技术公告

构建性能更高、成本更低、按需自由扩展的云网络

——基于 Scale-wide 架构的交换网络

数据中心交换网络架构的第一次演进

在云计算勃然兴起的最初，云中的网络架构师并未意识到这将给底层支撑网络带来什么样的挑战。最初的云网络的载体——数据中心交换网络——依然延用了以往园区网络的架构：采用核心、汇聚、接入的三层架构、利用 STP 来避免环路产生、“上半部分三层路由+下半部分二层交换”的混合结构。随着云的规模越来越大、虚拟机数量的爆炸性增长、云中业务的大范围动态迁移、东西向流量的指数级增长，三级架构的底层支撑网络根本已经无法支撑云计算的继续发展。

以 CLOS 交换架构作为理论支撑的 Spine-Leaf 底层支撑网络应运而生。

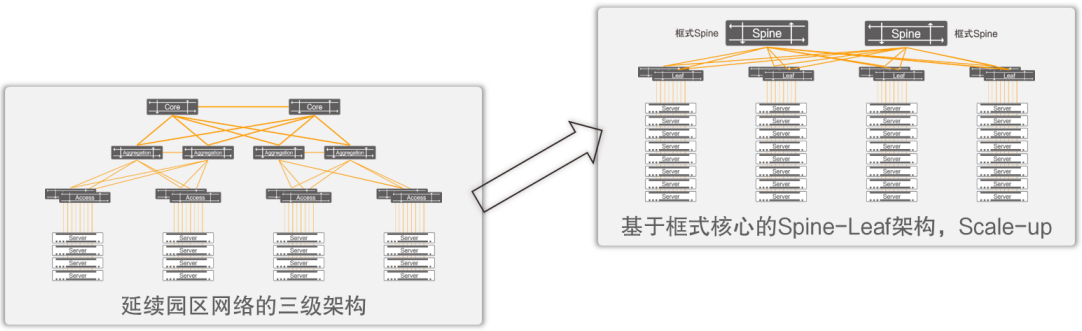

图 1：数据中心交换网络架构的第一次演进

在 Spine-Leaf 网络中，网络从三级被简化到二级，并且通过在 Spine 和 Leaf 交换机之间构建的 full-mesh 连接和充分利用 IP 网络非常成熟的负载分担技术（即 ECMP Routing，Equal-Cost Multi Path Routing），成功地化解了三级架构不能解决的问题：

- 将虚拟计算节点之间的通信以虚拟网络的模式承载在底层支撑网络之上，使得云的便捷和虚拟计算节点的数量不再受限于底层支撑网络；
- 云中业务在虚拟网络可达的范围之内可以任意迁移，迁移的过程中能够确保 IP 地址不发生变化，从而为业务的不中断打下坚实的基础；
- 两级网络架构极大地简化了交换路径，为虚拟计算节点之间的通信（即东西向流量）提供了高性能、高可靠的交换通道；

- 将交换域（即二层广播域）限制在 Leaf 交换机以下的部分，不再需要 STP 一类的协议进行环路检测，充分利用所有的带宽传递云中业务；
- 易于扩展，只需要扩展框式 Spine 交换机的接口卡和对应的 Leaf 交换机即可扩展底层支撑网络的规模。

至此，数据中心交换网络架构的第一次演进顺利完成；因其可扩展性主要通过框式 Spine 交换机的接口卡的纵向扩展来进行，一般也将其称为 Scale-up 架构的云网络。

