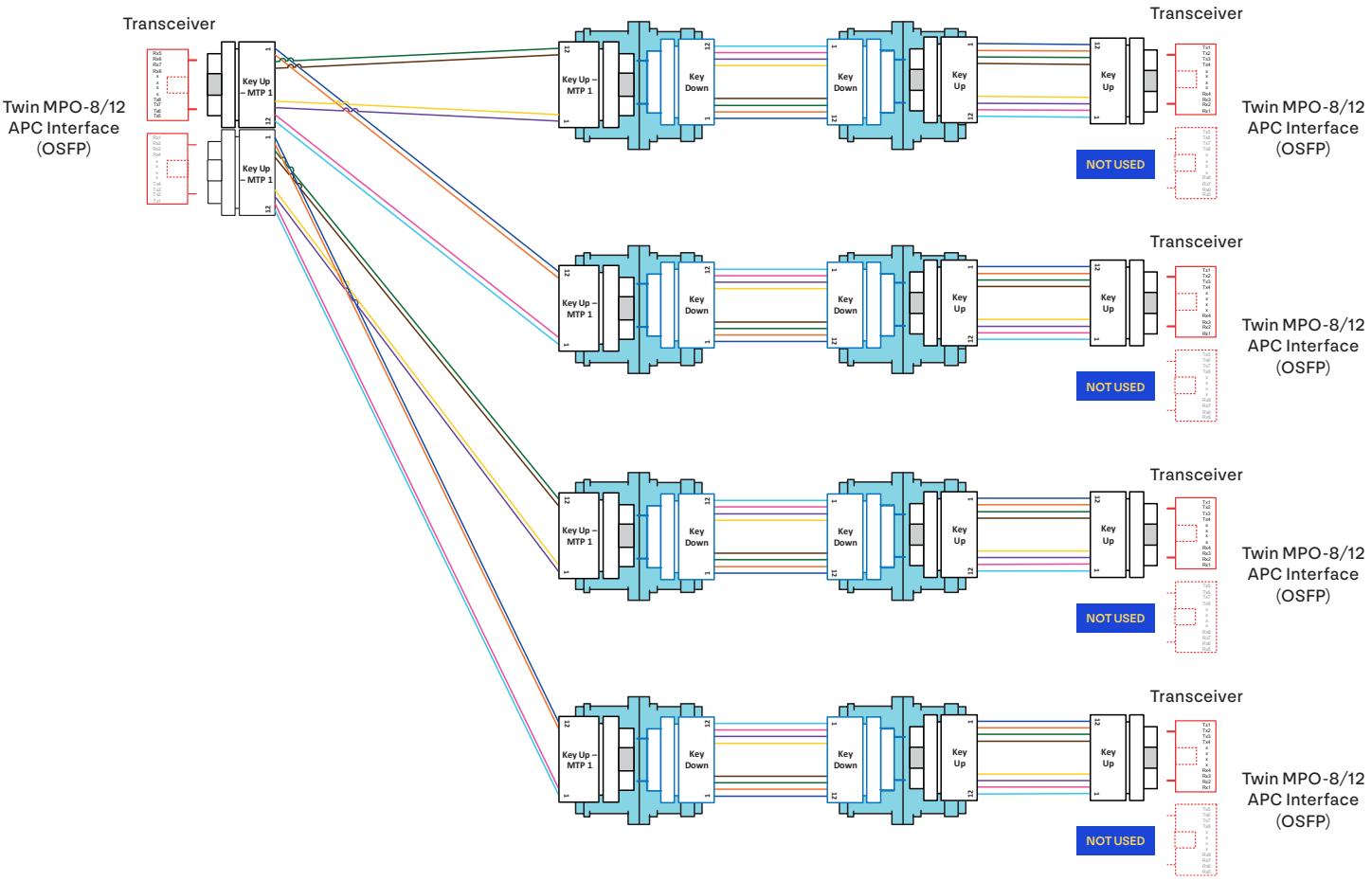


Figure 46. Scenario 4 – 1600G, 800G, 400G and 200G – Switch to Switch Across DC with Trunk. Use Case A.

