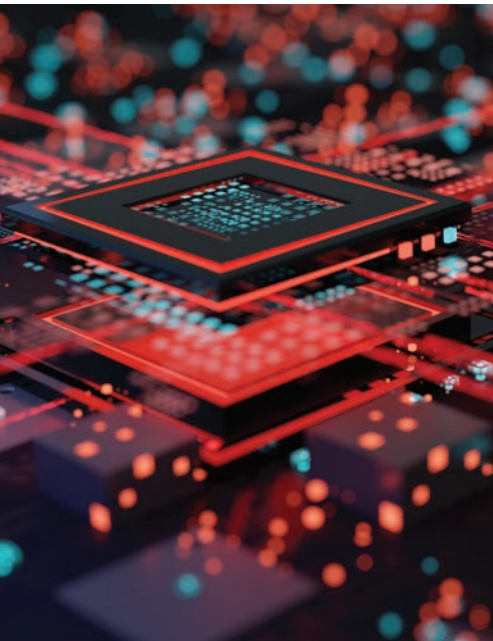


生成式人工智能解决方案指南

超越新兴人工智能网络的要求

生成式人工智能革命已经来临。 大家准备好了吗？



多年来，人工智能(AI)和机器学习(ML)重塑了各行各业，赋予人们强大的能力来解决了许多复杂的全球性问题。这些具有变革性的力量，过去被称为HPC(高性能计算)，已经推动了各种规模的企业组织进行数字化转型，不仅提高了生产力和效率，还增强了解决问题的能力。随着深度学习和神经网络支持的高度创新的生成式人工智能模型的涌现，游戏规则正在被进一步颠覆。通过生成原创内容和应对复杂的挑战，生成式人工智能不仅将彻底改变组织的运作方式，还将彻底改变创新本身的结构。

这些数据和计算密集型机器学习及生成式人工智能应用的增加，对数据中心基础设施提出了前所未有的要求，需要可靠的高带宽、低延迟数据传输、更高等级的布线和机柜功率密度，以及先进的冷却方法。随着数据中心为生成式人工智能做好准备，用户需要创新、强大的网络基础设施解决方案，以帮助他们轻松设计、部署和扩展后端、前端和存储网络架构，以适应复杂的高性能计算人工智能环境。值得庆幸的是，西蒙为您提供开展生成式人工智能革命所需的一切，而且所有这些都拥有经验丰富和专业的数据中心服务作为后盾。

人工智能在各行各业的蓬勃发展推动了35%以上的市场增长，预计至2032年，仅生成式人工智能一项的市场规模就将达到1.3万亿美元。

(Precedent Research, 2023)

先进的人工智能要求重新考量数据中心的设计

加速生成式人工智能和机器学习模型由训练型(学习新能力)和推理型(将能力应用于新数据)组成。这些深度学习和神经网络模仿人脑的架构和功能，在分析大量复杂数据集的模式、细微差别和特征的基础上，学习和生成新的原创内容。大型语言模型(LLM)，如ChatGPT和Google Bard，就是这些生成式人工智能模型的例子，这些模型在海量数据的基础上进行训练，以理解和生成合理的语言反应。

按顺序执行控制和输入/输出操作的通用CPU无法有效地从多种来源并行提取大量数据并进行快速处理。因此，加速机器学习和生成式人工智能模型依赖于图形处理单元(GPU)，后者使用加速并行处理来同时执行数千次高吞吐量计算。单个基于GPU的服务器的计算能力可媲美数十个基于CPU的传统服务器的性能！

