

## 数据中心高速网络升级平台：基础设施相关问题、趋势、推动力和建议

---

# 目录

<b>引言</b>	<b>3</b>
<b>不断变化的网络架构</b>	<b>4</b>
<b>不断演化的标准</b>	<b>5</b>
<b>升级选项</b>	<b>6</b>
40G 还是 25G 通道?	6
调制方案	6
收发器技术	6
串行还是并行传输	6
预端接线缆还是现场端接线缆	7
单模还是多模	7
OM4 还是 OM5 (宽带)	8
智能系统	9
<b>康普的观点</b>	<b>9</b>
<b>结语</b>	<b>10</b>

对于数据中心来说，速度就是一切。目前面临的挑战在于，您需要展望未来，了解近期及往后需要准备交付的成果，并描绘出最有利、最灵活的发展道路。可用技术和应用标准的发展速度越快，这项任务的难度就越大。

根据近期数据中心的发展趋势，预计数据中心的流量和带宽要求将以每年 25%-35% 的速度增长。为满足不断上升的网络容量需求，只有采用更高的交换速度，而这正是当今市场的发展趋势。据 Dell'Oro Group 调查，2017 年第 1 季度，25 Gbps 和 100 Gbps 端口的发货量增至 100 万以上。Dell'Oro 预测，到 2020 年底，以太网交换机收入将持续增长，其中 25G 和 100G 端口占很大份额。<sup>1</sup>

升级策略也在不断演进。100G 交换机链路（多模和单模）成本不断降低，使许多公司能够将交换网络从 10G 直接更新至 100G，完全跳过 40G。随着 25G 通道交换机变得越来越普遍，向 25G 通道的过渡也在顺利进行。与此同时，专有的和基于标准的 PAM-4 调制技术迎来了 50G 通道速率时代。25G 和 50G 端口的不断普及继续影响着 40G 服务器的部署。

展望未来，预计到 2020 年，通道容量将继续翻倍增长到 100G，以支持下一代高速矩阵交换机连接。吞吐速度激增由多个因素造成。

- 服务器密度以每年约 20% 的速度增加。
- 处理器能力也在不断提升，英特尔近期就推出了 22 核处理器。
- 虚拟化密度正以 30%<sup>2</sup> 的速度增长，这提升了上行链路至交换机的速度。
- 数据中心的东-西流量远远超过南-北流量。<sup>3</sup>

“未来的想法是：顺应以太网交换机的发展趋势，将通道速度提升到 25 Gb/s，然后是 50 Gb/s，再到 100 Gb/s，同时将光纤数量控制在 8 个以内。”

— 《The Next Platform》，2016 年 3 月

网络设计必须体现这一巨大流量，而且重要的是，必须能让服务器、存储设备和网络容量独立扩展，且中断和重新配置尽可能的少。因此，专业的数据中心现在必须支持更大的服务器密度，部署更多光缆，并加快计划使它们的核心和汇聚网络向更高的速率升级。数据中心内的网络基础设施必须能够扩展支持这些重大变化。

“采用骨干 (Spine) 和分支 (Leaf) 这一类的网络矩阵不但会刺激带宽需求，还会影响网络规模，要求为布线基础设施配备更多光缆。”

— 《Data Center Journal》，2016 年 4 月 25 日

## 不断变化的网络架构

数据中心流量和方向的变化要求网络设计能够适应东-西数据流量的快速增长。传统数据中心架构使用 3 层拓扑 (图 1)。核心层一般位于主配线区 (MDA), 各种网络交换机就是经由此处彼此连接, 以及与数据中心外的网络资源连接。

