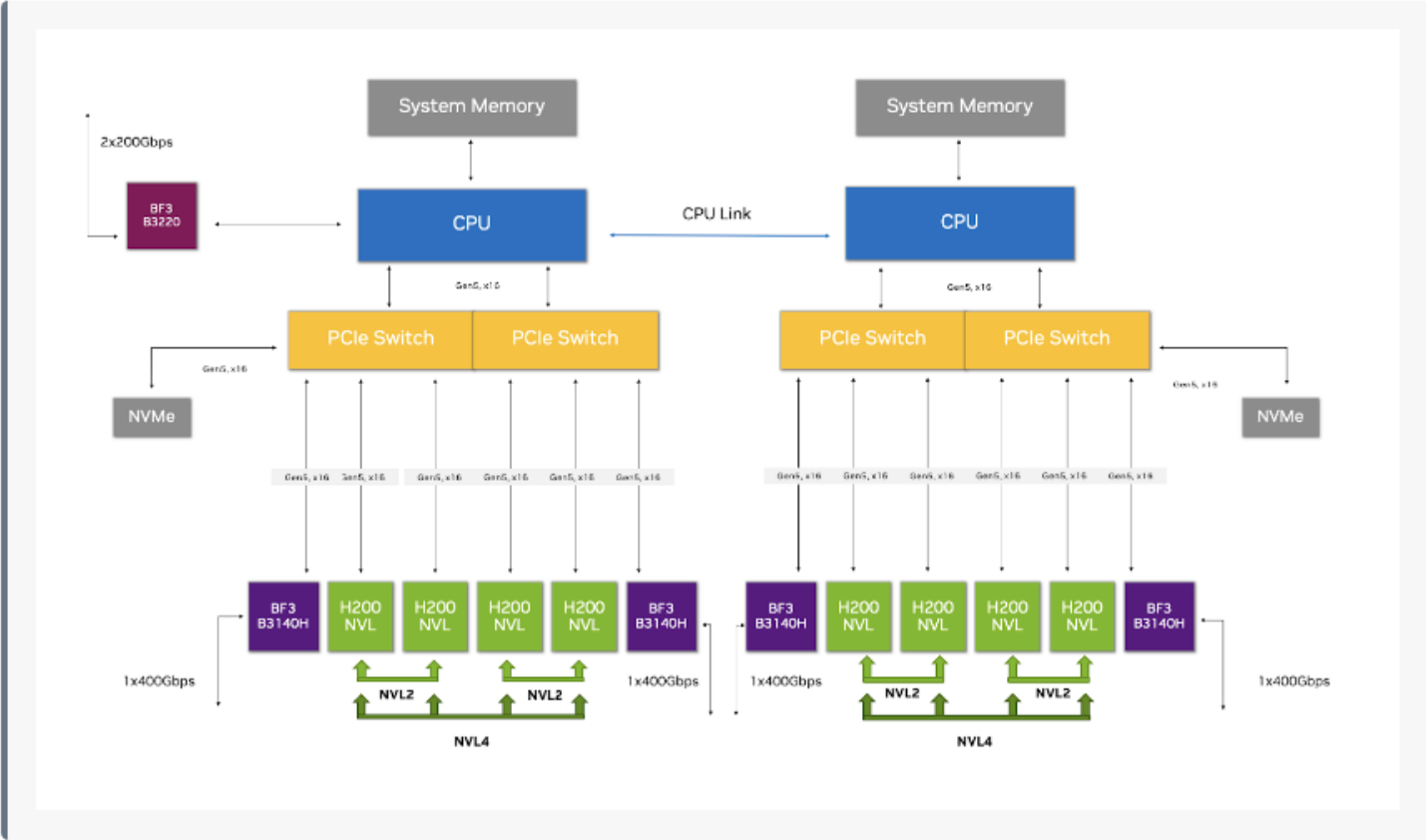


Diagrama y componentes de configuración optimizada NVIDIA H200 NVL



Components for the H200 NVL NVIDIA-Certified system are listed in the below table.

Parameter	System Configuration
GPU configuration	<p>GPUs are balanced across CPU sockets and root ports.</p> <ul style="list-style-type: none"> › Inference servers: 2x, 4x, and 8x GPUs per server › Training and DL servers: Minimum 8 GPUs per server <p>See the topology diagram above for details</p>
NVLink Interconnect	<p>H200 NVL supports NVL4 and NVL2 bridges. Pairing of GPU cards under the same CPU socket is best; pairing of GPU cards under different CPU sockets is acceptable but not recommended. See topology diagram above for NVLink bridging recommendations.</p>
CPU	<p>Intel Emerald Rapids, Intel Sapphire Rapids, Intel Granite Rapids and Intel Sierra Forest AMD Genoa and AMD Turin</p>
CPU sockets	<p>Two CPU sockets minimum</p>
CPU speed	<p>2.0 GHz minimum CPU clock</p>
CPU cores	<p>Minimum 7 physical CPU cores per GPU</p> <ul style="list-style-type: none"> › For configuration using MIG, 2 CPU cores required per MIG instance › For OS kernel or virtualization, additional two cores per GPU
System memory (total across all CPU sockets)	<p>Minimum 128 GB of system memory per GPU</p>
DPU	<p>One NVIDIA® BlueField®-3 DPU per server</p>
PCI Express	<p>One Gen5 x16 link per maximum two GPUs. Recommend one Gen5 x16 link per GPU</p>
PCIe topology	<p>Balanced PCIe topology with GPUs spread evenly across CPU sockets and PCIe root ports. NIC and NVMe drives should be under the same PCIe switch or PCIe root complex as the GPUs. Note that a PCIe switch may not be needed for low-cost inference servers; direct-attach to CPU is best if possible. See the topology diagram above for details</p>
PCIe switches	<p>Gen5 PCIe switches as needed (where additional link fanout is not required, direct attach is best).</p>
Compute (E-W) NIC	<p>Four NVIDIA® BlueField®-3 SuperNICs per server Up to 400 Gbps</p>
Local storage	<p>Local storage recommendations are as follows:</p> <ul style="list-style-type: none"> › Inference Servers: Minimum 1 TB NVMe drive per CPU socket. › Training / DL Servers: Minimum 2 TB NVMe drive per CPU socket. › HPC Servers: Minimum 1 TB NVMe drive per CPU socket
Remote systems management	<p>SMBPBI over SMBus (OOB) protocol to BMC PLDM T5-enabled. SPDM-enabled</p>
Security	<p>TPM 2.0 module (secure boot)</p>