CORNING

IT 服务数据中心的设计浅析

康宁光通信企业网亚太区市场经理 马锐

随着云计算技术安全性稳定性的大幅提高，越来越多的企业开始使用第三方云服务来处理各种数据业务需求。因此，云服务对网络基础设施的投入在未来几年将会迅速增加，成为数据中心市场发展的主要驱动力。相比较而言，传统企业 IT 基础设施费用则呈现出逐年下降的趋势（如图 1 所示）。

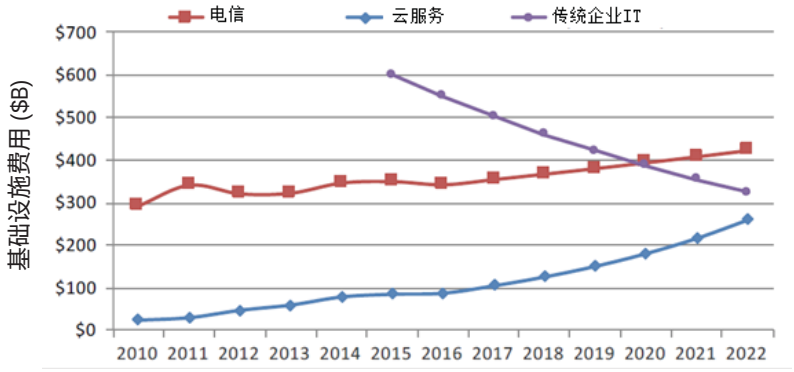


图 1. 全球网络基础设施费用细分市场对比趋势

来源：LightCounting and Forbes

用于第三方 IT 服务的云数据中心与传统企业数据中心对比，具有很多新的特点。云数据中心需要更高的带宽，更低的时延，更大的网络规模和更快的升级频率等。2016 年大约全球 80% 的 40GbE 光传输设备销售额来自互联网云服务企业。从 2017 年开始由于云数据中心的驱动 100GbE 光传输设备需求迅速增长。预计到 2022 年云服务对以太网收发器的需求将达到全球市场的 70%（如图 2 所示），并且在 2018 年后开始带动 200GbE 和 400GbE 的销售。相对而言，传统企业对以太网收发器的需求将会保持平稳，并将长期以 10GbE 网络为主。本文中我们将根据 IT 服务类用户的需求特点对数据中心光纤类型的选择，传输方式的选择和布线结构的选择进行浅析。

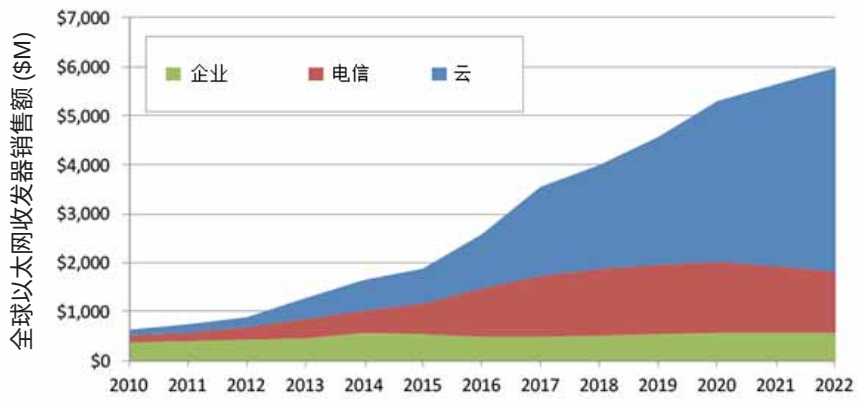


图 2. 全球以太网收发器销售额分布 ($M)

来源：LightCounting

光纤类型的选择

在数据中心以太网设计中，光纤类型的选择需要根据数据中心对传输速率，传输距离等的需求进行综合考虑。目前云计算数据中心已经普遍部署 40G 乃至 100G 以太网来支持日益增长的数据业务需求。多模光纤系统因技术成熟度高，总体成本较低等优点，仍旧被大多数数据中心所采用。然而业界普遍认为传统的 OM3/4 多模光纤主要用来支持 100m-150m 的短距离传输。因此为了延长多模光纤在高速网络中的传输距离，相关国际标准组织正式命名了基于波分复用技术的 OM5 多模光纤。