Use Case B

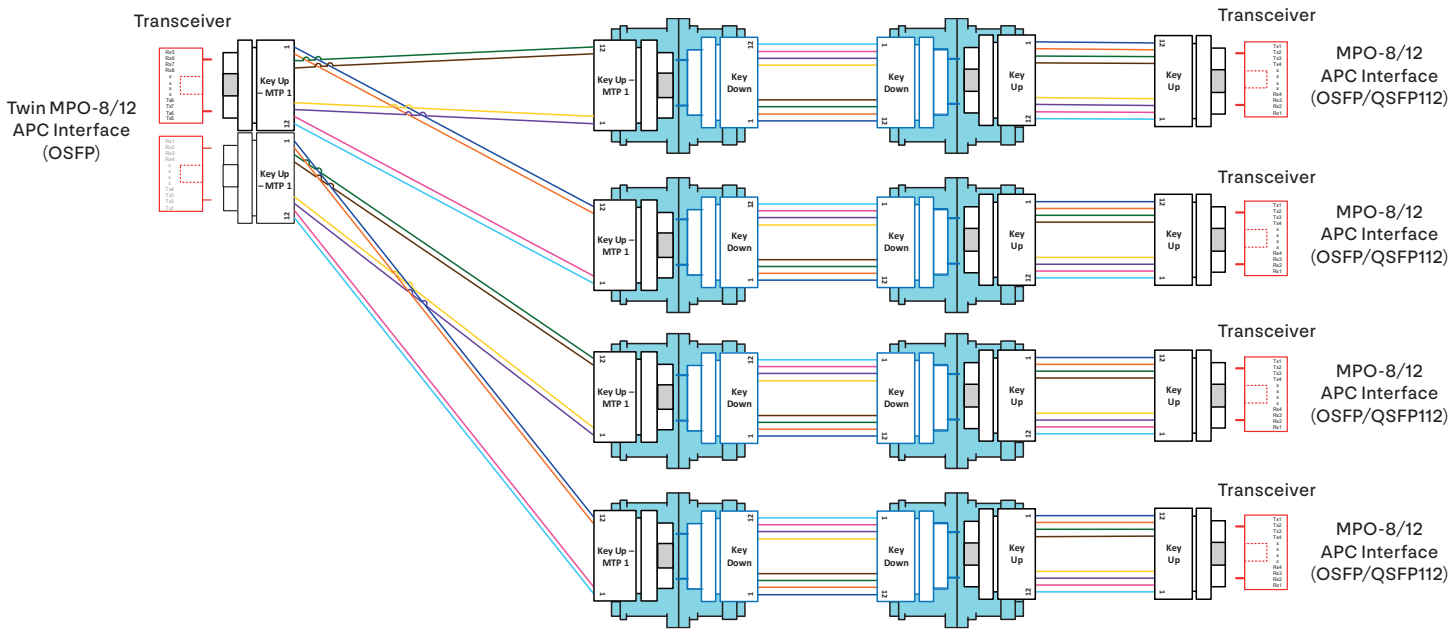


Figure 47. Scenario 4 – 1600G, 800G, 400G and 200G – Switch to Switch Across DC with Trunk. Use Case B.

Scenario 5 – 800G and 400G - Switch to Switch Applications LC-Duplex to LC-Duplex using Point-to-Point Cabling

Use Case A

Figure 48. Scenario 5 – 800G and 400G – Switch to Switch Local. Use Case A.