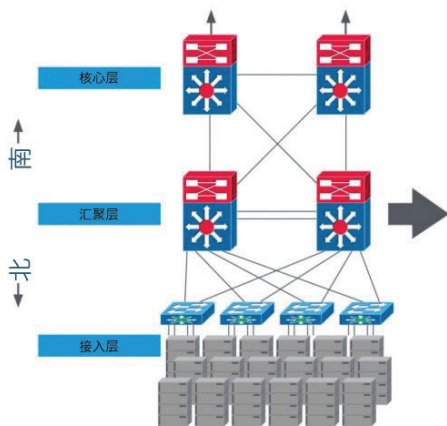


图 1: 传统的 3 层拓扑

核心层也为连接各种接入交换机的汇聚层提供资源。在大型企业和云数据中心内, 汇聚层一般位于中间配线区 (IDA)。在规模较小的数据中心中, 一般位于水平配线区 (HDA) 或设备配线区 (EDA)。接入网络连接数据中心的存储和计算资源。这种模型设计为可扩展的数据中心网络提供了可预测的基础, 但如果要用来支持如今低延迟的虚拟化应用, 就不尽如人意了。因此, 网络矩阵就出现了向“分支 (Leaf) 和骨干 (Spine)”矩阵的快速重大转变 (图 2)。经过优化的分支 (Leaf) 和骨干 (Spine) 模型以东西流向移动数据, 允许服务器互相配合, 交付基于云的应用。在这种拓扑中, 网络跨多个分支 (Leaf) - 骨干 (Spine) 交换机分布, 使得分支 (Leaf) - 骨干 (Spine) 交换层成为最大规模交付和性能的关键。

每个分支 (Leaf) 交换机连接各个骨干 (Spine) 交换机, 构建出一个具备高度弹性的“任意对任意” (any-to-any) 结构。光纤链路网状连接可以形成高容量网络资源或“矩阵”, 并与连接的所有设备共享。所有矩阵连接都以相同速度运行。速度越高, 网状网络 (通常称为“矩阵网络”) 的容量就越高。

矩阵网络需要大量光缆连接, 尤其是在分支 (Leaf) 交换层。设备供应商不断致力于提升其板卡密度, 以保证跟上趋势。随着密度的不断提高, 布线连接和管理就变得更加重要。矩阵网络要求整个网络实现高速连接, 而该网络通常遍布整个数据中心。部署速度更高且连接范围更长的更多链路已成为物理网络设计的新常态。

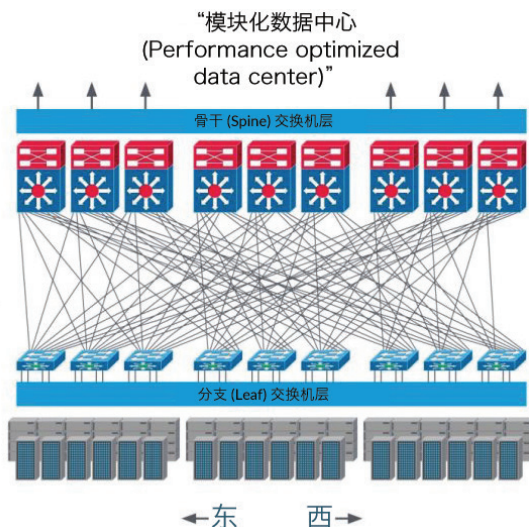


图 2: 两层骨干 (Spine) - 分支 (Leaf) 架构

## 不断演化的标准

应用标准组织, 即 IEEE 802.3 (以太网) 和 ANSI/T11 (光纤通道委员会) 一直忙于更新建议的准则, 以便跟上带宽快速增长的节奏。这些标准组织的目标不仅仅是促进向不断提高的线路速率发展; 还鼓励开发速度更高的应用, 以便提高数据中心设备之间链路的成本效益。为此, 多种中间速度有待开发, 用以弥补 10G、40G、100G 和 400G 之间的空白。表 1 列出了各种以太网标准。仍在制定的标准高亮显示为紫色。