	40G				100G			
光纤类型	标准 SR4	COC eSR4	COC BiDi	Finisar SWDM	标准 SR4	COC eSR4	Prelim BiDi	Finisar SWDM
OM3	100	330	100	240	70	200	70	75
OM4	150	550	200	350	100	300	100	100
OM5	150	550	200	440	100	300	150	150

表 1. 使用不同光纤类型和收发器支持的 40G/100G 以太网最大传输距离对比

*COC= 使用 Corning 产品测试

CORNING

LANscape®
Pretium® Solutions

面对三种多模光纤，数据中心设计人员应该根据数据中心的实际需求进行选择。在表 1 中我们列出了不同光纤类型使用不同收发器时能够支持 40G/100G 以太网的最大传输距离。目前因只有 40G SR4 和 100G SR4 是 IEEE 正式发布的国际标准，因此使用其他收发器类型的传输距离均为企业测试结果。根据表 1 数据可以看出在不使用 SWDM 收发器的前提下，OM5 光纤与 OM4 光纤支持的 40G/100G 最大传输距离相同。而且通过使用 eSR4 收发器的并行通信方式，OM4 光纤可以支持 550m(40G) 和 300m(100G) 的最大传输距离，均长于使用 SWDM 收发器下的 OM5 光纤的传输距离。因此在数据中心传输距离需要超过 150m 时 OM5 光纤并非唯一选择。此外，康宁根据对过去三年部署的 OM3 和 OM4 数据中心通道长度的统计发现，大约 95% 已部署 OM3 系统 90% 已部署 OM4 系统长度在 100m 以内（如图 3 所示）。总体来说 OM3/OM4 光纤能够满足 90% 的 40G/100G 数据中心连接需求。对于少数需要支持 300m 传输的数据中心，则可以使用 OM4/eSR4 或 OM5/SWDM 的组合方式。

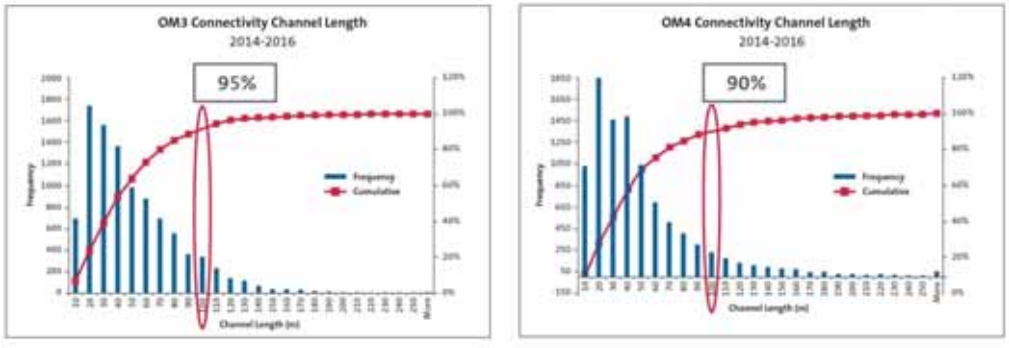


图 3. 数据中心以太网 OM3/4 连接通道长度分布

传输方式的选择

近年来云数据中心的发展持续驱动对高速数据传输的需求。传统上我们一直在使用加快光源开关速率的方式在单一通道内增加串行传输速率，例如从 1G 增加到 10G。但随着对更高速率网络需求的大幅提升，传统的串行传输方式已经无法满足数据中心发展的需要。目前能够支持 40G/100G 网络的传输方式主要可分为两类，一类是基于 WDM 技术的串行传输方式，一类是通过在多根光纤上同时传输数据的并行传输方式。我们在表 2 中简述了两种传输方式的优缺点和主要应用环境。

