



# **NVIDIA Spectrum-X 网络平台架构**

首个专为加速 AI 工作负载而设计的以太网网络

白皮书

# 目录

提高 AI 性能和效率.....	5
NVIDIA Spectrum-X 网络平台概述.....	8
深入了解 NVIDIA Spectrum-X 网络平台.....	8
Spectrum-X 的主要优势.....	9
NVIDIA Spectrum-X 以太网交换机.....	9
NVIDIA BlueField-3 SuperNIC.....	10
NVIDIA 端到端物理层.....	11
NVIDIA 加速软件.....	12
NetQ.....	12
Spectrum 软件开发套件.....	12
NVIDIA DOCA.....	13
在以太网上部署 AI 的挑战.....	14
需要 RDMA 和无损网络.....	14
需要 RoCE 动态路由来避免拥塞.....	15
熵的重要性.....	15
NVIDIA RoCE 动态路由确保负载均衡.....	17
使用 NVIDIA RoCE 拥塞控制.....	18
NVIDIA Spectrum-X 网络平台的主要特性.....	18
NVIDIA RoCE 动态路由的工作原理.....	19
借助 NVIDIA 直接数据放置实现 NVIDIA RoCE 动态路由.....	19
RoCE 动态路由总结.....	22
使用 NVIDIA RoCE 拥塞控制实现性能隔离.....	24
RoCE 性能隔离.....	25

NVIDIA 全栈优化.....	27
从参考架构获得的 Spectrum-X 性能结果.....	28
RDMA 网络性能 .....	28
NCCL 集合性能.....	29
AI 模型性能.....	30
虚拟化对性能的影响 .....	31
专为 AI 设计的以太网.....	32

## 图列表

图 1. 管理网络/用户访问网络与 AI 网络架构之间的差异.....	6
图 2. NVIDIA Spectrum-X SN5600 以太网交换机.....	10
图 3. NVIDIA BlueField-3 以太网 SuperNIC.....	10
图 4. 典型的 Spectrum-X 网络拓扑.....	11
图 5. NVIDIA NetQ 遥测仪表板.....	12
图 6. RDMA 使 BlueField 能够直接访问 GPU 内存.....	14
图 7. 采用静态负载均衡, AI 工作负载只使用了最大吞吐量的 60%.....	15
图 8. 为什么即使是非阻塞网络也会出现长尾延迟.....	16
图 9. 流完成时间可能会有显著差异.....	16
图 10. Incast 需要 NVIDIA RoCE 拥塞控制.....	18
图 11. RoCE 动态路由缩短了流完成时间.....	22
图 12. 拥塞控制造成受害流.....	24
图 13. Spectrum-X 通过流计量和拥塞遥测解决拥塞问题.....	25
图 14. Spectrum-X 实现噪声隔离机制.....	26
图 15. Spectrum-X 交换机采用全局共享缓冲区设计.....	27
图 16. Spectrum-X 为 RDMA 二分法实现更高的有效带宽和更低的时延.....	29
图 17. Spectrum-X 为 AI 云中运行的 NCCL 集合提供更高的总线带宽.....	30
图 18. Spectrum-X 在嘈杂的 AI 云环境中加速训练 AI 工作负载.....	31
图 19. Spectrum-X 在使用和不使用网络虚拟化的情况下性能表现同样出色.....	32

---

## 提高 AI 性能和效率

云 AI 工作负载需求正以前所未有的速度增长，同时生成式 AI 的应用也在激增。每年都会建设更多的 AI 工厂，这些专用于 AI 技术开发和运营的设施正日益扩展到云服务提供商（CSP）和云规模数据中心。



AI 工厂专为实现更高水平的性能而构建，能够训练万亿参数的基础模型，但通常仅包含一个或几个租户。这些工厂需要高度优化的网络，需要将 NVIDIA® NVLink® 高速 GPU 互连与 NVIDIA Quantum InfiniBand 相结合，以实现这种高水平的性能。相比之下，AI 云旨在满足更广泛的用户需求，并具有一些独特的要求，包括集成到以太网服务网络和管理框架的需求。

AI 云的主要特定要求包括：

- > **AI 在软件即服务 (Software as a Service, SaaS) 中的迁移和 AI 即服务 (AI as a Service)：** SaaS 产品越来越多地融入 AI 并提供 AI 即服务，因此需要多租户安全性和多作业性能隔离。
- > **云规模软件定义网络 (SDN)：** 此配置模型 (Provisioning Model) 利用基于以太网的协议和容器编排 (例如 Kubernetes)。SDN 支持从小型集群到大型集群的可扩展配置，并支持将计算和网络资源重新配置到多个小型集群中。
- > **集成和兼容性：** CSP 的 AI 服务需要与其他云服务紧密集成，在云中使用以太网来管理完整数据流水线。
- > **安全性和合规性：** CSP 实施强健的安全措施，包括数据包代理 (Packet Brokering) 和基于以太网的入侵检测/防御设备，以保护客户数据并遵守当地法规。
- > **开源：** 支持 SONiC 等开放网络操作系统。

针对共享云计算基础设施环境进行优化的 CSP 网络对于 AI 云需求来说并不理想。图 1 概述了管理网络/用户访问网络与 AI 网络架构之间的一些差异。

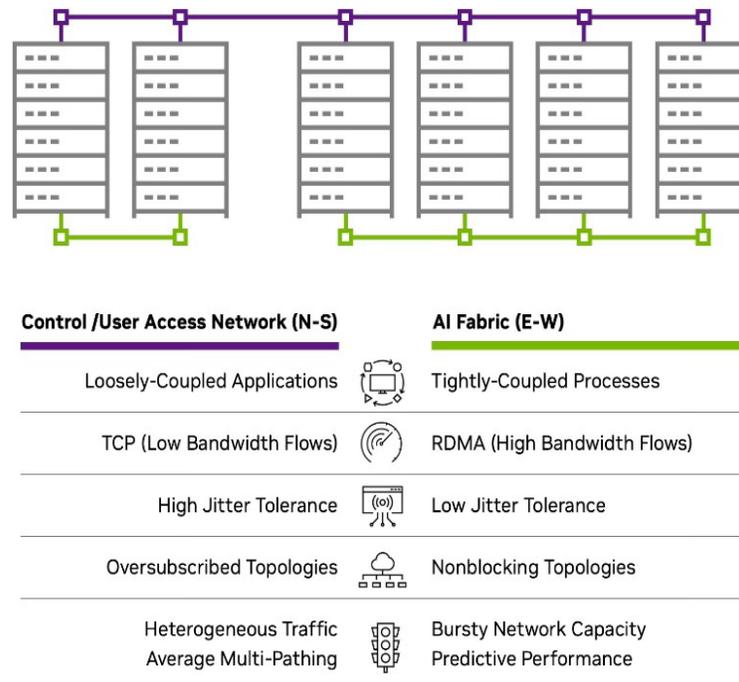


图 1. 管理网络/用户访问网络与 AI 网络架构之间的差异

当 AI 云使用传统的 IP /以太网作为计算网络时，它们只能部分达到经过优化的网络所能实现的

MLPerf 性能水平。在同时运行多个 AI 作业的多租户环境中，传统以太网难以提供性能隔离，这可能会导致一个租户的作业与另一个租户的作业发生干扰。对于使用以太网部署 AI 云的用户，NVIDIA 研发了 Spectrum-X [网络平台](#)来提高性能，同时提高基于以太网的 AI 云的可预测性和能效。

---

# NVIDIA Spectrum-X 网络平台概述

NVIDIA® Spectrum® -X 网络平台是首个专为提高基于以太网的 AI 云性能和效率而设计的以太网平台。这一突破性技术基于开放标准，可为大型 AI 工作负载（如 LLM 和推理）提供明显优于传统以太网解决方案的网络性能。它还提高了能效，并确保在多租户环境中实现一致且可预测的性能。Spectrum-X 利用网络创新来实现 Spectrum-X 以太网交换机与 NVIDIA BlueField®-3 SuperNIC 的紧密结合。这种集成有助于提供专为 AI 工作负载而构建的端到端网络功能，从而缩短基于 Transformer 的大规模生成式 AI 模型的运行时间。因此，网络工程师、数据科学家和云服务提供商可以更快地获得结果并做出明智决策。

## 深入了解 NVIDIA Spectrum-X 网络平台

有效的 AI 计算网络取决于其支持和加速 AI 工作负载的能力。针对从 SuperNIC 到交换机、线缆/光纤、网络和加速软件的网络各个方面，优化都是至关重要的。为了满足这些需求，NVIDIA 创造了许多全新的创新，以在负载和规模下实现更高的有效带宽：

1. Spectrum-X 交换机上的 NVIDIA RDMA over Converged Ethernet (RoCE) 动态路由
2. Spectrum-X SuperNIC 上的 NVIDIA 直接数据放置 (DDP)
3. Spectrum-X 交换机和 SuperNIC 上的 NVIDIA RoCE 拥塞控制
4. NVIDIA AI 加速软件
5. 端到端 AI 网络可视化

这些创新技术作为全栈解决方案的一部分协同工作，已经经过 NVIDIA 的测试、调优和基准测试，可确保提供更高水平的性能。将这些组件集成到 Spectrum-X 网络平台具有许多优势。

## Spectrum-X 的主要优势

- > **提升 AI 云性能：** Spectrum-X 可显著提升多租户 AI 云网络性能。
- > **标准以太网连接：** Spectrum-X 提供完全基于标准的以太网，并且完全可与基于以太网的堆栈互操作。
- > **提高能效：** 通过提高性能，Spectrum-X 有助于打造更节能的 AI 环境。
- > **增强的多租户保护：** 多租户环境中的性能隔离可确保每个租户的工作负载以一致的方式出色执行，从而提高客户满意度并提高服务质量。
- > **更好的 AI 网络架构可见性：** 对 AI 云数据流的可见性有助于识别性能瓶颈，这是现代自动化网络架构验证解决方案的关键。
- > **更高的 AI 可扩展性：** Spectrum-X 在单跳中可扩展至 128 个 400GbE 端口，在采用两层叶脊网络拓扑中可扩展至 8K 个端口，从而支持 AI 云扩展，同时保持高水平的性能。
- > **更快的网络设置：** 高级网络功能的自动化端到端配置已针对 AI 工作负载进行全面调优。