表 1 - IEEE 802.3 以太网光纤标准-已完成和正在制定中（紫色）

应用	标准	IEEE 编号	介质	速度	目标距离
10G 以太网	10GBASE-SR	802.3ae	MMF	10 Gb/s	33 米 (OM1) 至 550 米 (OM4)
	10GBASE-LR		SMF		10 千米
	10GBASE-LX4		MMF		300 米
	10GBASE-ER		SMF		40 千米
	10GBASE-LRM	802.3aq	MMF		220 米 (OM1/OM2) 至 300 米 (OM3)
25G 以太网	25GBASE-SR	P802.3by	MMF	25 Gb/s	70 米 (OM3) 至 100 米 (OM4)
40G 以太网	40GBASE-SR4	802.3bm	MMF	40 Gb/s	100 米 (OM3) 至 150 米 (OM4)
	40GBASE-LR4		SMF		10 千米
	40GBASE-FR		SMF		2 千米
	40GBASE-ER4		SMF		40 千米
100G 以太网	100GBASE-SR10	802.3bm	MMF	100 Gb/s	100 米 (OM3) 至 150 米 (OM4)
	100GBASE-LR4		SMF		10 千米
	100GBASE-SR4		SMF		70 米 (OM3) 至 100 米 (OM4)
	100GBASE-ER4		SMF		40 千米
50Gm 100G 和 200G	50GBASE-SR	802.3cd	MMF	50 Gb/s	100 米 (OM4)
	以太网		SMF		2 千米
	50GBASE-LR		SMF		10 千米
	100GBASE-SR2		MMF	100 Gb/s	100 米 (OM4)
	100GBASE-DR2		SMF		500 米
	100GBASE-FR2		SMF		2 千米
	200GBASE-SR4		MMF		100 米 (OM4)
200G 以太网	200GBASE-DR4	P802.3bs	SMF	200 Gb/s	500 米
	200GBASE-FR4		SMF		2 千米
	200GBASE-LR4		SMF		10 千米
400G 以太网	400GBASE-SR16	P802.3bs	MMF	400 Gb/s	70 米 (OM3) 至 100 米 (OM4)
	400GBASE-DR4		SMF		500 米
	400GBASE-FR8		SMF		2 千米
	400GBASE-LR8		SMF		10 千米

## 升级选项

围绕向更高线路速率升级的讨论不仅复杂，而且发展快速。它包含各种各样的决策，与光纤类型、调制和传输方案、接头类型，当然还有成本考量有关。图 4 显示了一种可能的升级路径，但还存在许多种其他路径。为给定环境选择最佳的路径需要仔细考虑各个方面。下方列出了必须要衡量的问题中的几种。



图 4: 40GBASE-SR4 链路，交换机和服务器采用并行光纤链路

## 40G 还是 25G 通道？

直到最近，广为接受的升级线路图才描绘了从 10G 通道到 40G 通道的预期升级。自从 IEEE 802.3by 标准获批以来，该行业已开始向作为下一代交换技术的 25G 通道过渡。这主要是由于新的 25G 通道更容易向 50G (2x25G) 和 100G (4x25G) 通道升级，并在一定程度上提高网络交换机中交换芯片的利用率。在相同的资本和运营成本下，与 10G 网络端口相比，25G 网络端口提供的容量更高。25 通道还可清晰和逻辑地支持 100G、200G 和 400G 速率。

## 调制方案

目前，更高效的新型调制方案也已经推出。方案已经建议在数据中心内部以及多个数据中心设施之间为光纤链路使用具备 4 个幅度等级 (PAM-4) 的脉冲幅度调制。如图 5 所示，PAM-4 采用四种不同的脉冲幅度来传输数据。与传统 NRZ 相比，在信号传输速率相同的情况下，PAM-4 支持两倍传输容量。然而，不足之处是需要更高的信噪比 (SNR)，这就对支持的物理基础设施提出了更严格的要求。尽管如此，PAM-4 的简便性和低功耗特性仍使其成为最具前景的 100G 及更高容量的调制技术之一。

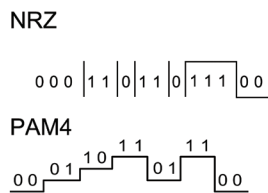


图 5: 6-4 和 NRZ 调制

## 收发器技术

除了通过更先进的调制方案来提升通道速度以外，各种波分复用 (WDM) 技术也在开发之中，用于增加单根光纤的波长数量。20 多年以来，人们一直使用 WDM，通过减少光纤数量来提升远距离网络的数据速率。它也被用在单模以太网应用中，例如 10GBASE-LR4 和 100GBASE-LR4，即利用粗波分复用 (CWDM) 技术，将 4 种波长融入到同一光纤之中。这种概念也借用到名为短波 WDM 或 SWDM (短波波分复用) 的技术，被延伸到多模光纤。如图 6 所示，SWDM (短波波分复用) 采用 850 nm 至 940 nm 的波长。

