

# Cisco® Nexus® 9000 Switches for Connecting Intel® Gaudi® 3 Accelerators

## Solution Overview

---

Cisco Nexus 9364E-SG2 switches are validated to connect Intel Gaudi 3 accelerators for running Large-Language Model (LLM) training, inference, or a similar Artificial Intelligence (AI) or Machine Learning (ML) workload.

**Q.** What are Intel Gaudi 3 servers?

**A.** Intel Gaudi 3 servers (see Figure 1) contain eight Gaudi 3 accelerators. Language models that can't fit in or train fast enough using eight accelerators, can benefit by connecting multiple Gaudi 3 servers via an 800 Gigabit Ethernet (GbE) non-blocking network allowing RoCEv2 transport. A Gaudi 3 server provides 24 x 200 GbE scale-out or back-end interfaces exposed via six OSFP ports for an inter-Gaudi 3 network, also known as a scale-out network or a back-end network. These ports, marked as OSFP-00 to OSFP-05 in Figure 1, are dedicated to interconnecting Gaudi 3 accelerators in other servers and are not connected to external networks. A Gaudi 3 server also provides additional interfaces for external, storage, and management connectivity. For more information, refer to the [Intel Gaudi 3 AI Accelerator White Paper](#).



**Figure 1.**  
Front-view of an Intel Gaudi 3 server

**Q.** What are Cisco Nexus 9364E-SG2 switches?

**A.** Cisco Nexus 9364E-SG2 switches (see Figure 2) provide 64 x 800 GbE or 128 x 400 GbE or 256 x 200 GbE or 512 x 100 GbE interfaces in 2 rack unit (RU) form factor. These switches are available in two models. Cisco Nexus 9364E-SG2-O offers ports in OSFP form factor, whereas Cisco Nexus 9364E-SG2-Q offers ports in QSFP-DD form factor. Except the port form-factor, both switch models are similar, and both can be used for connecting Intel Gaudi 3 servers. These switches can be connected to the Intel Gaudi 3 servers using 4 x 200 GbE breakouts.



**Figure 2.**  
Front view of the Cisco Nexus 9364E-SG2-O switch

**Q.** What cables and transceivers are validated for connecting Cisco Nexus 9364E-SG2 switches with Intel Gaudi 3 servers?

**A.** Cisco OSFP-800G-DR8 transceivers are validated for the Intel Gaudi 3 server scale-out ports. The same transceiver is also validated to be plugged into the Cisco Nexus 9364E-SG2-O switch for building a scale-out or backend network. For more information, refer to [Cisco OSFP 800G Transceiver Modules Data Sheet](#). MPO cables are validated for connectivity. For ISL connections, the same transceivers were used for the validation, but any other qualified transceiver listed on the [Cisco Optics-to-Device Compatibility Matrix](#) can also be used. In addition to the scale-out ports, Intel Gaudi 3 servers use additional high-speed NICs for frontend connectivity. Cisco Nexus 9364E-SG2 switches can be used even for front-end networks, allowing consistent operations for backend, frontend, and storage networks.



**Figure 3.**  
Cisco OSFP-800G-DR8 transceiver

**Q.** What software versions were used for this validation?

**A.** Table 1 shows the validated software versions.

**Table 1.** Validated software versions of Intel Gaudi 3 servers and Cisco Nexus 9364E-SG2 switches

Component	Version
Intel Gaudi 3 Server	Supermicro SYS-822GA-NGR3
Operating System on Intel Gaudi 3 Server	Ubuntu 22.04.5
Intel Gaudi Software Suit (Synapse) on Intel Gaudi 3 Server	1.22.1
Cisco Nexus 9364E-SG2 switches	NX-OS 10.6(1)

**Q.** How to buy Cisco Nexus switches?

**A.** Cisco part number, N9364E-SG2-O, can be used to buy Cisco Nexus 9364E-SG2-O switches. We recommend using Cisco Nexus Dashboard to simplify the configuration and monitoring of all networks of an AI cluster. Cisco part number, OSFP-800G-DR8, can be used to buy transceiver for Intel Gaudi 3 server scale-out ports and Cisco Nexus 9364E-SG2-O switch. For more information, refer to the [Cisco Nexus 9364E-SG2 Switches Data Sheet](#), [Cisco Nexus Dashboard Data Sheet](#), and [Cisco OSFP 800G Transceiver Modules Data Sheet](#).

**Q.** What is the typical sizing of a Gaudi 3 accelerator scale-out network?

**A.** Table 2 shows sizing of an Intel Gaudi 3 accelerator scale-out network using a non-blocking and non-oversubscribed design.

**Table 2.** Intel Gaudi 3 scale-out network sizing using Cisco Nexus 9364E-SG2 switches

# Gaudi 3 accelerators	# Gaudi 3 servers (1)	# scale-out/backend OSFP ports (2)	# Cisco Nexus 9364E-SG2 leaf switches (3)	# Cisco Nexus 9364E-SG2 spine switches (3)	# Total Cisco Nexus 9364E-SG2 switches	# OSFP-800G-DR8 transceivers in Gaudi 3 servers	# OSFP-800G-DR8 transceivers in switches	# of MPO cables (4)
8	1	6	0	0	0	0	0	0
16	2	12	1	0	1	12	12	24
32	4	24	1	0	1	24	24	48
64	8	48	1	0	1	48	48	96
128	16	96	3	0	3	96	96	192
256	32	192	6	3	9	192	576	768
512	64	384	12	6	18	384	1152	1536
1024	128	768	24	12	36	768	2304	3072
2048	256	1536	48	24	72	1536	4608	6144
4096	512	3072	96	48	144	3072	9216	12288

<sup>1</sup> Gaudi 3 server has eight Gaudi 3 accelerators

<sup>2</sup> Each Gaudi 3 server has six OSFP ports for connecting to an inter-Gaudi scale-out network

<sup>3</sup> Each Nexus 9364E-SG2 switch can connect up to 64 x 800 GbE or 128 x 400 GbE or 256 x 200 GbE interfaces

<sup>4</sup> OSFP-800G-DR8 is a twin-port transceiver, consuming two MPO cables

**Americas Headquarters**  
Cisco Systems, Inc.  
San Jose, CA

**Asia Pacific Headquarters**  
Cisco Systems (USA) Pte. Ltd.  
Singapore

**Europe Headquarters**  
Cisco Systems International BV Amsterdam,  
The Netherlands

Cisco has more than 200 offices worldwide. Addresses, phone numbers, and fax numbers are listed on the Cisco Website at <https://www.cisco.com/go/offices>.

Cisco and the Cisco logo are trademarks or registered trademarks of Cisco and/or its affiliates in the U.S. and other countries. To view a list of Cisco trademarks, go to this URL: <https://www.cisco.com/go/trademarks>. Third-party trademarks mentioned are the property of their respective owners. The use of the word partner does not imply a partnership relationship between Cisco and any other company. (1110R)