Spectrum-X 网络平台是基于以下关键组件构建的全栈解决方案：

## NVIDIA Spectrum-X 以太网交换机

Spectrum-X 交换机基于 51.2Tbps 的 ASIC 而构建，在单台 2U 交换机中支持多达 64 个 800GbE 端口或 128 个 400GbE 端口。SN5600 树立了新的标准，它是首款专为 AI 工作负载而设计的以太网交换机，通过将专业的高性能、低时延架构与标准以太网连接相集成，可提供无与伦比的效率和可靠性，确保为要求严苛的 AI 应用提供无缝操作和卓越性能。

Spectrum-X 交换机为 AI 提供 RoCE 扩展，并具有独特的增强功能：

- > RoCE 动态路由
- > RoCE 性能隔离
- > 在大规模标准以太网上实现更高的有效带宽
- > 低时延、低抖动和短尾时延

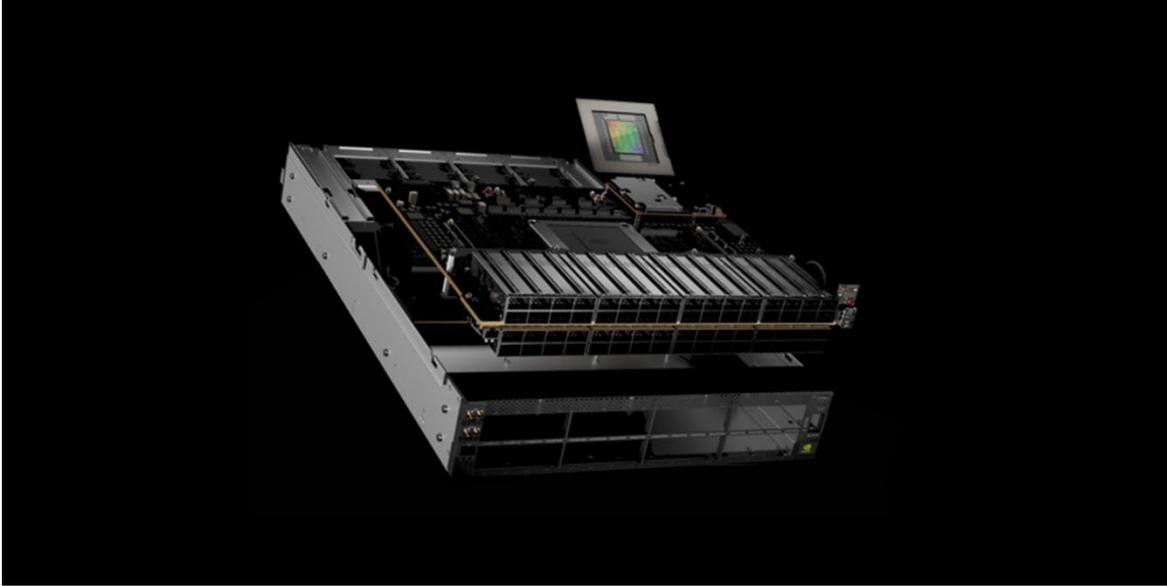


图 2. NVIDIA Spectrum-X SN5600 以太网交换机

## NVIDIA BlueField-3 SuperNIC

该 NVIDIA® BlueField-3® SuperNIC 是第三代片上数据中心基础设施，可助力组织构建从云到核心数据中心再到边缘的软件定义、硬件加速的 IT 基础设施。BlueField®-3 SuperNIC 可在 GPU 服务器之间提供高达 400Gb/s 的 RoCE 网络连接，优化了 AI 工作负载峰值效率，并在作业和租户之间提供确定且隔离的性能。



图 3. NVIDIA BlueField-3 以太网 SuperNIC

BlueField-3 专为 AI 加速而打造，配备适用于 AI 的集成 all-to-all 引擎、NVIDIA GPUDirect<sup>®</sup> 和 NVIDIA<sup>®</sup> Magnum IO GPUDirect<sup>®</sup> Storage (GDS) 加速技术。此外，BlueField-3 SuperNIC 在本地以网卡 (NIC) 模式运行，从而利用本地内存来加速大型 AI 云。这些云需要大量可在本地寻址的队列对 (QP)，而无需使用系统内存。最后，BlueField-3 还采用了 NVIDIA 直接数据放置 (Direct Data Placement, DDP) 技术，可增强 RoCE 动态路由功能。

## NVIDIA 端到端物理层

Spectrum-X 是唯一能够提供从交换机到 SuperNIC 再到 GPU 无缝连接的以太网网络平台。这种端到端集成可确保出色的信号完整性和最低的误码率 (BER)，从而大幅降低 AI 云的功耗。利用 NVIDIA Hopper GPU，Spectrum-X 交换机和 SuperNIC，Spectrum-X 可确保出色的能效和无损的性能。

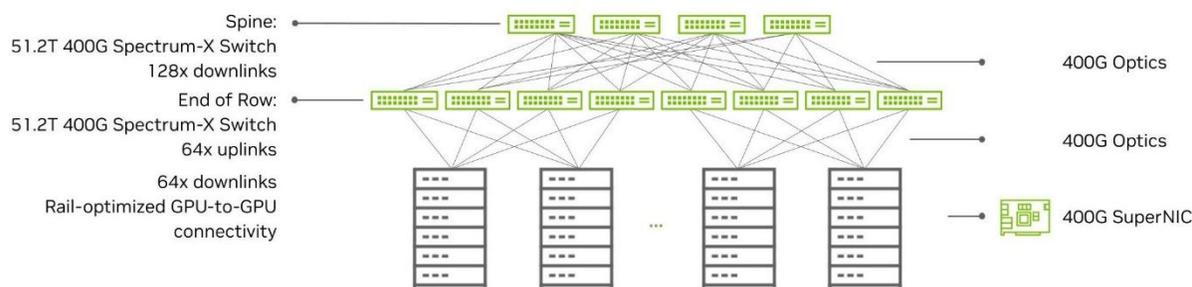


图 4. 典型的 Spectrum-X 网络拓扑

## NVIDIA 加速软件

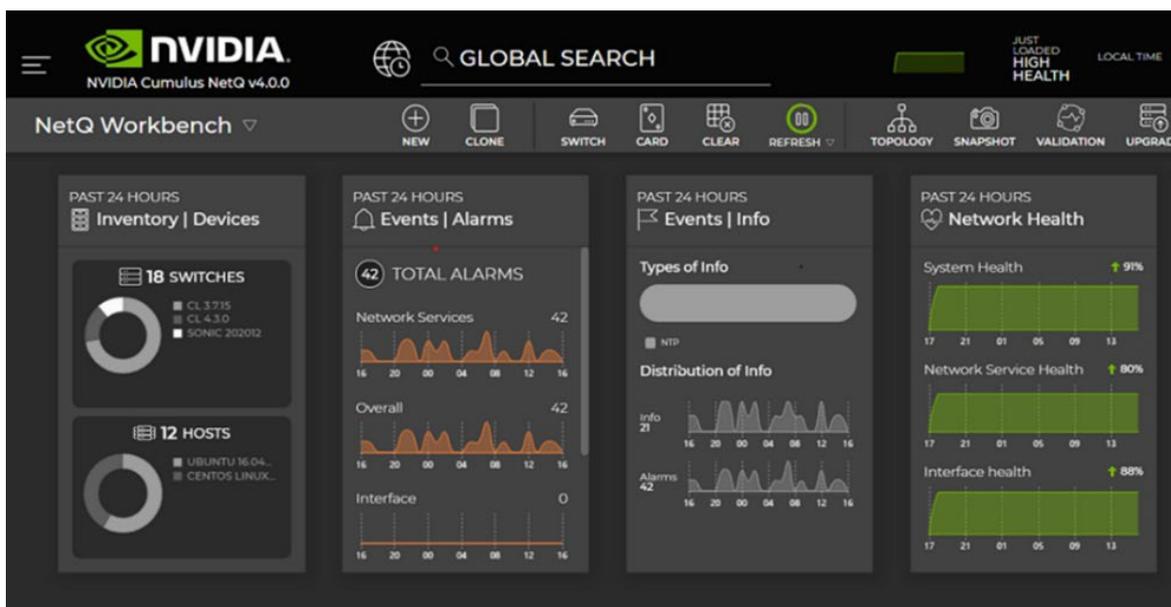


图 5. NVIDIA NetQ 遥测仪表板

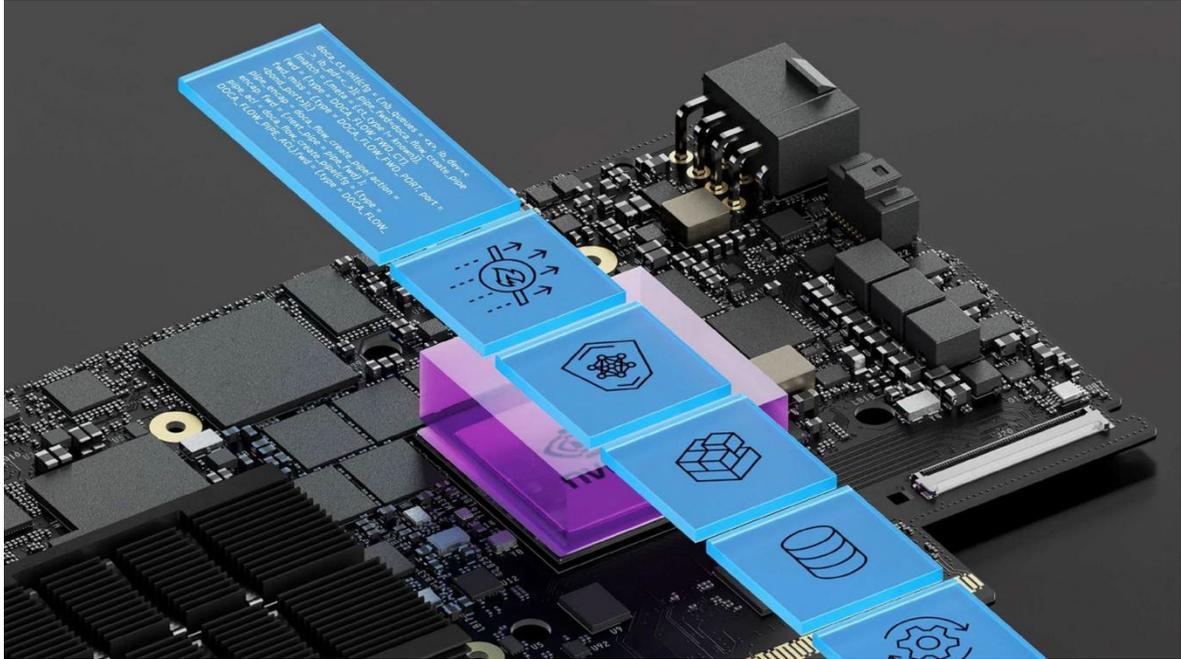
## NetQ

NVIDIA NetQ™ 是一个高度可扩展的网络运营工具集，用于实时 AI 网络可视化、故障排除和验证。NetQ 利用“NVIDIA What Just Happened”交换机遥测数据和 NVIDIA® DOCA 遥测，提供有关 Spectrum-X 交换机和 SuperNIC 运行状况的可行见解，将网络集成到组织的 MLOps 生态系统中。