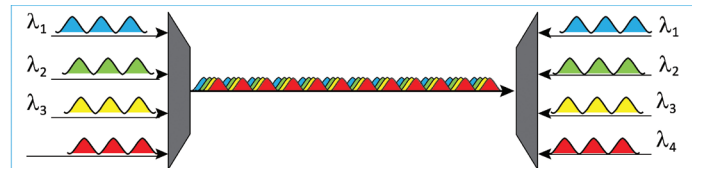


图 6: 融合 850 nm 至 940 nm 的 4 个波长的 SWDM (短波波分复用)

## 串行还是并行传输？

由于数据速率已随应用需求增长而提高，市场亦已转而采用并行光纤。这种趋势背后的支撑是对基于 MPO 主干的一贯需求，十多年以来，它一直是数据中心的主流产品。通过使用激光优化多模光纤 (LOMMF)，串行光纤能够经济有效地支持高达 10G 的速度。但在过去，使用串行传输支持 25G 或 40G，需要采用更昂贵的单模收发器。然而，并行光纤可为升级到 40G 提供经济高效的解决方案，并允许多个 25G 通道提供 100G 的连接。与此同时，通过结合串行和并行传输，在单模和多模光纤上建立适合 200/400G 以太网的未来升级路径。

越来越多地使用 MPO 接头也进一步加速了向并行光纤的转变。根据预测，在北美地区，2020 年之前，40/100GbE MPO 接头的销量每年将上涨 15.9%，到 2020 年将达到 1.26 亿美元。<sup>4</sup> 但是，随着新技术的实施可以更有效利用双工光纤，并行光纤的趋势可能出现起伏。

与此同时，SWDM4 等经济高效的技术也开始推动使用四个 25G 通道的双工 100G 应用发展。在不久的将来，50G PAM-4 通道还将在多模光纤上提供 100G 的连接。SWDM4 和 PAM-4 均可节省更多支出，因为它们使用的纤芯数更少。

MM	标准/ (光纤数量)	最长距离
	100GBASE-SR4 (8)	OM3 70 米 OM4/OM5 100 米
100GBASE-SR10 (20)	OM3 100 米 OM4/OM5 150 米	
100GBASE-eSR4 (8)	OM3 200 米 OM4/OM5 300 米	
100G-SWDM4 (2)	OM3 75 米* OM4 100 米* OM5 150 米	
100G-eSWDM4 (2)	OM3 200 米* OM4 300 米* OM5 400 米	

图 6: 数据中心的 100G 短距离应用  
\* 仅在 850nm 时指定 OM3/OM4 有效模式带宽

### 预端接线缆还是现场端接线缆

快速部署网络服务的必要性增加了预端接布线系统的价值和需求。据某些估算数据，与现场端接系统相比，预端接线缆的即插即用能力可以帮助节省 90% 的施工时间，而在网络维护方面，预端接系统比现场端接解决方案的速度快 50%。<sup>5</sup> 随着网络中采用的光纤连接数量增加，其价值也相应增长。工厂端接系统也是针对支持高速光纤链路所需的超低损耗系统唯一可行的解决方案。无论是单模还是多模连接系统，预端接解决方案中的 MPO 光纤方案都已经成为主流的方案，原因在于其高性能、易用性、部署速度和布线密度。

### 单模还是多模

数据中心管理者面临的最复杂决策之一就是，何时以及在何处部署单模或多模链路。可插拔单模、光模块的成本持续降低，从而使 100G 以太网能够在数据中心交换机端口市场中占较大份额。在超大规模和企业数据中心也是如此。

但是，关于这三种传输类型的讨论必须不再局限于可插拔光模块的成本，还应包括信道总成本的分析，以及数据中心的预期增长及其升级路线图。以下是在做任何决策之前应考虑并彻底了解的一些问题。

