



NVIDIA Spectrum-X

专为以太网 AI 云而打造。

NVIDIA Spectrum™-X 是首个专为提高基于以太网的 AI 云性能和效率而设计的平台。这项突破性技术可将整体 AI 性能和能效提高 1.7 倍，并在多租户环境中实现一致且可预测的性能。Spectrum-X 依托于 NVIDIA Spectrum™-4 以太网交换机与 NVIDIA® BlueField®-3 DPU（数据处理器）的紧密耦合所支持的网络创新而构建。所提供的端到端 AI 功能可缩短基于 Transformer 的大规模生成式 AI 模型的运行时间，并使网络工程师、AI 数据科学家和云服务提供商获得准确的结果并更快地做出明智的决策。

传统以太网在 AI 方面的挑战

使用传统以太网构建其计算网络架构的 AI 云，只能通过优化的网络实现部分 MLPerf 性能。在具有多个 AI 任务同时运行的多租户环境中，性能隔离对于防止性能进一步下降是至关重要的。如果出现链路故障，传统以太网网络架构可能会导致集群的 AI 性能减半。这是因为传统以太网主要是针对日常企业工作流程进行了优化，而不是为满足依赖于 NVIDIA 集合通信库（NCCL）的高性能 AI 应用程序的需求而设计的。

这些性能问题是由传统以太网的固有因素造成的：

- > 更高的交换机时延，这在商用 ASIC 中十分常见
- > 分离缓冲区交换机架构，这可能会导致带宽不公平
- > 针对 AI 工作负载生成的大型流进行次优化的负载均衡
- > 性能不能隔离和相邻噪声问题

基于 Spectrum-4 以太网交换机和 BlueField-3 DPU 的 Spectrum-X 网络平台可以解决传统以太网的这些问题。

主要特性

NVIDIA Spectrum-X 网络平台组件

- > Spectrum-4 以太网交换机 (SN5000 系列)
- > BlueField-3 DPU
- > NVIDIA LinkX® 线缆和收发器
- > Spectrum-X 许可

Spectrum-X 支持 RoCE 扩展功能

- > RoCE 动态路由
- > RoCE 拥塞控制
- > RoCE 性能隔离

NVIDIA Spectrum-X 的主要优势

- > **提升AI云性能：** Spectrum-X 可将 AI 云性能提升至 1.7 倍或更高，从而加速 AI 工作负载的处理、分析和执行的环节，进而加速 AI 解决方案的开发和部署。
- > **标准以太网连接：** Spectrum-X 采用 NVIDIA 创新技术，完全兼容标准的以太网，可与传统以太网的堆栈完全互操作。
- > **提高能效：** 通过提升性能，Spectrum-X 有助于打造更节能的 AI 环境，从而降低 AI 集群的功耗和运营成本。
- > **增强多租户性能：** 多租户环境中的性能隔离可以确保每个租户的工作负载都能持续出色地执行，从而提高客户满意度和服务质量。
- > **更好的AI网络架构可视性：** 对 AI 云中的数据流具有可视的特性有助于识别性能瓶颈，是现代自动化网络架构验证解决方案的关键部分。
- > **更高的AI可扩展性：** 在单跳中可扩展至 256 个 200G 端口，或在两层叶脊网络拓扑中可扩展至 16K 个端口，从而支持 AI 云的扩展需求，同时保持高水平的性能，使其成为满足不断发展的 AI 基础设施需求的理想解决方案。
- > **更快的网络设置：** 完全针对 AI 工作负载进行端到端调优的高级网络功能可以实现自动化配置。
- > **更高的弹性：** 借助更高的弹性，可以消除丢失链路时出现的性能级联问题，从而将带宽损失限制在单个链路上。
- > **安全的基础设施：** 借助 BlueField-3 DPU，Spectrum-X 可支持加密和深度数据包检测等功能，确保控制平面的安全。

提高能效

由于对计算资源的需求与日俱增，并需要控制能源成本，在数据中心的常见做法就是设置功率上限。总体上，提高每瓦的 AI 性能对在功率预算范围内实现更高的计算效率是至关重要的。这在深度学习等应用中尤为重要，因为在这些应用中，训练模型可能需要进行大量的计算，从而消耗大量电力。除了强大的性能优势外，Spectrum-X 还优于其他以太网解决方案，提供 1.7 倍的卓越能效（每瓦的性能）。

优化 NCCL 的网络

在拥有 Spectrum-X 许可的前提下，Spectrum-4 交换机和 BlueField-3 DPU 可以紧密协作。它们共同构成了**优化 NCCL 的网络架构**，旨在使用一套端到端的创新技术来优化 AI 集群性能：