此外，NetQ 流遥测可映射交换机端口和 RoCE 队列中的流路径和行为，以分析特定的应用程序流。NetQ 对数据包进行采样，分析并报告每台交换机的时延（最大、最小和平均），以及流路径中的缓冲区占用详细信息。NetQ GUI 可报告所有可能的路径、每个路径详细信息和流行为。将 What Just Happened 与流遥测相结合，有助于网络运营商主动识别服务器和应用程序问题的根本原因。

## Spectrum 软件开发套件

NVIDIA 以太网交换机软件开发套件（SDK）具有成熟的可编程性，对于数据包速率、带宽或时延方面在不损失性能的情况下，灵活地实施任何交换和路由功能。借助该 SDK，服务器和网络 OEM 以及网络操作系统（NOS）供应商可以利用以太网交换机系列芯片的高级网络功能，构建灵活、创新且成本优化的交换解决方案。



## NVIDIA DOCA

NVIDIA® DOCA™ 软件框架是释放 NVIDIA BlueField® 网络平台潜力的关键。借助 DOCA，开发者可以创建具有零信任保护的软件定义、云原生、BlueField 加速服务，对未来的数据中心基础设施进行编程，以满足现代数据中心日益增长的性能和安全需求。

# 在以太网上部署 AI 的挑战

## 需要 RDMA 和无损网络

AI 需要 GPU 的出色运行，而为了使多个 GPU 能够出色的协同运行，就需要远程直接内存访问（RDMA）技术。RDMA 是一种网络技术，允许从一台计算机的内存访问另一台计算机，而无需涉及任何一台计算机的 CPU。

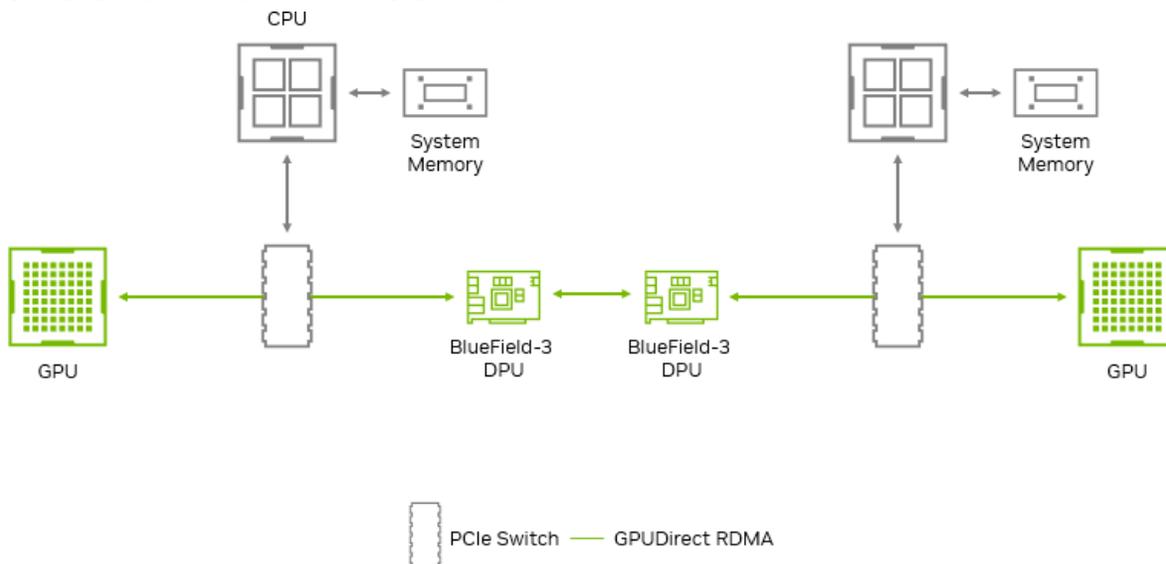


图 6. RDMA 使 BlueField 能够直接访问 GPU 内存

RDMA 使 NVIDIA BlueField 等外围 PCIe 设备能够直接访问 GPU 内存。GPUDirect RDMA 专为 GPU 加速而设计，可跨远程系统实现 NVIDIA GPU 之间的直接通信，从而绕过系统 CPU，且不需要再通过系统内存进行数据缓冲复制。在 RoCE（RDMA Over Converged Ethernet）上运行时，GPUDirect RDMA 当在无损网络上部署时可实现出色性能，这对于确保其可靠运行至关重要。

## 需要 RoCE 动态路由来避免拥塞

可扩展 IP 网络的一个关键属性是它们能够在多个交换机层级之间分配大量流量。

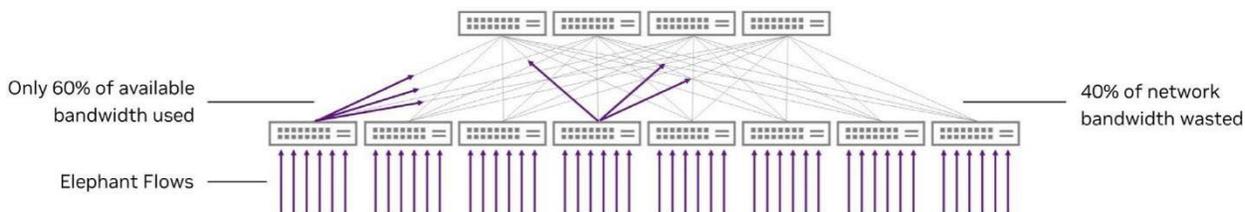


图 7. 采用静态负载均衡，AI 工作负载只使用了最大吞吐量的 60%

在理想环境中，数据流之间是完全不相关的，因此能够在多个网络链路上实现良好的分布式且平稳的负载平衡。这种方法依赖于现代哈希和多路径算法，包括等价多路径（ECMP）。现代数据中心构建在高端口数量、固定外形规格的交换机之上，且广泛利用 ECMP 来构建超大型网络。然而，在很多情况下，通常包括 AI 和存储等无处不在的现代工作负载，ECMP 基于流的哈希算法并不起作用。问题在于在同一路径上发送了过多的大象流会导致有限的熵和相关的哈希冲突，即在同一路径上发送过多的大象流。熵是指测量网络数据包或网络流中数据的随机性，它可以标示协议头中呈现的信息量或可变性。

传统的云服务会产生数千个流，这些流随机连接到全球各地的客户端，这使得云服务网络具有很高的熵。然而，AI 和存储工作负载往往会产生非常大的流，但数量并不多。这些大型 AI 流主导了每个链路的带宽使用，显著减少了流的总数量，并导致网络的熵非常低。这种低熵流量模式也称为大象流（Elephant Flow）分布，在 AI 和高性能存储工作负载中尤为常见。

## 熵的重要性

在传统 IP 路由中，为了在所有可用的等价路径上实现负载均衡，并避免乱序数据包，交换机使用了逐流（通常为 3 元或 5 元组）的静态哈希算法。这需要高的熵才能在多条链路上均匀分布流量，而不会出现拥塞。但是，在大象流的场景中，当熵较低而流量大小很大（在 AI 网络中很常见）时，多个流会被哈希到同一条链路上，从而导致链路超额负载。这种超额负载会导致拥塞、时延增加、数据包丢失和重传等问题，并最终导致应用程序性能不佳。

对于许多应用程序而言，性能不仅取决于网络的平均带宽，还取决于流完成时间的分布，完成时间分布中的长尾或异常值将显著降低应用程序性能。以下是 AI 网络中常见的网络拓扑，它们极有可能会由于长尾延迟而导致性能下降。

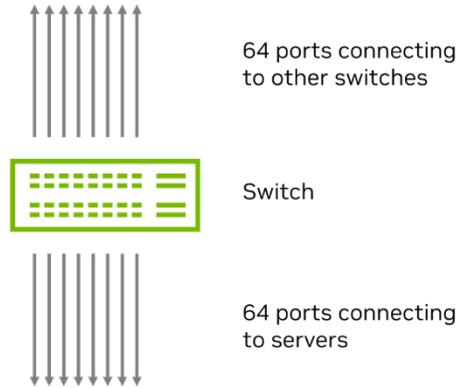


图 8. 为什么即使是非阻塞网络也会出现长尾延迟

图 8 中的示例包含一台具有 128 个 400GbE 端口的叶（Leaf）交换机。

- > 64 个 400GbE 下联端口，连接到服务器
- > 64 个 400GbE 上联端口，连接到脊（Spine）交换机
- > 每个下联端口接收到 4 个相同带宽的流量：每个流 100G，总计 256 个流
- > 上联链路端口的所有流量分配均通过基于流的 ECMP 哈希算法处理