Use Case B

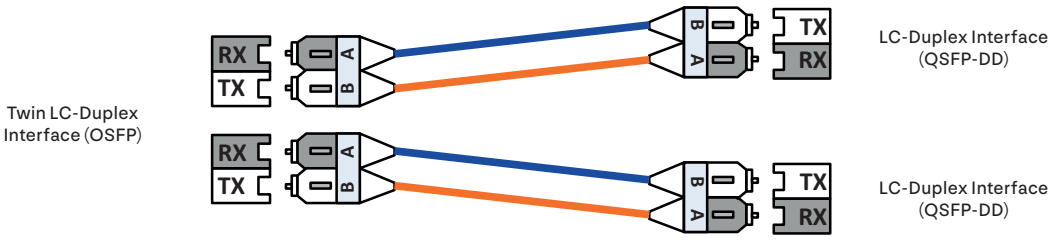


Figure 49. Scenario 5 – 800G and 400G – Switch to Switch Local. Use Case B.

Scenario 6 – 800G and 400G - Switch to Switch Applications
LC-Duplex UPC to LC-Duplex UPC using Structured Cabling Across DC with Trunk

Use Case A

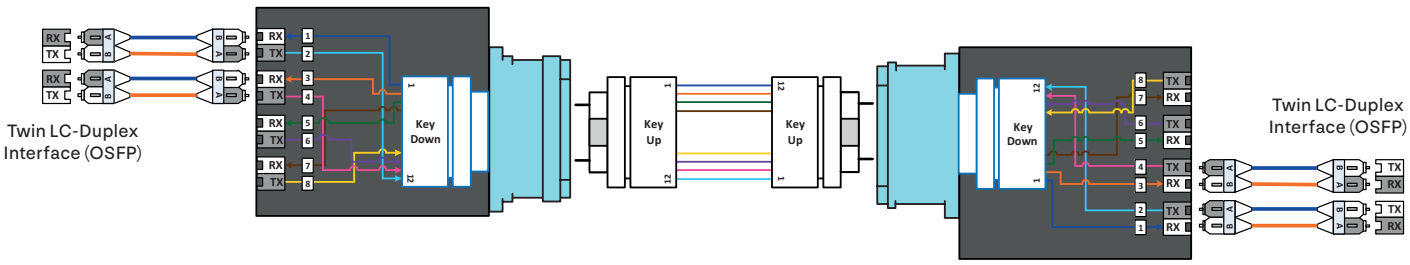


Figure 50. Scenario 6 – 800G and 400G – Switch to Switch Across DC with Trunk. Use Case A.

Use Case B

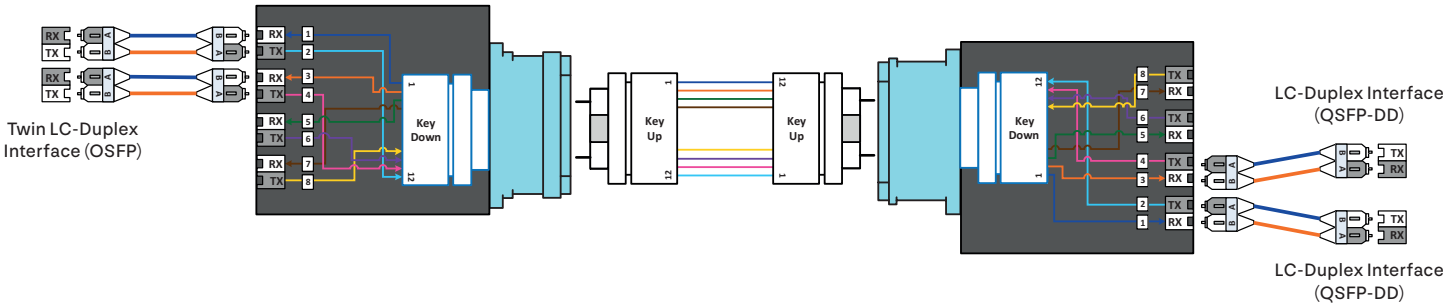


Figure 51. Scenario 6 – 800G and 400G – Switch to Switch Across DC with Trunk. Use Case B.