- > **RoCE 动态路由**通过动态路由使得 AI 中的大象流远离拥塞点，从而避免发生拥塞。这种方法提高了网络资源的利用率。Spectrum-4 交换机采用细颗粒度负载均衡，灵活改变活动数据流的路由以消除拥塞。此外，BlueField-3 DPU 可协同工作来处理乱序数据包，并在目标内存中按顺序放置数据包。RoCE 动态路由支持客户自定义配置以实现高效配置和自动化部署。
- > **RoCE 拥塞控制**借助带内网络遥测收集网络性能数据。BlueField-3 DPU 使用从交换机遥测收集到的数据来优化网络数据速率。BlueField 算法使用深度学习模型进行数据计量、优化多任务、多租户系统的设置。
- > **借助 NVIDIA NetQ™ 的端到端可视性**追踪从 GPU 到 DPU 的流级性能，并呈现交换机端口以及 RoCE 队列中的路径和每跳行为。
- > **NVIDIA 全栈集成**包括 NVIDIA NetQ、NCCL、Nsight™、H100 Tensor Core GPU、BlueField-3 DPU 和 Spectrum-4 交换机，所有这些都可以通过 NVIDIA Bright Cluster Manager 进行配置，以实现简单的生产就绪部署并加快 AI 的实现。
- > **超低时延**对于需要实时处理的 AI 和机器学习工作负载是至关重要的。Spectrum-4 交换机可以提供业内超低时延的 200/400GbE 交换，确保 256 个端口的超低时延和抖动。
- > **优化 NCCL** 交换机行为阈值可以为 AI 工作负载优化缓冲区和拥塞阈值。这确保了同步集合通信操作，减少了拥塞和数据包丢失的可能性。

Spectrum-X 为构建基于以太网的多租户 AI 云的组织提供了突破性解决方案。Spectrum-X 增强了各种应用中 AI 云的性能和能效，从而实现了更高的可预测性和一致性。这可以加快上市时间并增强竞争优势。

NVIDIA Spectrum-X 网络平台

Spectrum-X 网络平台由如下组件组成。

Spectrum-4 以太网交换机设计用于智能叶交换机、脊交换机和超级脊交换机，在密集的 2U 外形规格中提供 64 个 800GbE 端口，并在 NVIDIA Spectrum-X 部署中发挥着至关重要的作用。SN5600 交换机支持标准叶和脊设计，并具有 TOR (top-of-rack) 交换机和多轨优化的 EOR (end-of-row) 拓扑。SN5600 提供 10-800GE 组合的多种连接，并拥有业界领先的 51.2Tb/s 的总带宽。

BlueField-3 DPU 提供以 400Gb/s 运行的高性能计算，并对软件定义网络、存储和网络安全进行线速处理。BlueField-3 DPU 结合了强大的计算能力、高速网络和广泛的可编程性，为 Spectrum-X 提供加速 RoCE，以及对控制平面至关重要的多租户和零信任安全性。

LinkX 收发器和线缆提供一些列可靠的直连铜缆 (DAC)、有源铜缆 (ACC)、有源光缆 (AOC)，以及用于连接 OSFP 和 QSFP 所需的光纤。Spectrum-4 交换机与 LinkX 产品线相结合，可实现市场上高能效的 AI 网络架构。

Spectrum-X 许可释放了 Spectrum-4 交换机和 BlueField-3 DPU 的全部功能，实现了 RoCE 扩展功能和优化网络参数设置，从而提供增强的 AI 性能。

Spectrum-X 网络平台订购规格

有关订购 Spectrum-4 交换机、BlueField-3 DPU 以及 LinkX 线缆和收发器的更多信息，请参阅 [SN5000 交换机 datasheet](#) 和 [BlueField-3 DPU datasheet](#)。

Spectrum-X 许可可以实现在以太网上的生成式 AI 所需的软件定义、硬件加速功能集。如果没有许可，这些功能（包括 RoCE 扩展）将不可用。

对于使用标准的、开箱即用 RoCE 的部署，不需要 Spectrum-X 许可。Spectrum-X 许可按照每台交换机进行许可授权，从而解锁从交换机到 DPU、再到 GPU 的端到端无损网络传输价值。

Spectrum-X 许可的部件号 (part number) 是绑定既有支持服务的。包括业务标准型 (business-standard) 的和业务关键型 (business-critical) 的服务级别协议 (SLA) 分别提供 3 年和 5 年的支持服务。建议 Spectrum-X 许可的支持级别与基本交换机硬件支持级别保持一致。也就是说，业务关键型 Spectrum-X 支持应与业务关键型硬件支持配对，反之亦然。如需动态续订部件号，请联系您的 NVIDIA 销售代表。

Spectrum-X 许可适用于 NVIDIA Cumulus® Linux 和 ONIE (Open Network Install Environment) Spectrum-4 交换机部署。许可可以在购买交换机时购买，也可以在需要 Spectrum-X 功能时单独购买来添加许可。

NVIDIA Spectrum-X 许可订购信息

OPN	描述
798-XSPEXZ-P3CMI36	Spectrum-X 许可订阅，提供为期 3 年的业务标准型支持
798-XSPEXZ-P3CMI60	Spectrum-X 许可订阅，提供为期 5 年的业务标准型支持
798-YSPEXZ-P3CMI36	Spectrum-X 许可订阅，提供为期 3 年的业务关键型支持
798-YSPEXZ-P3CMI60	Spectrum-X 许可订阅，提供为期 5 年的业务关键型支持
798-SPEXEVAL	Spectrum-X 许可评估 (90 天)

准备好开始了吗？

要了解有关 NVIDIA Spectrum-X 网络平台的更多信息，
请访问：nvidia.com/en-us/networking/spectrumx

© 2023 NVIDIA Corporation.保留所有权利。NVIDIA、NVIDIA 徽标、BlueField、Cumulus、NetQ、Nsight 和 Spectrum 均为 NVIDIA Corporation 在美国和其他国家 / 地区的商标和 / 或注册商标。其他公司和产品名称可能是其各自关联公司的商标。其他所有商标均为其各自所有者的财产。2761336. 2023 年 5 月