随着带宽水平的提高，出现拥塞的可能性会增加，同时流完成时间也会相应增加。在最坏情况下，流的完成时间可能比理想情况下的时间长 2.5 倍（如图 9 所示）。

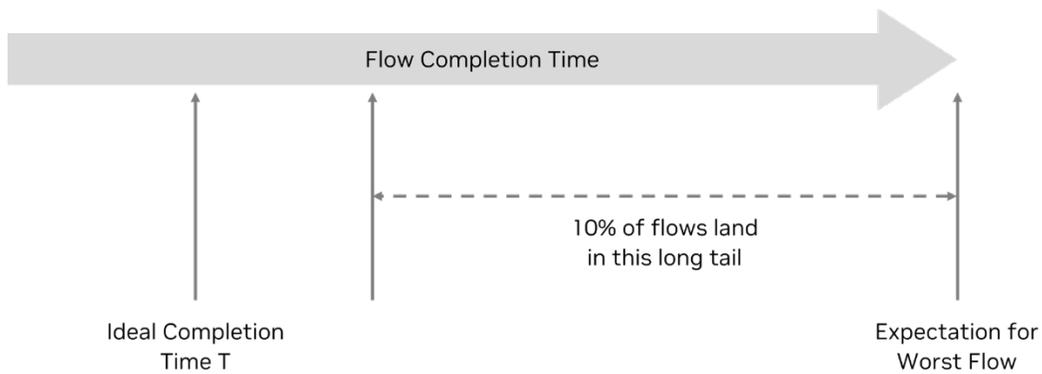


图 9. 流完成时间可能会有显著差异

在这种情况下，一些端口已经出现拥塞，而另一些端口却未被使用。最后一个流（最差流）的预期持续时间是第一个流预期持续时间的 250%。也就是说，存在一个长尾流，其完成时间比预期的要长。为了以高置信度（98%）避免拥塞，您必须将所有流的带宽都减少到 50% 以下。

由于静态 ECMP 哈希处理无法感知带宽，许多流的完成时间都很长，因此交换机上的一些端口高度拥塞，而另一些端口则未得到充分利用。当流完成传输并释放一些端口带宽时，滞后流会通过相同的拥塞端口，从而导致更多的拥塞。这是因为在哈希处理完包头后，转发路径是静态选取的。

## NVIDIA RoCE 动态路由确保负载均衡

RoCE 动态路由可在 Spectrum-X 交换机上使用。借助动态路由，向 ECMP 组的流量转发可选择最不拥塞的端口进行传输。系统会根据出口队列负载评估拥塞，从而确保 ECMP 组时刻保持均衡，无论熵大小如何。向多个服务器发出多个请求的应用程序会在极短时间内接收数据。

对于转发到 ECMP 组的每个数据包，交换机都会选择其出口队列中负载最小的端口进行转发。被评估的队列是那些与数据包的流量类别相匹配的队列。当相同流的不同数据包通过网络的不同路径传输时，它们可能会乱序到达目的地。在 RoCE 传输层，Spectrum-X SuperNIC 负责处理乱序数据包，并按顺序将数据转发给应用程序。这也使得应用程序能够透明的受益于 RoCE 动态路由。

在发送端，SuperNIC 可以动态标记流量，使其符合数据包重新排序的条件，从而确保在需要时可以强制执行数据包间排序。交换机动态路由分类器会对这些标记的 RoCE 数据包进行分类，并应用独特的转发机制。

## 使用 NVIDIA RoCE 拥塞控制

当多个应用程序在 AI Cloud 系统上同时运行时，由于网络拥塞，它们可能会遇到性能不佳和运行时间不一致的情况。这可能是由应用程序本身的网络流量或来自其他应用程序的后台网络流量所引起的。这种类型的拥塞的主要原因是多对一拥塞或 incast 拥塞，其特征是具有多个数据发送端和一个数据接收端。

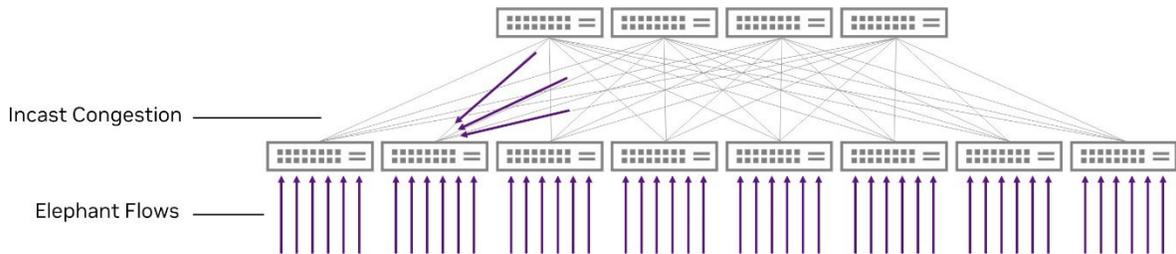


图 10. Incast 需要 NVIDIA RoCE 拥塞控制

遗憾的是，由于无法绕过 RoCE，RoCE 动态路由无法解决 incast 拥塞问题。Spectrum-X 可以使用 NVIDIA RoCE 拥塞控制解决 incast 拥塞问题。

## NVIDIA Spectrum-X 网络平台的主要特性

AI 工作负载就是以少量的大象流为特征，这些大象流负责 GPU 之间的大量数据移动，而其尾部延迟会严重影响整体应用程序性能。使用传统的网络路由机制来承载此类流量模式可能会导致 AI 工作负载的 GPU 性能不一致且未得到充分利用。Spectrum-X RoCE 动态路由是一种细粒度负载均衡技术，它动态地调整 RDMA 数据路由以避免拥塞，并提供优化的负载平衡以实现更高的有效数据带宽。

## NVIDIA RoCE 动态路由的工作原理

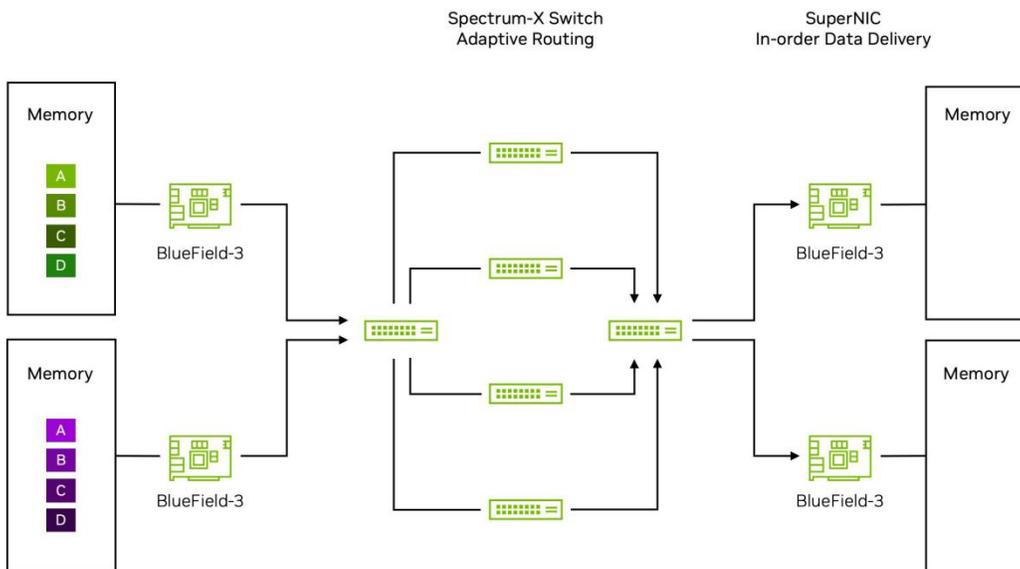
RoCE 动态路由是由 Spectrum-X 交换机和 SuperNIC 联合支持的端到端功能。Spectrum-X 交换机负责为逐个数据包选择最不 congest 的端口进行数据传输。当相同流的不同数据包通过网络的不同路径传输时，它们可能会乱序到达目的地。Spectrum-X SuperNIC 会在 RoCE 传输层处理所有的乱序数据，以透明方式向应用程序提供按序数据。

Spectrum-X 交换机基于出口队列负载来评估拥塞，确保所有端口都保持很好的均衡。Spectrum-X 交换机采用数据包喷洒技术，在其出口队列中为每个网络数据包选择负载最小的端口。Spectrum-X 交换机还接收来自相邻交换机的状态通知，这可能会影响转发决策。参与评估的队列与相应的流量类别相匹配。因此，Spectrum-X 在负载和规模上为超大规模系统提供高达 95% 的有效带宽。

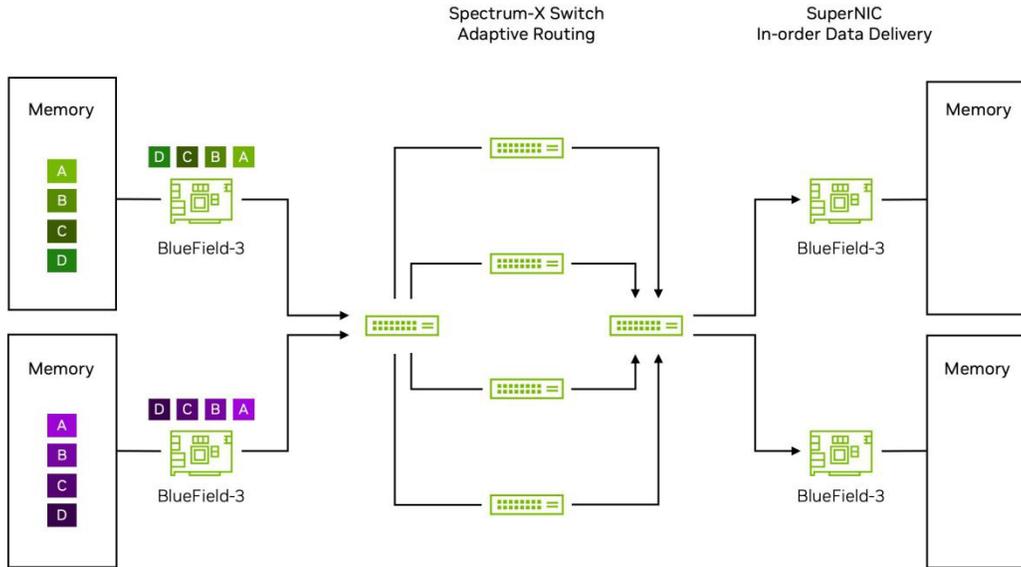
## 借助 NVIDIA 直接数据放置实现 NVIDIA RoCE 动态路由

以下数据包级演示展示了 AI 流如何跨 Spectrum-X 网络在 GPU 之间移动。它描述了 Spectrum-X 交换机上的 RoCE 动态路由如何与 Spectrum-X SuperNIC 上的 NVIDIA 直接数据放置 (Direct Data Placement, DDP) 技术协同工作：