Annex 3 – References and Contact

This section contains a partial list of references to NVIDIA Overview Whitepapers. For more detailed information on NVIDIA's products, please visit www.docs.nvidia.com

Transceivers:

- **200G Optical Lanes (XDR)**
  - **MMS4A00** 1600 Gbps Twin-port OSFP 2x800Gb/s **Single-mode 2 x DR4,** 500 m
    www.docs.nvidia.com/networking/display/9iahx00xmosfptcvr1600
  - **MS4A20-XM800** 800Gbps Single-port OSFP 1x800Gb/s **Single-mode DR4,** 500 m
    www.docs.nvidia.com/networking/display/9iat0mosfp800sprhs

- **100G Optical Lanes (NDR)**
  - **MMS4X00-NM** 800Gbps Twin-port OSFP 2x400Gb/s **Single-mode 2 x DR4,** 500 m
    www.docs.nvidia.com/networking/display/mms4x00nm800g500m/application+overview
  - **MMS4X00-NS** 800Gbps Twin-port OSFP 2x400Gb/s **Single-mode 2xDR4,** 100 m
    www.docs.nvidia.com/networking/display/800gmms4x00ns/overview
  - **MMA4Z00-NS** 800Gb/s Twin-port OSFP, 2x400Gb/s **Multimode 2xSR4,** 50 m
    www.docs.nvidia.com/networking/display/800gmma4z00ns/overview
  - **MMS4X50-NM** 800Gb/s Twin-port OSFP, 2x400Gb/s **Single-mode 2xFR4,** 2 km
    www.docs.nvidia.com/networking/display/mms4x50nm800g2kmpub
  - **MMS1X00-NS400** 400Gb/s Single-port QSFP112, 1x400Gb/s **Single-mode DR4,** 100 m
    www.docs.nvidia.com/networking/display/mms1x00ns400/overview
  - **MMA1Z00-NS400** 400Gb/s Single-port QSFP112, 1x400Gb/s **Multimode SR4,** 50 m
    www.docs.nvidia.com/networking/display/mms1z00ns400sr4

NVIDIA Cables:

- **MFP7E30-Nxxx,** Single-mode, **Straight** Crossover Fibers Cable
  www.docs.nvidia.com/networking/display/mfp7e30nxxxpub/specifications
- **MFP7E40-Nxxx,** Single-mode, **Splitter** Crossover Fibers Cable
  www.docs.nvidia.com/networking/display/mfp7e40nxxxpub/specifications
- **MFP7E10-Nxxx,** Multimode, **Straight** Crossover Fibers Cable
  www.docs.nvidia.com/networking/display/mfp7e10nxxx/specifications
- **MFP7E20-Nxxx,** Multimode, **Splitter** Crossover Fibers Cable
  www.docs.nvidia.com/networking/display/mfp7e20nxxx/specifications

NVIDIA Architectures and reference pages:

- **NVL72 GB200**
  www.nvidia.com/en-us/data-center/gb200-nvl72/
- **NVL72 GB300**
  www.nvidia.com/en-us/data-center/gb300-nvl72/
- **DGX H100**
  www.docs.nvidia.com/dgx-superpod/reference-architecture-scalable-infrastructure-h100/latest/dgx-superpod-architecture
- **DGX B200**
  www.docs.nvidia.com/dgx-superpod/reference-architecture-scalable-infrastructure-b200/latest/dgx-superpod-architecture
- **DGX B300**
  www.docs.nvidia.com/dgx-superpod/reference-architecture/scalable-infrastructure-b300/latest/abstract
- **DGX GB200**
  www.docs.nvidia.com/dgx-superpod/reference-architecture-scalable-infrastructure-gb200/latest/dgx-superpod-components

CORNING

For questions, please contact Corning's Technical Support Line:

Americas
Phone: 800-743-2671
Email: dutyeng@corning.com

EMEA and APJ
Phone: +49 305 303 2134
Email: engineer.en.emea@corning.com

Corning Optical Communications LLC • 4200 Corning Place • Charlotte, NC 28216 USA
800-743-2675 • FAX: 828-325-5060 • International: +1-828-901-5000 • www.corning.com/opcomm