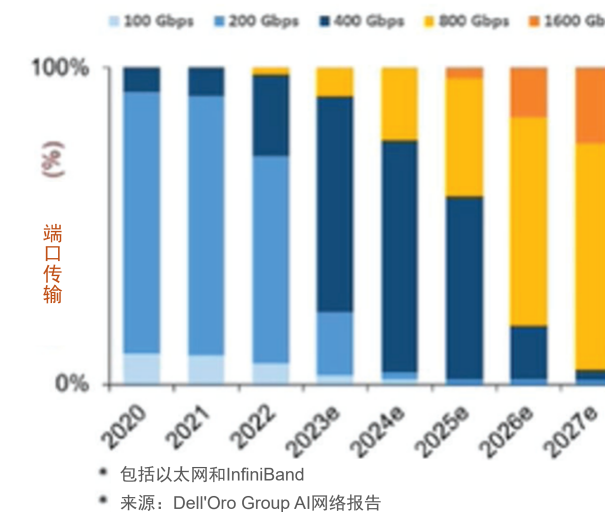
独特的机器学习和生成式人工智能特性

为了从海量数据集中提取信息，机器学习和生成式人工智能模型在后端HPC多节点集群中运行，这些集群拥有成百上千个相互连接的GPU，并有一些独特的要求：

- **非常高的带宽** – 服务器端100G、200G、400G，甚至是800G的速率，交换机至交换机连接快速迁移至800G和1.6T速率。
- **超低延迟** – 实时(< 20 毫秒)节点间东西向数据传输。
- **功耗大幅增加** – 基于GPU的服务器需要10倍以上的电力，导致机柜功率密度达到30-100kW甚至更高。
- **先进的冷却技术** – 数据中心正在评估更高效的冷却方法，如芯片级液冷和浸入式液冷，以处理服务器产生的高热量。
- **InfiniBand和以太网协议** – InfiniBand的高吞吐量和低延迟性能在后端的GPU间连接中占主导地位，而以太网的兼容性、安全性和管理功能则是前端接口的理想选择。以太网的不断进步将使二者在后端并存。
- **高密度、高性能的布线系统** – 节点之间、以及用于存储、管理和交换的高速连接需要更多的高带宽布线。

由于功耗和冷却要求较高，支持人工智能通常需要将GPU分配到不同的机柜中，并从柜顶交换 (ToR)配置转移到列末或列中(EoR/MoR)配置。较高的功率要求可能还需要先进、高效的液体冷却方法。

人工智能网络向高速迁移(后端)

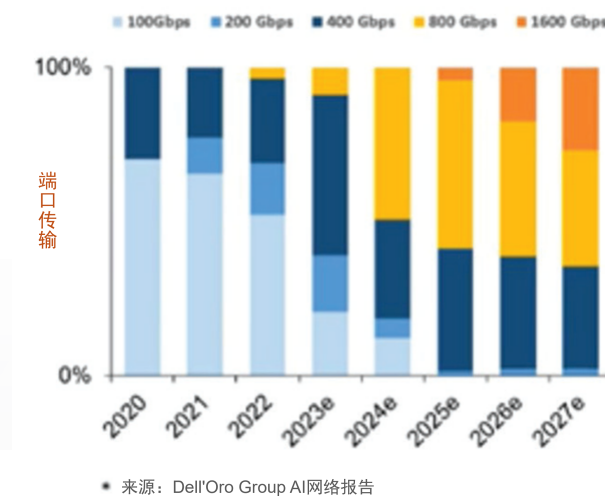


• 包括以太网和InfiniBand
• 来源：Dell'Oro Group AI网络报告

初步预测(2023-2027)

- InfiniBand和以太网将共存。
- 到2027年，几乎所有端口将达到800G及以上的速率。
- 网络带宽达到三位数的复合年增长率。

人工智能网络向高速迁移(前端)



• 来源：Dell'Oro Group AI网络报告

初步预测(2023-2027)

- 所有端口为以太网。
- 到2027年，几乎2/3的端口将达到800G及以上速率。

我们已为人工智能做好准备

无论是需要在由数千个互连GPU组成的大型HPC集群内支持更高的机柜功率密度和ToR配置的人工智能云服务提供商，还是希望在现有现场或主机托管数据中心内构建自己的特定业务人工智能基础设施的大型企业。西蒙提供用户所需的一切，以支持人工智能网络的设计、交付和日常运营，包括：



高密度、端到端的LightVerse单模和多模MTP光纤系统，可为后端、前端和存储架构提供高达800G及以上的高性能超低损耗(ULL)传输。



全面的直接铜缆(DAC)和有源光缆(AOC)系列，用于以太网、RoCE和InfiniBand网络人工智能集群内的后端点对点高速、低延迟连接。



能够在ToR和EoR/MoR配置中同时支持InfiniBand和以太网协议。



凭借专业知识和经验，帮助企业设计和部署优化的人工智能InfiniBand和以太网铜缆和光纤基础设施，以满足计算、存储、带外和带内管理等所有网络的需求。



数据中心服务，根据具体用例、预算、现有基础设施和未来需求来提供基础设施设计建议。



两大数据中心巨头强强联手

长期以来，以太网一直是满足大多数网络需求的实际协议，而InfiniBand则是HPC网络中的常见协议，由于其远程直接内存访问(RDMA)技术可以处理多个高带宽、低延迟的并行连接，因此已经成为人工智能集群中后端GPU互连的强大替代方案。InfiniBand交换机的延迟约为100纳秒，而以太网交换机的延迟为230纳秒。随着对人工智能支持的改善，全球InfiniBand市场预计将在2023年至2028年间增长40%以上。