	WDM 串行传输	并行传输
传输方式	WDM 是一种利用多个激光器在单条光纤上同时发送多束不同波长激光的技术。如果将数据包比作车里的乘客，WDM 是通过增加每辆车的乘客数来提高运载能力。	并行传输是数据同时在多条光纤通道中发射和接收的一种通信方式。如果将数据包比作车里的乘客，并行传输是通过增加车道来提高运载能力。（车内的乘客数和车速保持不变）
优点	保持传统双芯 LC 连接，布线结构简单	可利用现有收发器从而降低成本
缺点	因需处理不同波长导致收发器的成本昂贵	布线系统相对比较复杂
应用	主要应用于长距离传输	主要应用于中短距离传输

表 2. WDM 串行传输与并行传输方式的对比

与选择光纤类型类似，选择哪种传输方式主要根据数据中心部署速率，传输距离以及整体预算等进行综合考虑。目前大多数已部署的 40G/100G 数据中心均采用并行传输方式。主要原因有如下几点。首先，采用并行传输可以满足绝大多数 40G/100G 数据中心对传输距离的要求。此外，并行传输使用的收发器可以利用现有收发器的市场规模优势降低成本。例如 40G 收发器只需使用 4 个相同的 10G 光源和光组件，既有产品市场规模大，成熟度高，相应的价格较低。如果使用 40G 波分复用收发器，则需要使用多个不同波长的光源，新波段的光源因处于产品生命周期的早期，成本较高。

并行传输方式被广泛使用的另一个主要原因是并行传输支持端口分支应用。使用并行传输方式的 40G/100G 光收发器既可以传输一路 40G/100G 信号，又可以同时传输 4 路独立的 10G/25G 信号。这种方式给用户带来的一个最显著的好处就是可以在相同的空间内提高 10G/25G 端口的密度。我们以 40G 网络为例来具体说明使用端口分支方式如何提高端口密度。

常规的带有 QSFP 收发器端口的 40G 交换机板卡可以提供 36 个 40G 端口。如果我们使用端口分支的方式将一个 40G 端口分支为 4 个 10G 端口，则 36 个 40G 端口可以支持 4×36=144 个 10G 链路。比较而言，常规的带有 SFP+ 收发器端口的 10G 交换机板卡可以提供 48 个 10G 端口。如果需要达到同样的 10G 端口数，则需要部署 3 个 48 口 10G 交换机。目前全球已出货的 40G QSFP 交换机大概有一半应用于 10G 端口分支。采用这种方式不但可以提高 10G 端口密度 3 倍左右，而且每个 10G 端口能耗可以降低 60%，成本可以降低 30%-45%。而且当数据中心未来需要升级为全 40G 网络时，无需重新购买 40G 收发器和交换机就可以平滑升级。同样的方式也适用于 100G 乃至未来的 200G/400G 网络（如图 4 所示）。因此，在可以预见的未来，并行传输方式仍将被大多数高密度数据中心所采用。