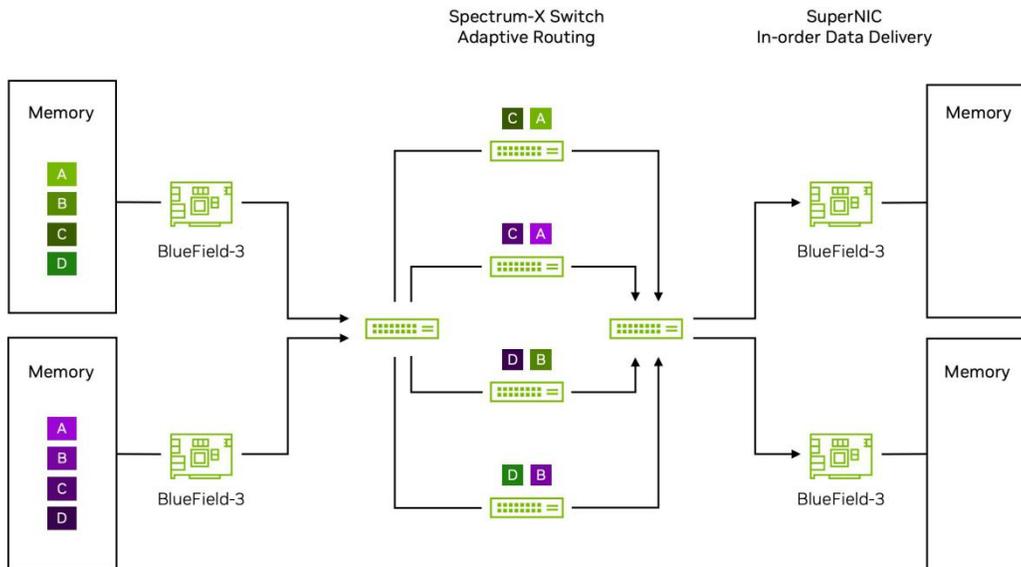
**第 1 步：**数据从图左侧的服务器或 GPU 内存发出，并发送到图右侧的服务器。



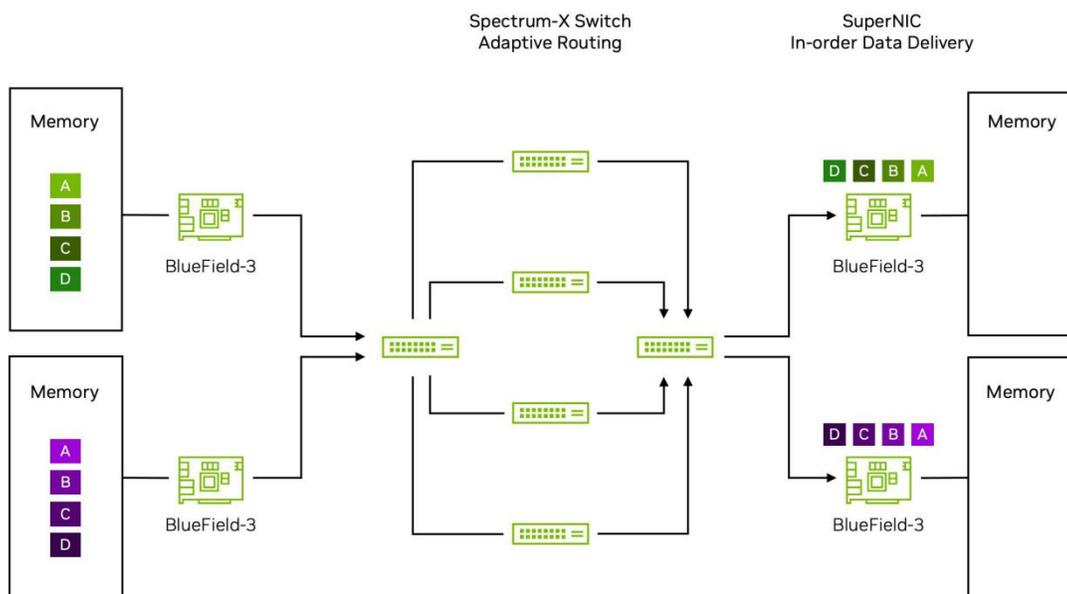
**第 2 步：** Spectrum-X SuperNIC 将数据放入数据包中，并将这些数据包发送到第一台 Spectrum-X 叶交换机，同时将这些数据包标记为适用于 RoCE 动态路由的安全数据包。



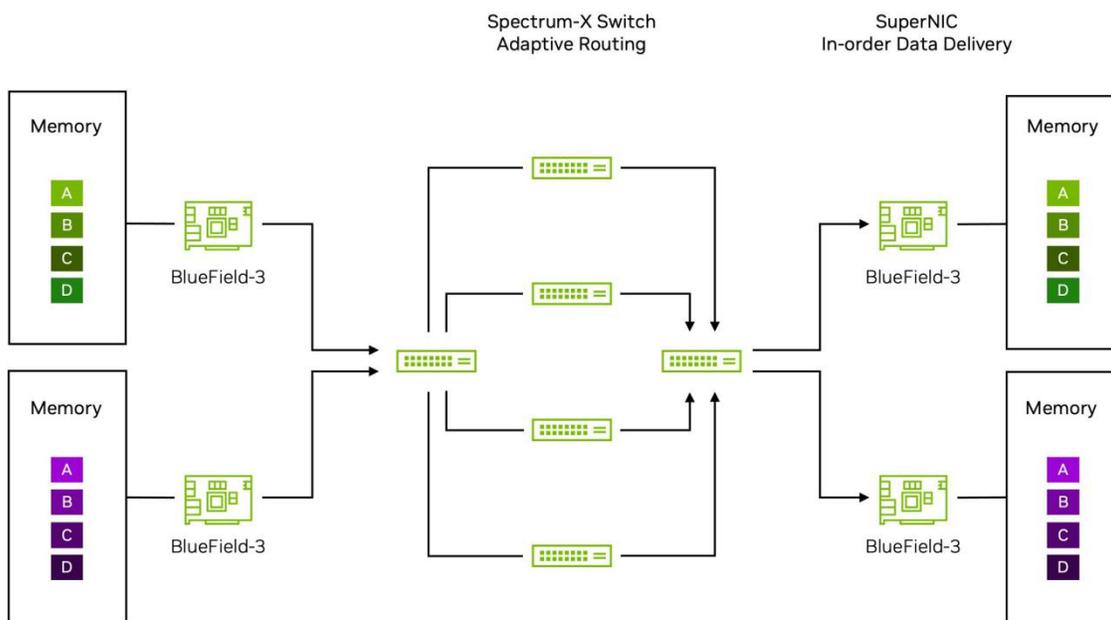
**第 3 步：** 左侧的 Spectrum-X 叶交换机使用 RoCE 动态路由来对来自绿色和紫色数据流的数据包在 4 台 脊交换机上进行负载均衡，从而将每个流的数据包发送到多台脊交换机。这将有效带宽从标准以太网的 60%提高到 Spectrum-X 的 95%。



**第 4 步：**部分数据包在到达右侧的 Spectrum-X SuperNIC 时可能会出现乱序到达的情况。



**第 5 步：**右侧的 Spectrum-X SuperNIC 使用 NVIDIA 直接数据放置（DDP）技术，将数据以正确的顺序放置在主机/ GPU 内存中。



## RoCE 动态路由结果

为了验证 RoCE 动态路由的效果，我们首先使用 RDMA 写入 (Write) 测试程序进行基础测试。在这些测试中，我们将主机分成两对，每对主机在很长一段时间内相互发送大量的 RDMA 写入流。

图 11 显示，基于哈希的静态转发在上联端口上发生冲突，导致流完成时间增加，带宽降低，以及流之间的公平性降低。转为采用动态路由后，所有问题都得到了解决。

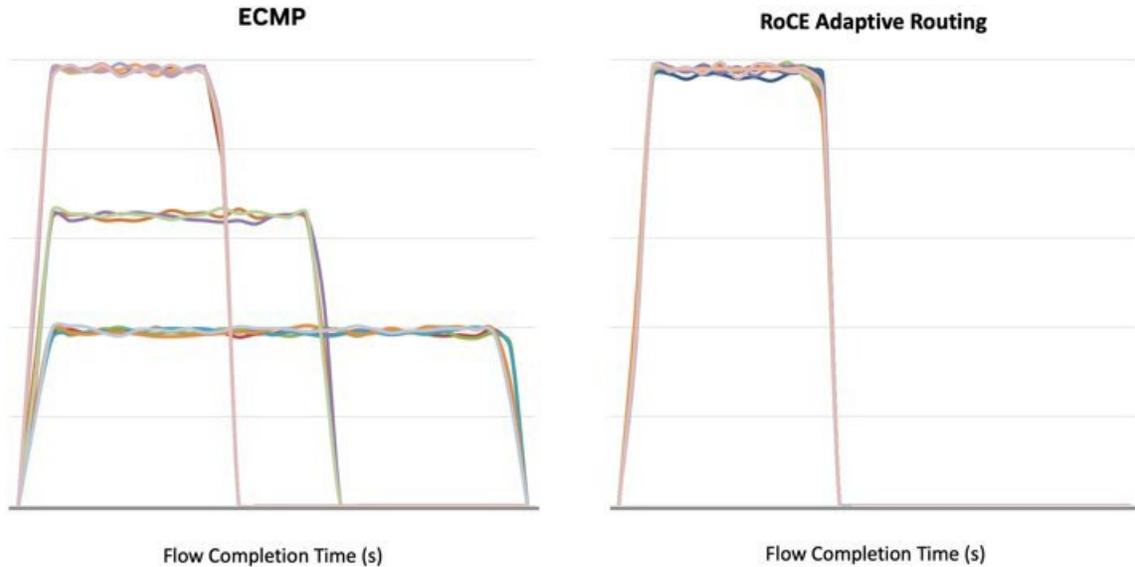


图 11. RoCE 动态路由缩短了流完成时间

在 ECMP 图中，一些流具有相同的带宽和完成时间，而另一些则发生冲突，从而导致更长的完成时间和更低的带宽。具体来说，在 ECMP 场景中，一些流实现了 13 秒的最佳完成时间  $T$ ，而最慢的流则需要 31 秒才完成，大约是最佳时间  $T$  的 2.5 倍。在 RoCE 动态路由图中，所有流都在大致相同的时间内完成，且具有相似的峰值带宽。

## RoCE 动态路由总结

在许多情况下，基于 ECMP 流的哈希会导致高拥塞和不可预测的流完成时间，从而降低了应用程序的性能。

Spectrum-X RoCE 动态路由通过提高网络的实际吞吐量 (goodput) 解决了此问题，最大限度地减少了流完成时间的不可预测性，从而提高了应用程序性能。通过将 RoCE 动态路由与 Spectrum-X SuperNIC 上的 NVIDIA 直接数据放置 (DDP) 技术相结合，使该技术对应用程序保