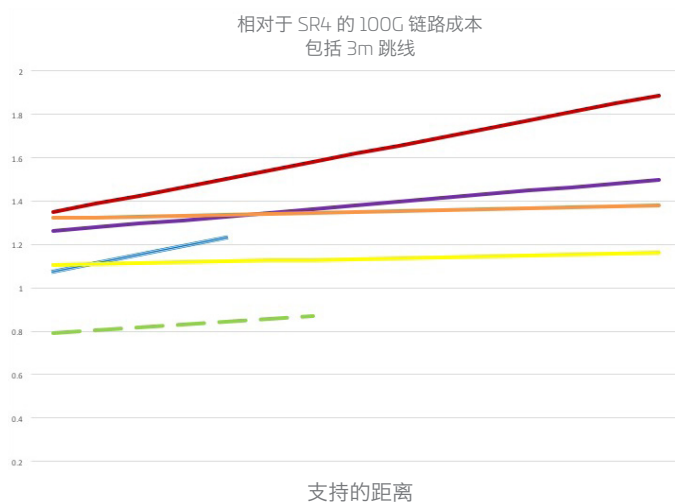


图 7: 单链路成本比较

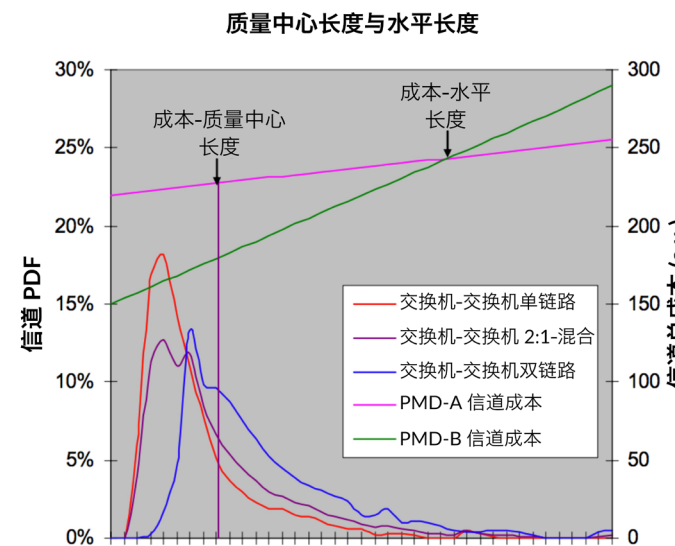


图 8: 根据拓扑结构估计适当的信道长度  
资料来源: IEEE 802.3 新一代 40G 和 100G 光学研究小组 2012 年 5 月

**链路距离:** 通常，数据中心要求使用大量的网络链路和相对短的距离。这样低成本多模光纤就更具吸引力了，但前提是它必须支持未来网络发展所需的速度。另一方面，单模光纤常见于数据中心接入间，且远距离能力使其成为数据中心和城域网/广域网之间链路的唯一选择。事实上，许多远距离、高速选项只能使用单模光纤。

**网络拓扑结构:** 一些数据中心可能拥有 100,000 多台服务器，而其他数据中心可能只有少数几台服务器。一些数据中心将网络设备集中放置在一起，而其他数据中心则将网络设备分布于整个数据中心。这些设计要求和选择决定了网络链路的数量以及网络链路必须支持的距离。

**信道总成本：**不同网络类型之间的链路成本比较需评估整个链路（收发器、主干光缆和跳线）的成本。为有助于比较不同网络链路类型的相对成本，已经开发出来各种成本模型。其中一些模型（如图 8 所示）根据所选拓扑结构对适当的链路长度提供了相关指导，且在平均链路长度未知的情况下比较有用。例如，该模型指出，长度约为 230 米时，PMD 信道 A 和 B 的相对成本相等。因此，了解链路长度使我们能够确定更低成本的解决方案。

当平均信道长度已知时，准确地比较不同链路类型之间的成本就更容易些。通过利用数据资源（如图 7 中的图表），评估相对信道总成本的流程就相当简单。图 7 显示了长度从 50 至 300 米的各种 100GBASE 链路的成本（收发器、主干光缆和跳线）比较。该模型还比较了使用 OM5 布线的 100GBASE-SWDM4 双工光纤与使用 OM4 的 100GBASE-SR4 的成本。另外，它还表明 SWDM 选项可实现更低的资本成本。因为 SWDM 采用 OM5，而 OM5 支持传输 100G 更远的距离。较之于短距离基于单模光纤的光模块（如 PSM4），近期推出的在 OM5 上支持 400 米传输的 100G eSWDM4 有过之而无不及。