相比之下，以太网的广泛TCP/IP生态系统支持、可扩展性、易用性以及安全和管理功能使其成为前端人工智能网络的理想选择，包括交换机到交换机连接、存储结构以及带外和带内管理网络。随着交换机技术和协议如RDMA over converged Ethernet (RoCE)和Ultra Ethernet Transport (UET)的发展以满足人工智能工作量的要求，InfiniBand和以太网将在后端人工智能集群中竞争和驻留，而前端人工智能网络仍将是以太网。归根结底，这两种技术在人工智能架构中各有所长，最高可支持800G的速度，其路线图可达1.6T。

通道	InfiniBand			Ethernet			连接选项
	接口类型	通道速率	链路速率	接口类型	通道速率	链路速率	
2	QSFP	EDR HDR NDR	50G 100G 200G	QSFP	25G 50G 100G	50G 100G 200G	1. 光收发器+光纤布线 2. 直连铜缆(DAC) 3. 有源光跳线(AOC)
4	QSFP OSFP	EDR HDR NDR	100G 200G 400G	QSFP OSFP	25G 50G 100G	100G 200G 400G	
8	OSFP	HDR NDR XDR	400G 800G 1.6T	QSFPDD OSFP	50G 100G 200G	400G 800G 1.6T	

西蒙数据中心设计服务

支持企业充分利用其数据中心人工智能环境的真正潜力。

无论您是受托为多个客户提供IT服务并履行SLA(服务水平协议)的服务提供商，还是投资人工智能和高性能计算(HPC)网络以加速业务发展的企业，数据中心的底层网络基础设施都是您发展壮大的不二法宝。

数据中心的布线基础设施是满足内部和外部客户对顶级人工智能网络和高性能计算可用性和性能期望的核心。您的团队还需要无忧的正常运行时间、可靠性和可扩展性保证，这样他们才可以专注于取得成功的必要条件。

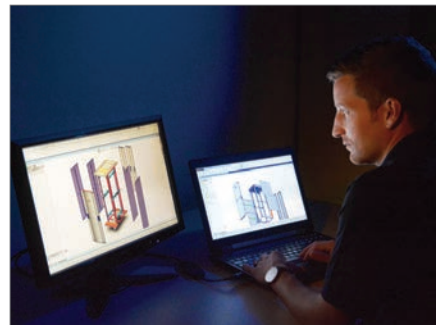
我们如何帮助您？

我们将数据中心的专业知识集中到全球服务网络中，旨在指导您选择和设计所需的底层物理基础设施，以确保您的数据中心为人工智能做好准备，同时为您提供所需的持续支持，以快速响应不断变化的需求、防止宕机并保持最高性能。



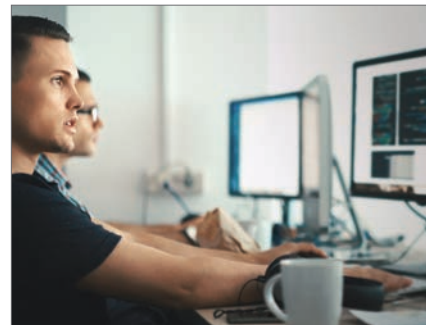
数据中心布线合规检查服务

获得对当前布线基础设施策略的总结，以及对可能进行的改进及节约的详细分析和指导建议。



数据中心设计服务

您是否希望设计、部署或升级数据中心空间，并需要专家的建议和帮助来完成整个过程？我们的数据中心设计专家团队随时准备为您提供支持。



技术服务团队

我们的专业技术服务团队遍布全球，为客户的整体西蒙体验提供支持。我们的团队与您一路同行。

业界领先的合作伙伴提供支持

多年来，我们建立了一个由数据中心合作伙伴组成的体系，他们都是各自领域的专家。西蒙以与全球人工智能和高性能计算领导者合作为荣，他们提供的互补性产品和服务与我们一流的IT基础设施解决方案相结合，为我们的客户提供额外的价值和支持。