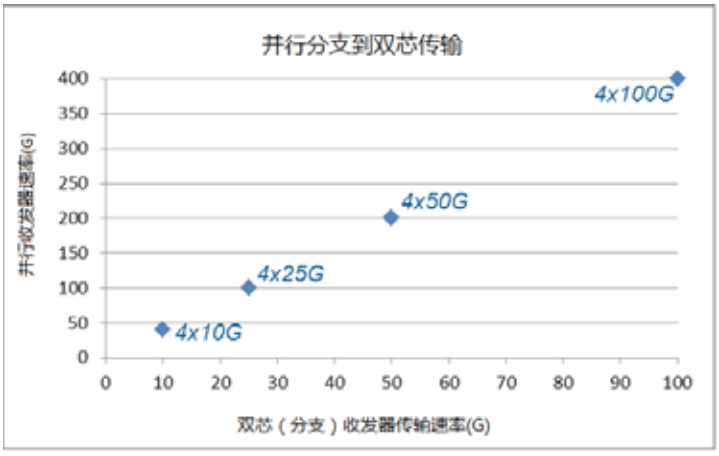


图 4. 并行分支应用演进

布线结构的选择

为了应对高速增长的数据处理需求，尤其是逐渐增加的数据中心内服务器之间东西向的数据传输需求，云数据中心开始采用有别于传统三层交换网络结构的新型网络来提高效率，降低时延。其中如图 5 所示的脊叶 (Spine-leaf) 两层网络结构近来被越来越多的云数据中心所采用。这种网络结构需要每一个脊交换机与每一个叶交换机相连，从而提高路由效率。随着云数据中心规模的不断增大，不但在一个数据中心内部需要大量的布线连接，而且在一个园区内的不同数据中心之间也需要大量的布线连接。数据中心内和数据中心之间的全交叉互联需求给布线结构的设计以及安装测试等都带来了新的挑战。

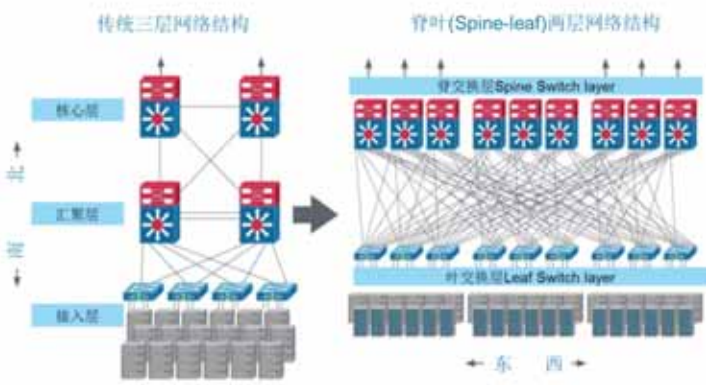


图 5. 传统三层网络结构与脊叶 (Spine-leaf) 两层网络结构的对比

CORNING

LANscape®
Pretium® Solutions

我们以图 6 所示的网络示意图为例来分析使用不同布线方案对用户的影响。如表 3 所示，我们对比三种不同的布线方式，LC 跳线端到端直连，72 芯 MTP 预端接光缆连接 MDA 和 HDA 配线区，576 芯 MTP 预端接光缆连接 MDA 和 HDA 配线区。通过对比我们可以发现，方案 2 和方案 3 采用结构化布线，预端接 MTP 光缆连接 MDA 和 HDA 配线区的方式相比较方案 1 跳线直连可以大大减少安装，标识，测试以及故障查找的工作量，从而降低安装测试和维护的成本，并且可以大量节省时间满足云计算数据中心对快速部署快速交付的要求。此外，在网络需要迁移增加和变更时，方案 1 需要重新在两层交换之间安装端到端的长跳线，工作量大安装难度高。而方案 2 和方案 3 只需在 MDA 或 HDA 配线区使用短跳线配置即可。