虽然任何链路的成本都取决于长度，但是一些链路的成本本来就更高，其原因在于使用的光纤数量更多，而这一差异必须在比较时加以考虑。同样重要的是，有些成本模型（如图 7 和 8 所示）适用于企业数据中心，但不适用于比较超大规模数据中心的链路成本。原因在于这些设施的规模和带宽要求都极高。

**其他考虑因素：**在许多情况下，信道距离可能比较短，以至于长度并不是确定成本的关键变量。在这种情况下，最佳传输媒介的决定通常取决于以下一个或多个因素。

**链路速度：**每个数据中心将（或应）拥有其自己的升级路线图，该路线图以组织预测的 IT 需求以及支持该 IT 需求所需的基础设施发展为基础。传输介质必须能够支持当前和未来应用的最高链路速度。

**信道运营支出：**运营成本应包括评估支持考虑使用的传输介质所需的人员、流程和供应商关系。每种技术的广泛功能和复杂性导致了技能的专业化、标准的流畅性以及其它核心能力。在没有必要资源进行管理的情况下，采用新的传输介质会增加风险并导致额外成本。

**基础设施生命周期：**理想情况下，基础设施应能够支持多代设备技术，以避免进行成本高昂的升级换代。

### OM4 还是 OM5（宽带）

在多模环境下，数据中心用户需做出其他复杂决策，涉及应部署何种多模技术。他们的选择包括 OM3、OM4 和 OM5 宽带技术。

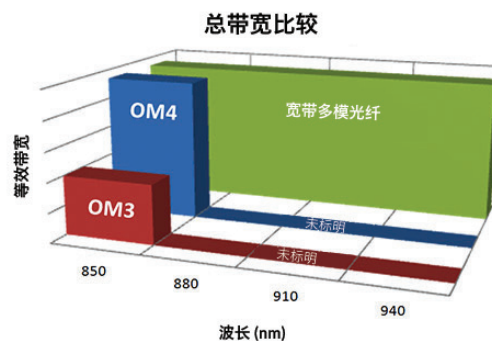


图 9：总带宽比较

OM3 激光优化光纤最初是为了支持 10GbE 链路而开发的。2002 年，OM3 实现了标准化，且其后继技术 (OM4) 于 2009 年实现了标准化。这两种多模光纤都针对 VCSEL 收发器进行过优化处理，可实现 850 nm 波长上的信号传输，且两者都使用相同的连接器。然而，较之于 OM3，OM4 在衰减和带宽方面进行了改进。



OM3 的最大衰减低于 3.5 dB/Km，但使用 OM4 传播时，损耗为低于 3.0 dB/Km。因此，OM4 能够支持更远的距离和更高的吞吐量。

所以，真正的决定在于选择 OM4 还是 OM5。康普于 2015 年推出的 OM5 近期通过了 ANSI/TIA-492AAA E 批准，也已经得到了 ANSI/TIA-942-B 的推荐。OM5 明显改进了对短波分复用 (SWDM) 技术的支持，进一步延长了传输距离。通过使用 OM5，数据中心运营商还可以将并行光纤数量减少到原来的 1/4 或更少。这意味着，欲支持 40 Gbps 和 100 Gbps 通道，只需 2 芯 OM5 光纤，但如果使用 OM4 光纤，则需 8 芯。图 9 显示了 OM3 多模光纤、OM4 多模光纤以及 OM4 宽带多模光纤之间的带宽比较结果。此外，OM5 还支持所有基于传统多模光纤的应用，并与 OM3 和 OM4 光纤兼容。随着 WDM 和 PAM-4 技术继续发展，OM5 支持 SWDM 的能力也将使其从传统多模光纤中分离出来。

## 智能系统

自动化基础设施管理 (AIM) 系统可以通过为物理层和所有连接设备提供精准映射，大大促进升级进程。由于 AIM 系统会自动监测和记录所有在用的端口和光纤，从双工向并行升级时，它们可以帮助确保有足够光纤可用。