对超低损耗连接的需求

后端和前端人工智能网络中的InfiniBand和以太网链路利用了经济高效的多模和短距单模光收发器技术的带宽能力。对于高速结构化布线链路，多模光纤可支持200G和400G至50米，而短距单模光纤可支持更长的500米和2000米距离。对于以太网部署，多模(SR和VR)和短距离单模应用(DR和FR)有严格的插入损耗要求，多模的最大信道损耗为1.9 dB，DR单模为3 dB，FR单模为4dB。

对于人工智能网络内的结构化布线，超低损耗(ULL) MPO/MTP连接可确保最大信道长度，同时确保持有余量以适应安装变量，并提供支持便捷跳接的灵活性，从而有助于提高可管理性、可扩展性和部署速度。在选择连接时，不同供应商的损耗值可能会有所不同，许多供应商提供标准损耗、低损耗和超低损耗连接。由于存在差异，因此必须确保第三方验证的最大插入损耗值能够真实反映性能。

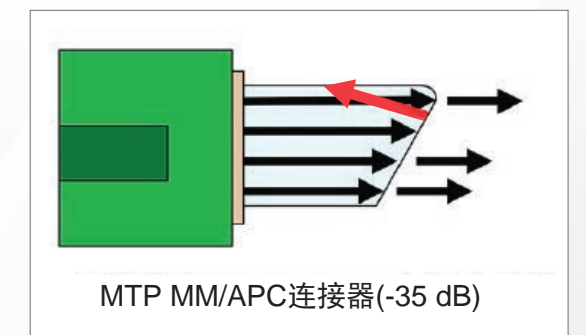
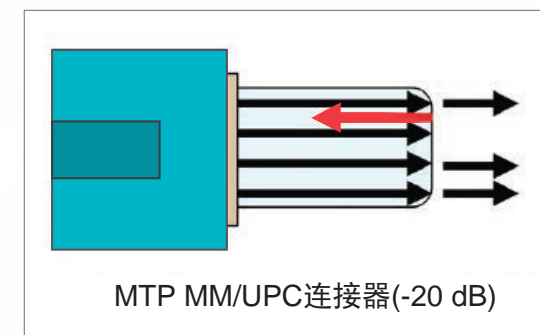
西蒙的ULL LightVerse® MTP多模和单模布线系统已通过第三方验证，可提供相当大的插入损耗余量，从而提高人工智能网络的性能，传输速度可达100G、200G、400G和800G。