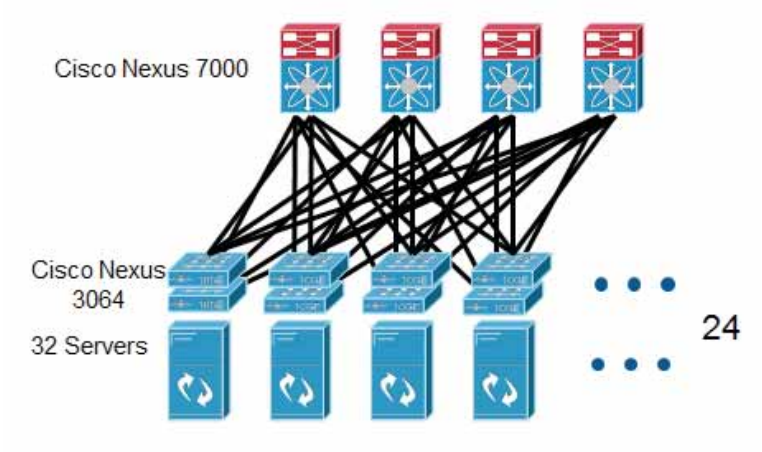


图 6. 脊叶两层交换网络配置示意图

（10G 网络，4 个脊交换机，48 个叶交换机，每个叶交换机带 32 个服务器，收敛比 1：1。）

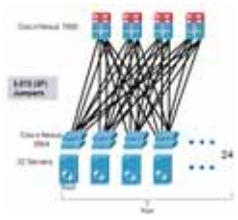
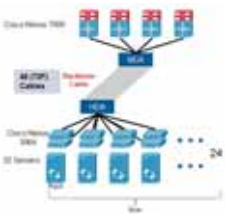
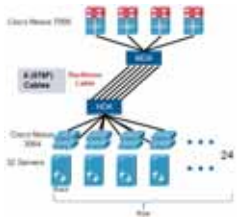
	方案 1 3072（2 芯跳线）	方案 2 48（72 芯 MTP 预端接光缆）	方案 3 6（576 芯 MTP 预端接光缆）
布线方案			
测试和清洁	6,144 (2F) 双工 LC 连接器	576 (12F) MTP 连接器	576 (12F) MTP 连接器
记录和标识	3,072 跳线 + 6,144 连接器	48 主干光缆 + 576 连接器	6 主干光缆 + 576 连接器
安装数量	3,072 根跳线	48 根主干光缆	6 根主干光缆
故障查找	3,072 个链路， >6000 连接器	48 个链路， 576 个连接器	6 个链路， 576 个连接器
网络迁移， 增加和变更	每次端到端安装一根长跳线	在 MDA/HDA 通过短跳线配置	在 MDA/HDA 通过短跳线配置

表 3. 布线方案对比

使用 MTP 预端接光缆的结构化布线方案对比 LC 跳线端到端直连的另一个优点是可以大量节省线槽资源（如图 7 所示）。对于服务于第三方的云计算数据中心来说，空间的节省就代表着成本的节省。在部署高密度的数据中心时，使用预端接方案，特别是使用高芯数预端接光缆可以有效节省线槽资源。不但便于安装铺设，更有利于整个数据中心系统的空气流动从而降低日常运行时的制冷成本。

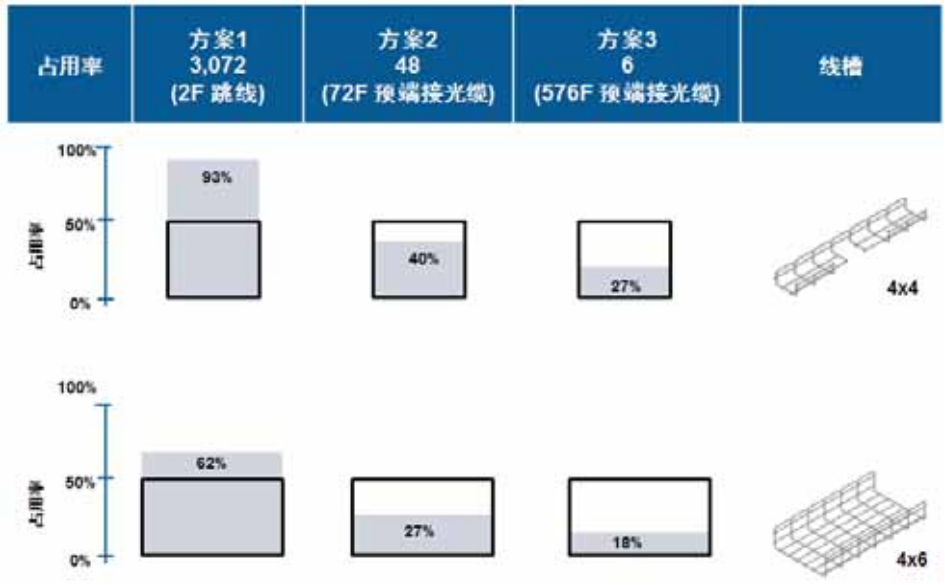


图 7. 不同布线方案线槽占用率对比

CORNING
LANscape®
Pretium® Solutions

康宁光通信中国
上海市漕河泾高新科技开发区桂箐路 111 号立明大厦 3 楼
电话：86 21 5450 4888
传真：86 21 5427 7898
www.corning.com