此外，AIM 可以帮助确认多余的线缆和交换机端口，并将它们用于并行-双工升级。针对 AIM 的 ISO/IEC 18598 标准以及欧洲标准 EN 50667 等已于 2016 年获得批准。部署 AIM 基系统的好处亦得到了 TIA 的肯定，因为该组织制定了 ANSI/TIA-5048 标准草案，该草案重复了（几乎逐字重复）ISO/IEC 18598 标准中的内容。



图 10: 具备各种光纤数量的 MPO 连接器

## 康普的观点

在评估这些选项及市场发展轨迹的过程中，以下建议代表了康普对本文讨论的一些问题的看法：

- 基于 MPO 的预端接光纤解决方案仍然将是高性能网络的首选。它们提供出色的工厂级端接性能，此外，其速度和灵活性也可以支持私有、云化的企业数据中心的扩展需求。
- SYSTIMAX® 超低损耗 (ULL) 单模和多模主干光缆和布线组件将大大增强对高速应用的支持，同时保持支持 TIA 942-B 结构化布线设计的灵活性。
- 已部署了数年的 MPO 12 芯系统将继续用于支持双工及并行应用。ULL 性能的提升将为部署和访问大多数数据中心应用提供卓越的灵活性，同时实现可靠统一运营。随着未来新应用的涌现，预计 MPO 12 芯系统将继续使用。
- 针对大容量、高密度应用，我们主张使用 MPO 24 芯多模系统。随着骨干 (Spine)-分支 (Leaf) 架构越来越成熟，MPO 24 芯系统亦可满足双工多模网络发展的高密度和高容量需求。MPO24 的另一个优势是可为 8 芯并行应用提供灵活的支持。
- 最后，我们预测选择性使用 MPO 8 芯系统。这包括支持采用 4X10G 或 4X25G 配置的常见 4 通道 QSFP 应用，主要用于存储设备和服务器之间网络连接。因为网络矩阵主干链路不需要利用扇出来的低速端口，2 芯双工链路（例如 100G SWDM4）可以成为非常具有吸引力的交换机-交换机链路选择。

无论您作何选择，康普解决方案支持 8 芯、12 芯和 24 芯并行以及 2 芯双工应用，为广泛的数据中心应用提供最佳支持。

## 结语

尽管了解大量技术选项和新兴解决方案非常重要，但必须考虑企业数据中心具体环境。什么是企业的发展轨迹？它如何影响数据中心的变速度和扩展要求？正在考虑的各种升级场景的总拥有成本是多少？

请记住，作为数据中心管理者，您不必单独行动。所涉及的研究工作量和决策数量可能会让人觉得心烦意乱。目前有各种知识渊博的资源中心，具备多种解决方案和丰富经验，可帮您做出正确决定，康普便是其中之一。凭借我们的技术专业知识和广阔视野，我们可以帮助您制定长期的升级策略，以保持数据中心的适应性、功能性和高效性。而不受事物变化影响。康普术业专攻，实现未来愿景。

### 参考资料

- 1 以太网路线图中的构建区域；《The Next Platform》；2016年3月24日
- 2 数据中心战略北美企业调查；Infonetics Research；2015年5月
- 3 Facebook 提供网络数据中心设计方面的经验教训；2014年11月
- 4 市场预测——40/100GbE MPO 接头；ElectroniCast Consultants；2015年12月
- 5 衡量预端接光纤系统的成本和效益；线缆安装与维护；2014年5月1日

COMMSCOPE®

commscope.com.cn

欲了解更多信息，请访问我们的网站或联系您的康普销售代表。

© 2018 CommScope, Inc. 版权所有。

本文件仅供规划设计之用，不涉及对任何康普产品或服务相关规格要求或保证的修改或补充。

所有标有®或™的商标均为康普公司相应的注册商标或商标。康普致力于最高标准的商业诚信和环境可持续发展，其全球诸多分支机构已获得ISO 9001、TL 9000、ISO 14001等国际认证。更多相关康普公司的承诺，请访问 <http://zh.commscope.com/About-Us/Corporate-Responsibility-and-Sustainability>。

WP-110615-22-ZH.CN (10/17)