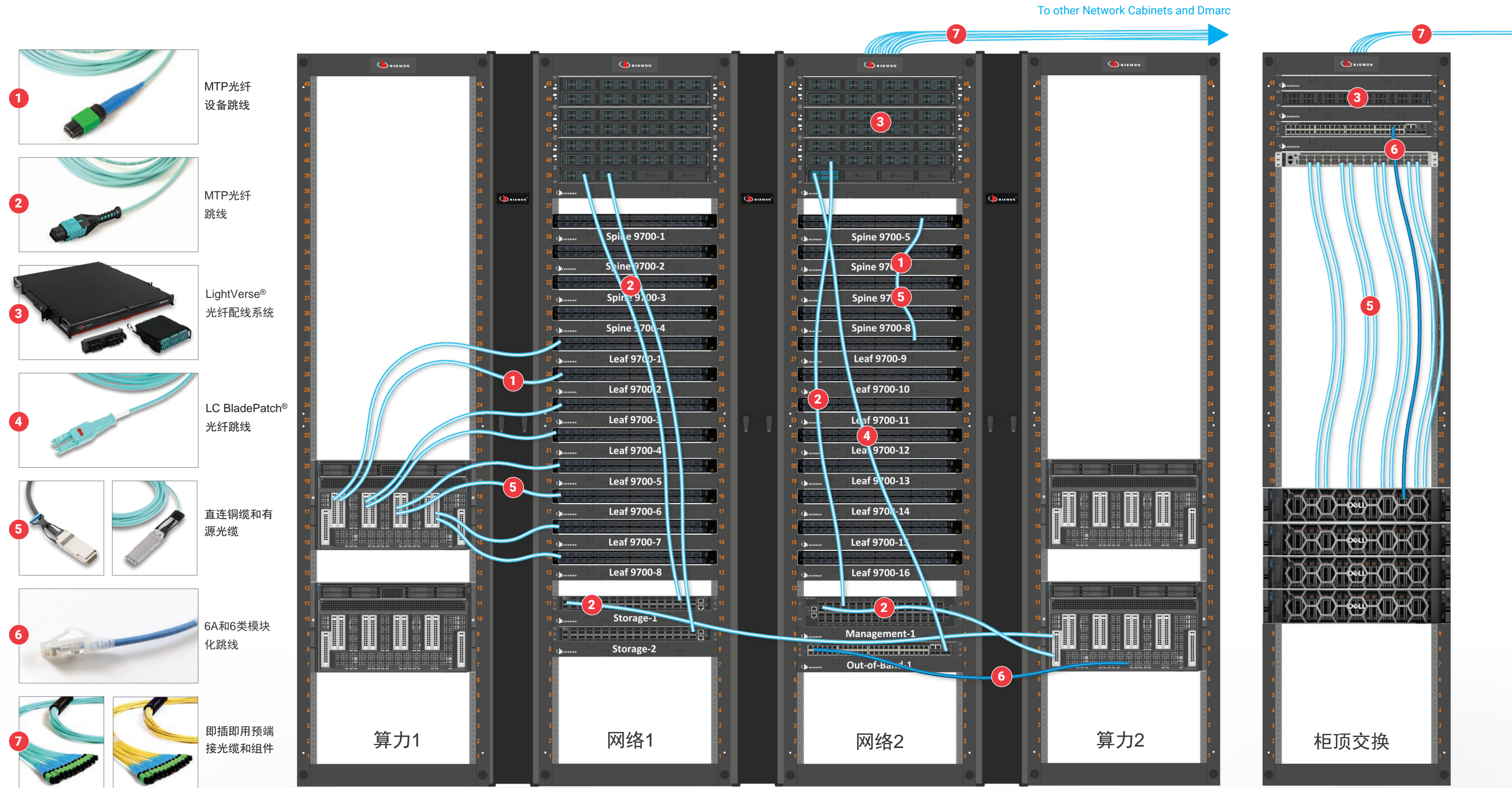
对多模和单模APC连接的需求

除了严格的插入损耗要求外，高速400G和800G应用以及未来的1.6T应用由于信噪比(SNR)较高，更容易受到反射率的影响。短距离DR和FR单模应用尤其容易受到影响，因此行业标准规定了基于信道中光纤耦合数量的反射率(回波损耗)值。

反射性能差会对信道插入损耗和传输性能产生不利影响。因此，除了用于高速AI链路的传统单模MTP连接外，许多云数据中心还指定使用多模斜面物理接触(APC)连接器。与采用圆形光纤端面的超物理接触(UPC)连接器不同，APC连接器采用8度角研磨，以减少反射信号量。多模MTP/UPC连接器的反射值通常为-20dB，而多模MTP/APC连接器的反射值提高到-35dB。



人工智能应用的布线解决方案





西蒙的优势

西蒙成立于1903年，在为全球客户制造和创新高质量、高性能数据中心解决方案方面是值得信赖的行业领导者。西蒙高级数据中心解决方案从底层开始设计，以满足当前和未来人工智能和高性能计算环境的苛刻需求。我们对数据中心需求的深刻理解推动了我们应对最新人工智能革命的战略转移。所有这些都以西蒙行业领先的质量、性能和可靠性为后盾，帮助您降低风险，最大限度地延长正常运行时间，并成功交付新的人工智能应用和服务。

- ▶ 西蒙已经成功地为其他几家公司设计并部署了HDR 200G和NDR 400G InfiniBand算力系统。
- ▶ 西蒙对行业领先的硬件厂商的参考架构设计有深刻理解，能为特殊用例场景提供布线设计建议。
- ▶ 西蒙针对AI低延迟需求有独特的解决方案，简化部署，并提供初始安装后的灵活性。
- ▶ 西蒙可以支持所有的通用网络：算力、存储、带外管理和带内管理。
- ▶ 不断优化提升的产品线以满足不断变化的数据中心需求，得益于持续改进、大量研发投入和对行业标准的领先参与的企业文化。
- ▶ 全球销售覆盖和完善的数据中心合作伙伴体系，以一系列领先的产品和服务为客户带来增值方案。
- ▶ 我们广泛的解决方案涵盖了InfiniBand和以太网光纤、铜缆、DAC和AOC，支持端到端安装。