持透明。这种集成可确保 NVIDIA Spectrum-X 以太网平台提供更大程度数据中心运营所需的加速以太网性能。

## 使用 NVIDIA RoCE 拥塞控制实现性能隔离

由于网络拥塞，在 AI 云系统上并发运行的应用程序可能会遇到性能降低和不可预测的运行时间。这可能是由应用程序自身的网络流量或来自其他应用程序的后台网络流量引起的。这种拥塞的主要原因是多对一拥塞，涉及多个数据发送端和单个数据接收端。

RoCE 动态路由无法解决此类拥塞问题，而是需要对每个端点进行数据流计量。Spectrum-X RoCE 拥塞控制是一种端到端技术，Spectrum-X 交换机提供实时拥塞数据的实时网络遥测信息。这些遥测信息由 Spectrum-X SuperNIC 处理，它可管理和控制数据发送端的数据注入速率，以更大限度地提高网络共享效率。如果没有这种拥塞控制，多对一场景会导致网络反压、拥塞扩散甚至丢包，从而严重降低网络和应用程序的性能。

在拥塞控制过程中，Spectrum-X SuperNIC 执行拥塞控制算法，每秒处理数百万个拥塞控制事件，以微秒响应级别做出细粒度的速率决策。Spectrum-X 交换机带内遥测可提供用于准确拥塞估计的队列信息，以及用于快速恢复的端口利用率指标。NVIDIA 的拥塞控制方法允许遥测数据绕过拥塞流中的排队时延，从而显著缩短拥塞发现和响应时间，同时仍可提供准确的并发遥测数据。

以下示例展示了网络如何经历多对一拥塞，以及 Spectrum-X 如何使用流量计量和带内遥测实现 RoCE 拥塞控制。

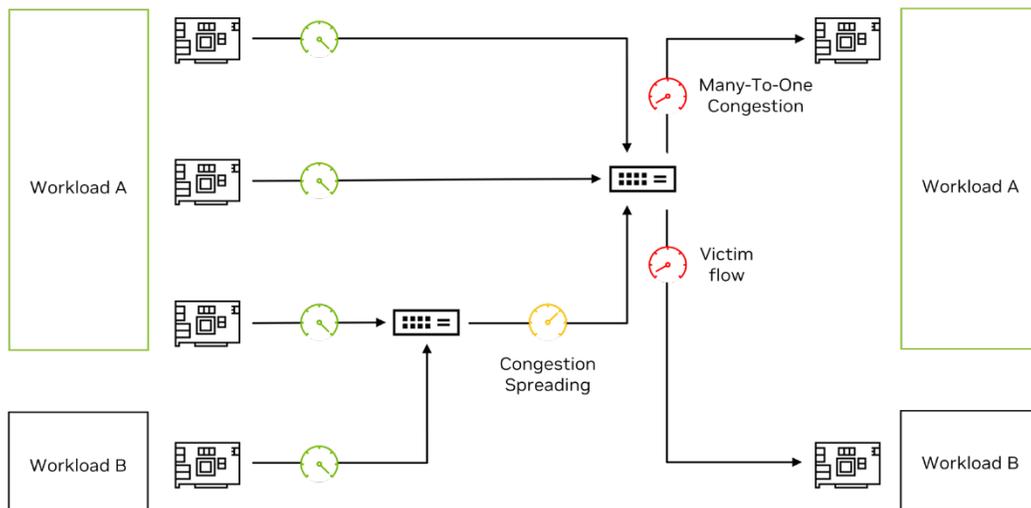


图 12. 拥塞控制造成受害流

图 12 展示了网络拥塞导致受害流 (Victim Flow) 的情况。4 个源 SuperNIC 将数据传输至 2 个目的 SuperNIC。源 1、2 和 3 传输到目的 1，每个源接收占用三分之一的可用链路带宽。

源 4 会通过源 3 相同的叶交换机传输到目的 2，因此接收占用三分之二的可用链路带宽。

在没有拥塞控制的情况下，源 1、2 和 3 会造成对源 4 的拥塞，因为它们都会将数据发送到目的 1。这种拥塞会造成反压，从目的 1 扩散到连接到源 1 和源 2 的叶交换机。源 4 则成为受害流，其到目的 2 的吞吐量降低到可用带宽的 33%（预期性能的 50%）。这对 AI 应用程序性能产生了不利影响，而性能依赖于平均和最差情况。

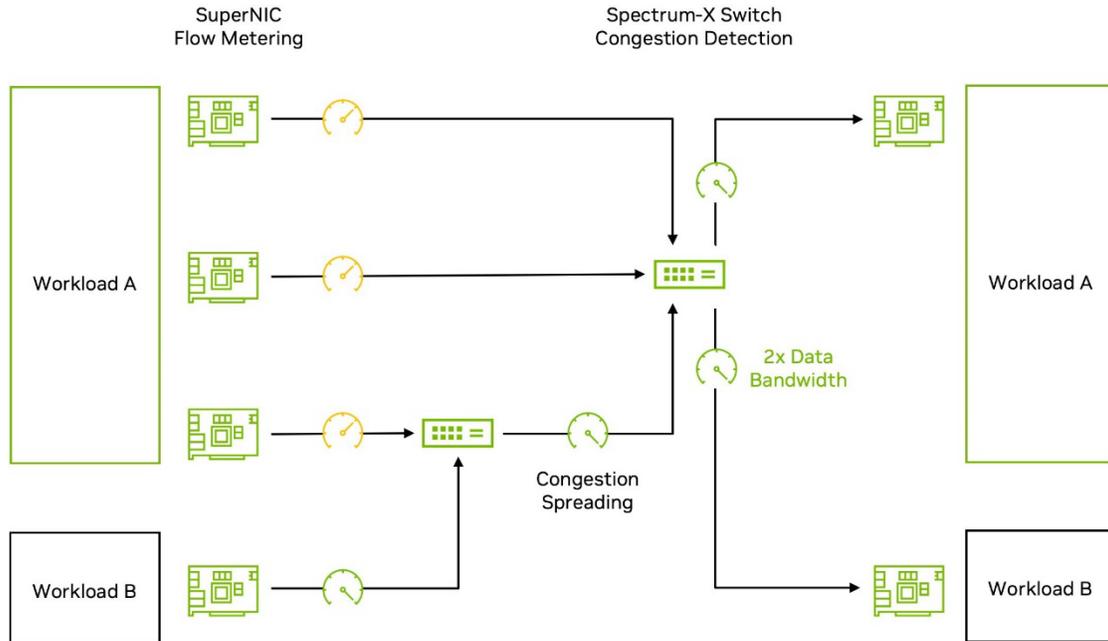


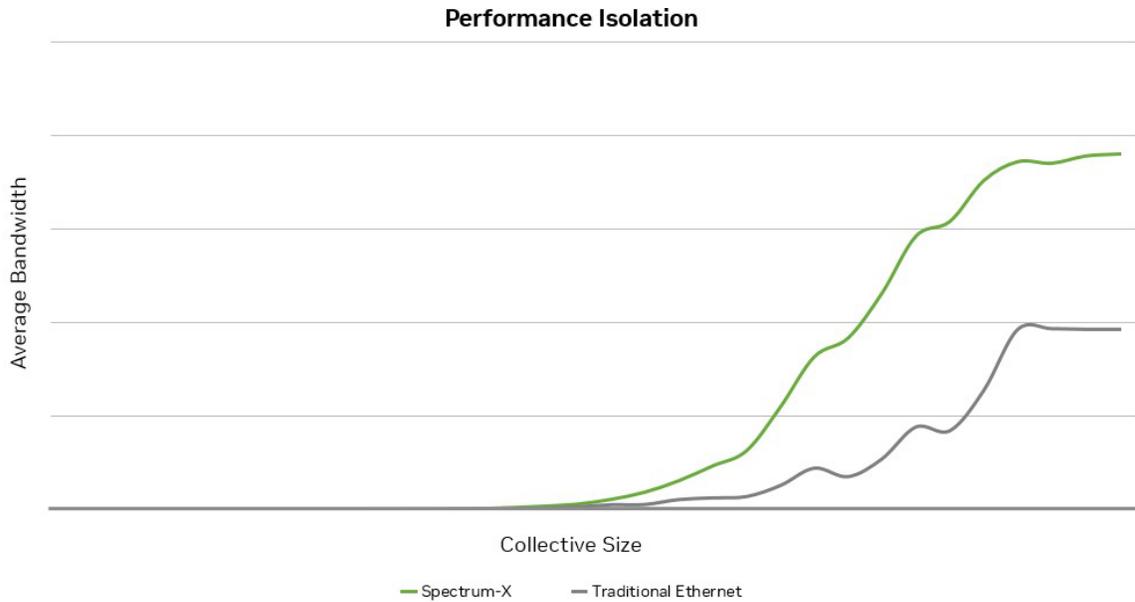
图 13. Spectrum-X 通过流计量和拥塞遥测解决拥塞问题

图 13 显示了 Spectrum-X 如何解决图 12 中的拥塞问题。此图具有相同的测配置：4 个源 SuperNIC 将数据发送到 2 个目的 SuperNIC。在这种情况下，通过在源 1、源 2 和源 3 进行流量计量来避免叶交换机的拥塞。这消除了影响源 4 的反压。源 4 按预期获得完整的三分之二有效带宽。此外，Spectrum-X 交换机使用通过 What Just Happened 生成的带内遥测来动态重新分配流路径和队列行为。

## RoCE 性能隔离

AI 云基础设施需要支持大量用户（租户）和并行应用程序或工作流程。这些用户和应用程序会生成流量，并争夺基础设施的共享资源（如网络），因此可能会相互影响性能。

此外，使用 NVIDIA 集合通信库（NCCL）增强 AI 网络以提高性能，需要协调和同步在云中同时运行的所有工作负载。对于 AI 应用程序而言，云计算的传统优势（例如弹性和高可用性）变得更加有限，因此性能下降成为全球关注的一个更重要的问题。



Spectrum-X delivers 2X Average Bandwidth of Traditional Ethernet

图 14. Spectrum-X 实现噪声隔离机制

Spectrum-X 平台包括多种机制，这些机制共同实现了性能隔离，确保一个工作负载不会对另一个工作负载的性能产生不良影响。这些服务质量机制可防止任何工作负载产生网络拥塞，从而阻碍另一个工作负载的数据移动。