生成式人工智能的混合布线系统: 超越传统的点到点

虽然高性能计算集群中的长距离链路得益于结构化布线，但GPU的短距离低延迟连接通常依赖于DAC等点对点解决方案。但是，铺展GPU和机柜的人工智能集群已经超出了DAC的长度极限。AOC和单根光缆可提供长达100米的机柜到机柜连接，但在大型集群中管理成百上千根光缆就成了一件麻烦事。使用结构化布线的一些好处如下：

增加灵活性：使用配线架和光跳线来连接任意两个GPU，更容易适应你不断变化的AI需求。

密集交换机连接：无缝管理高密度leaf至spine交换机连接的线路优化设计。

保护关键设备：在移动、增补和变更操作中避免触及重要的、昂贵的设备端口。

经济高效且可扩展：简化日后的运维工作，可扩展至更高速率而无需重新布线，节约时间和费用。

易于查障排错：配线架上的标准化标签和文档可简化故障排除工作，减少宕机时间。

改善气流和空间：减少线缆拥塞，改善气流，方便设备访问。

认证的性能：符合行业标准的系统保证布线性能。

与领先协会共同开拓人工智能

西蒙是很多塑造人工智能未来的领导组织的活跃成员。



“西蒙与InfiniBand的合作加强了我们在全球推进网络基础设施解决方案的承诺。我们认识到InfiniBand在满足人工智能和加速卡不断升级的需求方面发挥着关键作用。西蒙致力于提供创新的布线和连接解决方案，以促进该技术的发展和应用。”

– Gary Bernstein | 西蒙全球数据中心高级销售总监



ethernet alliance

“以太网联盟长期以来一直致力于通过行业标准和多厂商互操作性来支持以太网的发展。这与西蒙长期参与行业标准制定并致力于提供基于标准的优质解决方案的理念不谋而合。”

– John Siemon | 西蒙公司首席技术官



“IEEE802.3工作组负责制定以太网网络标准。西蒙积极参与所有重要工作组的工作，为众多人工智能网络制定200、400、800G和1.6T标准。”

– Dave Valentukonis | 西蒙公司北美技术服务经理

西蒙高级数据中心解决方案



高密度光纤配线箱

西蒙的高密度LightVerse®光纤布线系统包括高密度配线箱包括Core、Plus和Pro选项以及MTP到LC模块、MTP和LC适配器板，以支持单模和多模光纤配线，以及两种不同的光纤熔接选项——全部旨在支持400G及以上的网络部署。

go.siemon.com/LightVerseEnclosures



超高密度配线箱

西蒙的LightStack™ 光纤配线箱旨在轻松集成到任意标准的19英寸机架或机柜中，LightStack™ 光纤配线箱由高品质钢材制造而成，专为在当今先进的数据中心和存储区域网络环境中轻松安装和快速部署即插即用光纤解决方案而设计。

LightStack will be available in North America in Spring 2024.

go.siemon.com/LightStackEnclosures



直连铜缆

西蒙高速互连产品和有源光缆是音视频(AV)设备机房高速直连的理想选择，提供一系列QSFP28、SFP28、QSFP+、SFP+接口，以半米为增量的0.5米到10米的长度选项，及多种颜色可选。

go.siemon.com/DACs



有源光缆

西蒙多模有源光缆(AOC)组件最长可达100米，是数据中心、机柜和机架外设备中远距离点对点连接的理想选择。与收发器组件相比，其节能设计耗电量更小，线束更小有利于改善气流，并降低冷却成本。

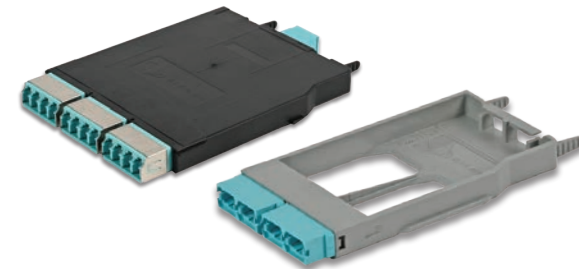
go.siemon.com/AOCs



模块和适配器

西蒙系列即插即用模块支持MTP至 LC/SC 光纤转换，可在单个模块中更快速、更轻松部署最多24芯光纤。LightVerse适配器板专为从配线箱前部单手轻松安装而设计，还可在配线箱内前后移动以完成装拆操作。

go.siemon.com/LightVerseModAdapt



超高密度模块和适配器

LightStack MTP到LC超低损耗(ULL)即插即用模块提供了一种快速高效的方式，可在紧凑的高密度包装中部署高性能光纤布线。LightStack超低损耗MTP直通适配器采用6端口设计，每个适配器板可支持最多72芯光纤，并且可提供正反两个方向的选择，以适应所有极性方式。