借助 RoCE 动态路由，可通过细颗粒度数据路径均衡来实现这种隔离，从而避免了流经叶脊交换机的数据流发生冲突。借助 RoCE 拥塞控制，可进一步通过计量和遥测来支持隔离，防止因多对一的主机流量而造成受害流。

此外，Spectrum-X 交换机通过其全局共享缓冲区设计来提高性能隔离。这种设计可在不同大小的流之间实现带宽公平，保护工作负载免受“嘈杂邻居”流的影响，并在多个流具有相同目的地端口时实现更大的微突发吸收。

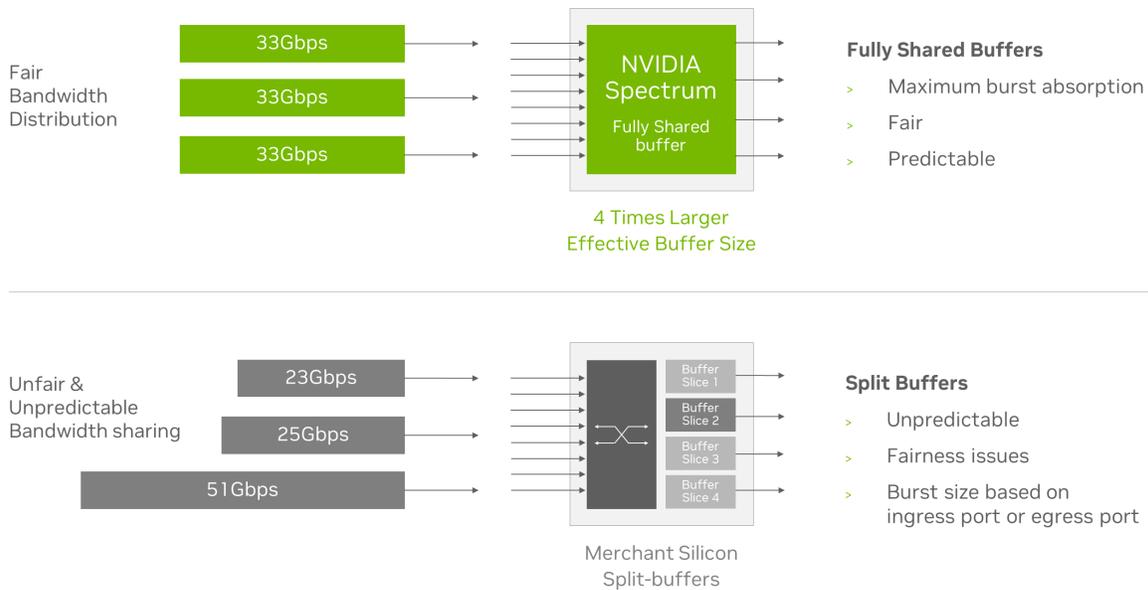


图 15. Spectrum-X 交换机采用全局共享缓冲区设计

## NVIDIA 全栈优化

AI 云是一种精密设计的机器，其性能可能会受到看似轻微的网络事件（例如链路或设备故障）的不良影响，而这些事件可能会在云控制网络上被忽视。单个迟缓的系统会产生连锁效应，从而减慢整个云的速度。此外，如果网络缺乏能够检测和缓解此类行为的警报监控系统，一个不稳定的 SuperNIC 就会干扰到相邻的设备。

NVIDIA 遵循多阶段流程来测试、认证和调优作为全栈 AI 解决方案一部分的 Spectrum-X 组件：

- 1. AI 性能基准测试：** NVIDIA 测试典型集群的整体 AI 性能，然后发布性能结果，以便其他人了解使用 Spectrum-X 网络构建的 AI 集群的预期性能水平。
- 2. 组件测试：** NVIDIA 首先对每个组件（交换机、SuperNIC、GPU 和 AI 库）进行单独测试，以确保它们均按预期运行并符合性能基准。
- 3. 集成测试：** 在验证各个组件的功能后，NVIDIA 会对其进行集成，以创建紧密结合的 AI 解决方案。集成系统经过一系列测试，以确保组件之间的兼容性、互操作性和无缝通信。
- 4. 性能调优：** 在组件集成并作为一个单元运行后，NVIDIA 专注于优化全栈解决方案的性能。这涉及调整参数、识别瓶颈和微调配置，以更大限度地提高网络的有效带宽。

- 5. 整体系统性能：**在此阶段，NVIDIA 还专门调整缓冲区和拥塞阈值，以满足 GPT、BERT 和 RetinaNet 等 AI 工作负载的需求，确保这些流行的深度学习模型获得出色性能。
- 6. 库和软件优化：**NVIDIA 优化了 NCCL 等 AI 库，以确保 GPU 与其他组件之间的高效通信。这一步骤对于在大规模深度学习应用程序中更大限度地减少时延和提高吞吐量是至关重要的。
- 7. 认证：**在测试和调整全栈 AI 解决方案后，NVIDIA 会执行一系列认证，以确保系统安全可靠地运行。该过程包括压力测试、安全测试，以及验证与热门 AI 框架和应用的兼容性。
- 8. 真实测试：**最后，NVIDIA 在真实场景中部署全栈 AI 解决方案，以评估其在各种条件和工作负载下的性能。此步骤有助于识别任何不可预见的问题，并确保解决方案随时可供客户广泛采用。

通过遵循这一全面流程，NVIDIA 可确保我们的全栈 AI 解决方案的稳健性、可靠性和高性能，为其客户提供无缝体验，尤其是对于 GPT、BERT 和 RetinaNet 等广泛使用的 AI 工作负载。

## 从参考架构获得的 Spectrum-X 性能结果

NVIDIA 的全栈优化为 AI 工作负载带来了显著的性能优势。为了全面调整和优化 Spectrum-X 推荐的配置，NVIDIA 使用一台生成式 AI 超级计算机。

这台生成式 AI 超级计算机由 4 个可扩展的 HGX 系统单元组成，这些系统单元配备了 B3140 SuperNIC，并通过三层 SN5600 以太网交换机连接，旨在为大规模 AI 云绘制蓝图，NVIDIA 客户和合作伙伴可以复制这些云，以便为自己的工作负载实现相同的性能。

以下基准测试将完全优化的 Spectrum-X 网络与针对 AI 优化的传统以太网的性能进行了比较。这种传统以太网网络利用 RoCE 和标准以太网流量控制技术，但没有利用 Spectrum-X 引入的 RoCE 扩展功能和 NCCL 感知网络优化。

### RDMA 网络性能

AI 工作负载及其网络流量模式因模型而异，但所有模型都需要高效、可靠和高性能的 RoCE 网络。第一个基准测试结果显示了 RDMA 二分法的有效带宽和时延。与传统以太网相比，Spectrum-X 提供了显著的性能（更高的有效带宽和更低的时延）。

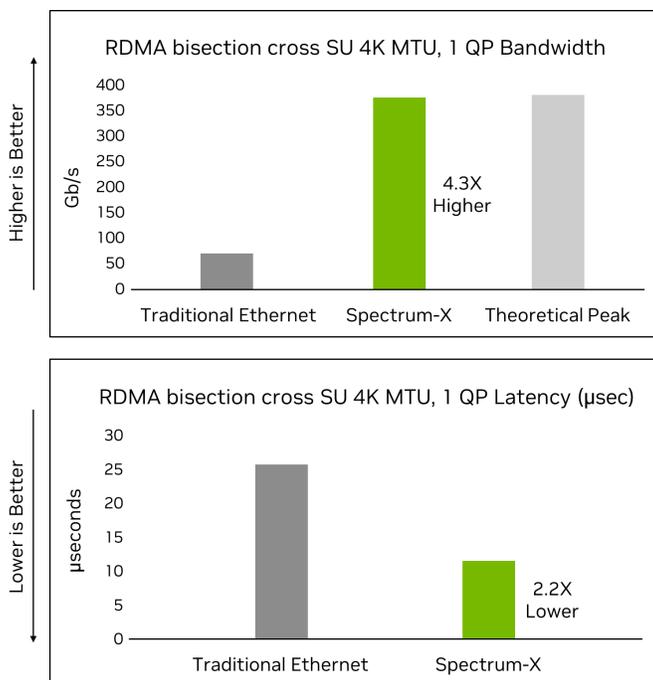


图 16. Spectrum-X 为 RDMA 二分法实现更高的有效带宽和更低的时延

## NCCL 集合性能

all-reduce 和 all-to-all 等 NCCL 集合对于分布式 AI 性能至关重要，也是 AI 训练的原子操作。Spectrum-X 被设计为 NCCL 感知网络架构，可确保对这些操作进行适当调整。以下基准测试结果显示了 AI 云场景中的 NCCL 操作性能，即在云环境中有多作业同时运行并产生噪声。Spectrum-X 为 all-reduce 和 all-to-all 提供更高的 NCCL 总线带宽，同时提供噪声隔离，从而实现几乎与“无噪点”场景（即没有其他作业同时在云上运行）中测量的性能一样高的一致性。

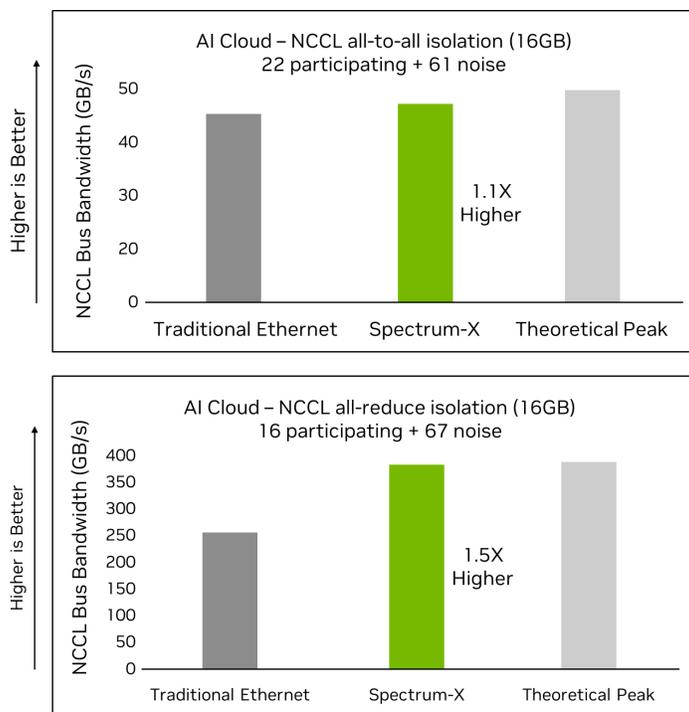


图 17. Spectrum-X 为 AI 云中运行的 NCCL 集合提供更高的总线带宽

## AI 模型性能

网络优化最重要的目标是提供更高的应用程序级性能。对于 LLM 训练等 AI 工作负载，这意味着减少迭代步骤时间，从而缩短作业完成时间并加快 AI 模型的价值实现。NVIDIA 针对几个常见的大语言模型（包括 GPT-4、FSDP Llama 70B 和 NeMo LLM 43B）对 AI 云训练迭代时间进行了基准测试。在所有情况下，Spectrum-X 都缩短了训练时间，并显示出一致且可预测的性能，且不受噪音影响。