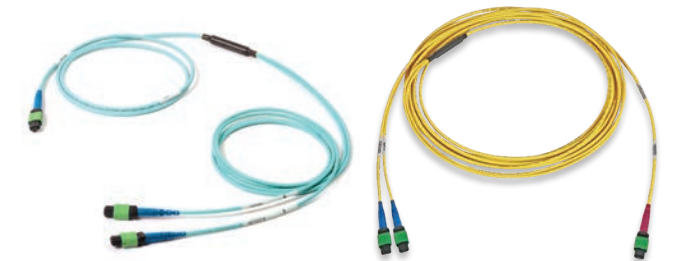
go.siemon.com/LightStackModAdapt



MTP预端接光缆

西蒙单模和多模即插即用预端接光缆和组件包括各种芯数的Base-8或Base-12 MTP至MTP预端接光缆、MTP至LC混合预端接光缆、MTP转换跳线、MTP跳线、LC或SC双工和单工扇出组件。近期还新发布了16芯MTP光跳线和从1根16芯MTP到2根8芯MTP的Y型转换跳线。

go.siemon.com/DCMTPTrunks



MTP光纤转换跳线

西蒙的转换跳线可确保在40和100G应用中达到100%的光纤利用率。有多种版本的转换跳线可供选择，包括将2根12芯MTP连接从主干预端接过渡到3根8芯MTP连接，以及将2根12芯MTP至MTP预端接缆从主干过渡到1根24芯MTP接头并连接到有源设备。

go.siemon.com/DCConversionCords

西蒙高级数据中心解决方案



LightVerse光/铜混合配线架

LightVerse®光/铜混合配线架旨在让用户在1U机架空间内轻松混合使用高性能光纤和铜缆连接。LightVerse光/铜混合配线架有平面和角型两种配置，支持单模和多模光纤，及屏蔽6A类和非屏蔽6A/6类铜缆连接。

go.siemon.com/CopperFiberPanels



模块化高密度光纤配线架

使用LightVerse模块化配线架有利于简化数据中心和智能建筑中的光纤部署。该类1U、2U或4U面板可为最多96根光纤提供灵活的高密度解决方案，支持无缝跳接、转换和熔接。

go.siemon.com/LightVerseHDPanels



先进的铜缆布线解决方案

使用西蒙先进的铜缆布线系统，打造卓越网络。我们的UltraMAX™和Z-MAX™系统为5e类、6类、6A类、7A类和8.2类选项提供新一代的端到端性能。探索双绞线线缆、RJ45/TERA®插座，现场安装插头、预端接光缆、配线架和各种跳线的潜力。放心地进行设计、部署和连接。

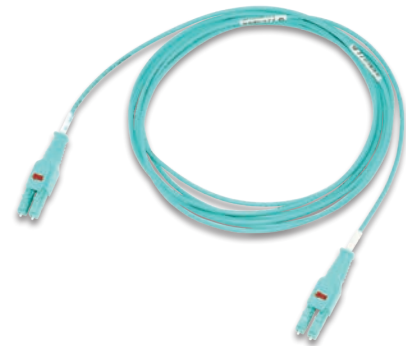
go.siemon.com/Copper



铜跳线

SkinnyPatch®模块化跳线可提供卓越的性能，且线缆直径更小，可显著改善气流，并提升高密度配线区的灵活性。其较小的28号线规多股铜缆结构可提供显著更小的弯曲半径，有利于轻松进行线路管理并强化线缆管理。SkinnyPatch有6A类屏蔽、6A类UTP和6类UTP版本可选。

go.siemon.com/DCSkinnyPatch



高密度光跳线

西蒙LC BladePatch®单模和多模双工跳线可为高密度光纤配线环境提供独特的解决方案，其革命性的推拉式UniClick™护套设计可控制搭扣，便于在紧凑空间内轻松进行跳线插拔。

go.siemon.com/DCFiberJumpers



MTP光跳线

西蒙MTP跳线用于将MTP预端接主干连接到有源设备。设计紧凑的MTP跳线和西蒙2mm直径的RazorCore光缆有利于实现更大的连接访问，减少线缆通道堵塞，并改善有源设备附近的气流。MTP跳线提供Base-8和Base-12两种版本可选。

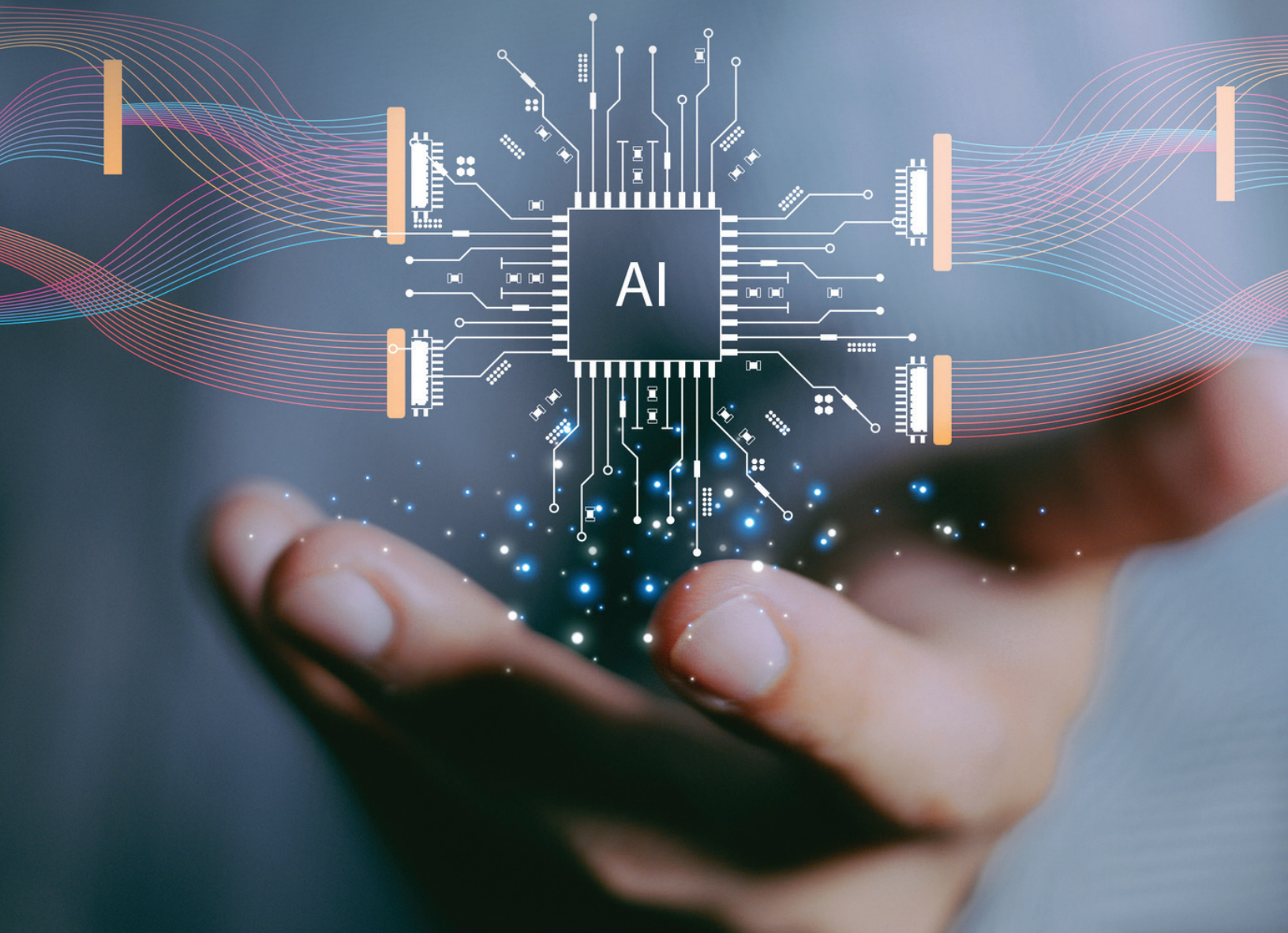
go.siemon.com/DCJumpers



机架及线缆管理

西蒙提供各式机架、垂直和水平线缆管理器及附件，以满足各种数据中心需求。

go.siemon.com/DCRacks



今天就开始您的西蒙人工智能之旅吧