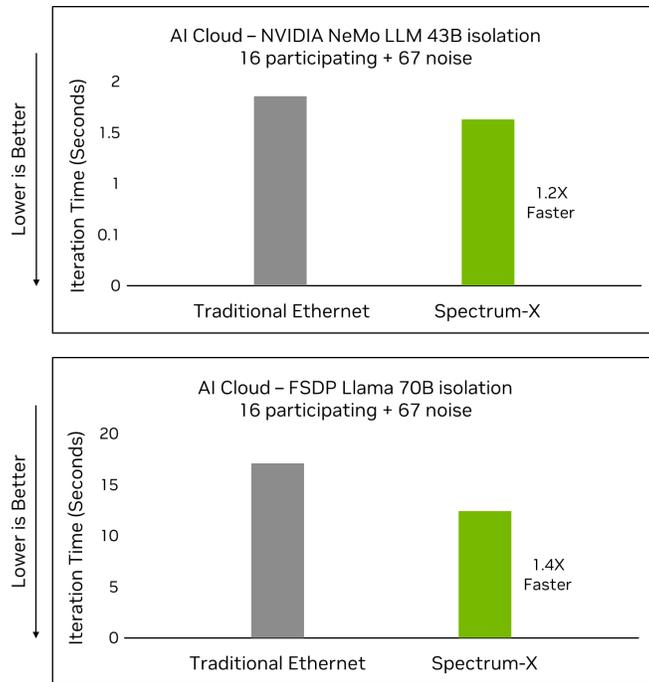


图 18. Spectrum-X 在嘈杂的 AI 云环境中加速训练 AI 工作负载

## 虚拟化对性能的影响

作为多租户架构，AI 云通常需要通过 EVPN VXLAN 等方法实现网络虚拟化。这些虚拟化技术侧重于优化可扩展性和可扩展性，而非性能。如果不仔细实施，虚拟化可能会带来严重的性能损失。

NVIDIA 通过 EVPN VXLAN 比较了支持和不支持虚拟化的 Spectrum-X 性能。在所有测试中，单租户和多租户环境的性能几乎相同，并且表现出极小的下降，这表明 Spectrum-X 是多租户 AI 云的理想以太网网络架构。

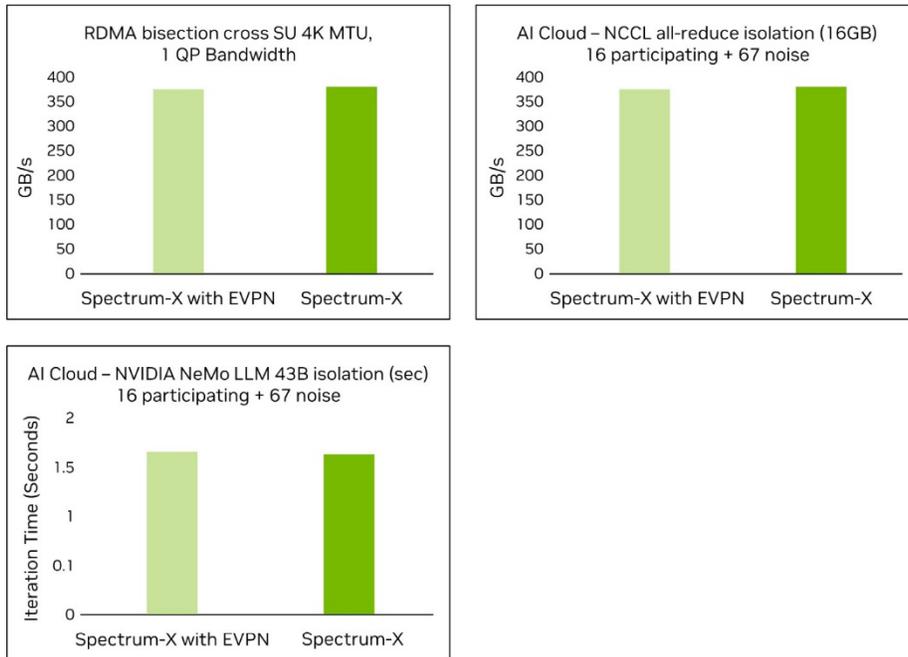


图 19. Spectrum-X 在使用和不使用网络虚拟化的情况下性能表现同样出色

## 专为 AI 设计的以太网

Spectrum-X 网络平台专为要求严苛的 AI 应用程序而构建，相较于传统以太网，可提供一系列优势。凭借更高的性能、更低的功耗、更低的总体拥有成本、无缝的全栈软件和硬件集成以及强大的可扩展性，Spectrum-X 成为当前和未来 AI 工作负载的终极以太网平台。

## Notice

This document is provided for information purposes only and shall not be regarded as a warranty of a certain functionality, condition, or quality of a product. NVIDIA Corporation ("NVIDIA") makes no representations or warranties, expressed or implied, as to the accuracy or completeness of the information contained in this document and assumes no responsibility for any errors contained herein. NVIDIA shall have no liability for the consequences or use of such information or for any infringement of patents or other rights of third parties that may result from its use. This document is not a commitment to develop, release, or deliver any Material (defined below), code, or functionality.

NVIDIA reserves the right to make corrections, modifications, enhancements, improvements, and any other changes to this document, at any time without notice.

Customer should obtain the latest relevant information before placing orders and should verify that such information is current and complete.

NVIDIA products are sold subject to the NVIDIA standard terms and conditions of sale supplied at the time of order acknowledgement, unless otherwise agreed in an individual sales agreement signed by authorized representatives of NVIDIA and customer ("Terms of Sale"). NVIDIA hereby expressly objects to applying any customer general terms and conditions with regards to the purchase of the NVIDIA product referenced in this document. No contractual obligations are formed either directly or indirectly by this document.

NVIDIA products are not designed, authorized, or warranted to be suitable for use in medical, military, aircraft, space, or life support equipment, nor in applications where failure or malfunction of the NVIDIA product can reasonably be expected to result in personal injury, death, or property or environmental damage. NVIDIA accepts no liability for inclusion and/or use of NVIDIA products in such equipment or applications and therefore such inclusion and/or use is at customer's own risk.

NVIDIA makes no representation or warranty that products based on this document will be suitable for any specified use. Testing of all parameters of each product is not necessarily performed by NVIDIA. It is customer's sole responsibility to evaluate and determine the applicability of any information contained in this document, ensure the product is suitable and fit for the application planned by customer, and perform the necessary testing for the application in order to avoid a default of the application or the product. Weaknesses in customer's product designs may affect the quality and reliability of the NVIDIA product and may result in additional or different conditions and/or requirements beyond those contained in this document. NVIDIA accepts no liability related to any default, damage, costs, or problem which may be based on or attributable to: (i) the use of the NVIDIA product in any manner that is contrary to this document or (ii) customer product designs.

No license, either expressed or implied, is granted under any NVIDIA patent right, copyright, or other NVIDIA intellectual property right under this document. Information published by NVIDIA regarding third-party products or services does not constitute a license from NVIDIA to use such products or services or a warranty or endorsement thereof. Use of such information may require a license from a third party under the patents or other intellectual property rights of the third party, or a license from NVIDIA under the patents or other intellectual property rights of NVIDIA.

Reproduction of information in this document is permissible only if approved in advance by NVIDIA in writing, reproduced without alteration and in full compliance with all applicable export laws and regulations, and accompanied by all associated conditions, limitations, and notices.

THIS DOCUMENT AND ALL NVIDIA DESIGN SPECIFICATIONS, REFERENCE BOARDS, FILES, DRAWINGS, DIAGNOSTICS, LISTS, AND OTHER DOCUMENTS (TOGETHER AND SEPARATELY, "MATERIALS") ARE BEING PROVIDED "AS IS." NVIDIA MAKES NO WARRANTIES, EXPRESSED, IMPLIED, STATUTORY, OR OTHERWISE WITH RESPECT TO THE MATERIALS, AND EXPRESSLY DISCLAIMS ALL IMPLIED WARRANTIES OF NONINFRINGEMENT, MERCHANTABILITY, AND FITNESS FOR A PARTICULAR PURPOSE. TO THE EXTENT NOT PROHIBITED BY LAW, IN NO EVENT WILL NVIDIA BE LIABLE FOR ANY DAMAGES, INCLUDING WITHOUT LIMITATION ANY DIRECT, INDIRECT, SPECIAL, INCIDENTAL, PUNITIVE, OR CONSEQUENTIAL DAMAGES, HOWEVER CAUSED AND REGARDLESS OF THE THEORY OF LIABILITY, ARISING OUT OF ANY USE OF THIS DOCUMENT, EVEN IF NVIDIA HAS BEEN ADVISED OF THE POSSIBILITY OF SUCH DAMAGES. Notwithstanding any damages that customer might incur for any reason whatsoever, NVIDIA's aggregate and cumulative liability towards customer for the products described herein shall be limited in accordance with the Terms of Sale for the product.

## Trademarks

NVIDIA, the NVIDIA logo, BlueField, GPUDirect, MagnumIO GPUDirect Storage, and Spectrum are trademarks and/or registered trademarks of NVIDIA Corporation and/or its affiliates in the U.S. and other countries. Other company and product names may be trademarks of the respective companies with which they are associated. All other trademarks are property of their respective owners.

## Copyright

© 2024 NVIDIA Corporation & Affiliates. All rights reserved. JULY24