Scale-up 架构的云网络再次陷入困局

Scale-up 架构的确很好地解决了云网络面临的各种问题。但是，在 Scale-up 云网络的具体实施和部署的过程中，新的问题又逐渐浮出水面。

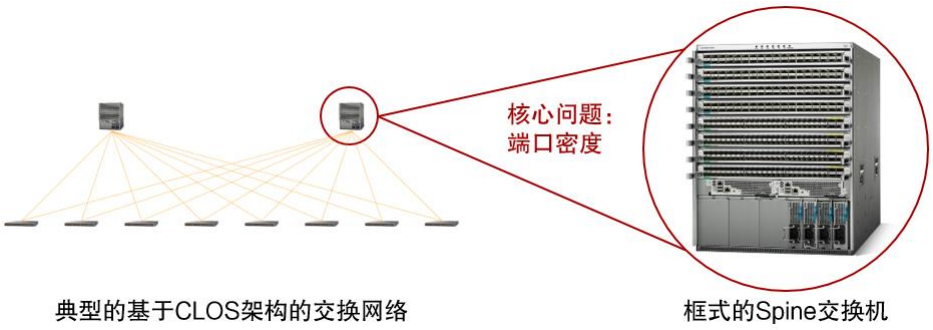

图 2：Scale-up 架构的云网络再次陷入困局

如上图所示，Scale-up 架构云网络的规模主要取决于 Spine 交换机的高速接口密度，为了提升密度，起初往往采用具有很高端口密度的框式交换机作为 Spine。这样的框式交换机存在三个问题：

- 高成本。因为需要在固定的空间内提供极高密度的高速接口，导致设备的设计复杂度大幅提升，而这样的设计复杂度最终折射在用户身上，就是动辄以百万人民币为单位的单机价格。
- 高功耗。因为集成度过高，因此框式设备的整机功耗远远高于可提供同样端口密度的盒式设备；同时，因为需要考虑各种条件下的散热设计，导致系统设计的复杂度进一步上升。
- 扩展性。框式设备的可扩展性主要取决于其接口卡的扩展能力，也就是说，当框式设备的接口卡槽位完全被占用后，其可扩展性（即 Scale-up）也随即消失，此时的再扩展就只能通过更换设备来完成。

另外，此类框式设备的运维也会带来额外的高成本，例如对此类的备件管理和运维流程都需要单独设计流程，无法与 Leaf 设备的相关流程保持一致性。

云网络通过 CLOS 架构找到了正确的发展方向，却在实际的实施和部署过程中，再次面临重重挑战；为了应对挑战、解决问题，需要新的思路来突破眼前的迷雾。

按需自由扩展的 Scale-wide 架构

云计算产业与生态中的先行者率先进行了勇敢的探索与大胆的尝试。这其中，对框式交换机体系结构的深入分析和合理解构为云网络的下一步发展指明了方向。

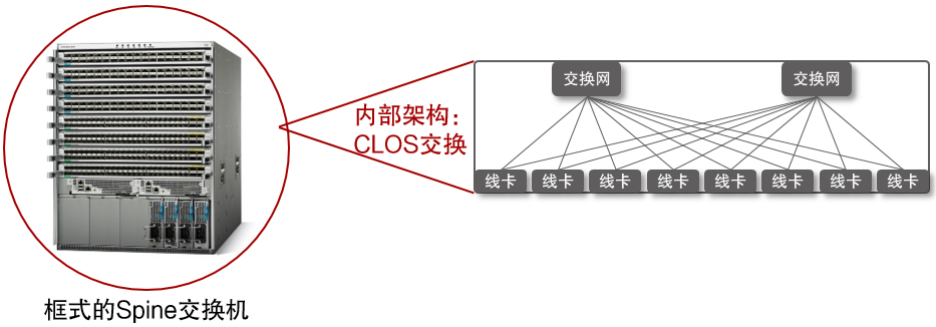

图 3：被禁锢在框式交换机机箱内的 CLOS

框式 Spine 交换机的内部体系结构，从本质上来说，仍然是一个 CLOS 架构的网络：基于大容量的交换芯片设计的交换网（也称为交换板、交换模块）是这个 CLOS 架构的 Spine 节点，为用户提供业务接口的每一个线卡（即接口卡）是这个 CLOS 架构的 Leaf 节点，而机箱内部的中置背板和正交连接器则是这个 CLOS 架构中连接所有 Spine 和 Leaf 的 full-mesh。框式交换机通过这个内部的 CLOS 架构，达到了提升端口密度、增大交换容量等目的，但同时也为此付出了更多的成本，最终使用户面临如前所述的困难。

在明确了这一点之后，一个自然而然的想法诞生了：为什么不把禁锢在机箱中的 CLOS 架构释放出来直接部署在网络中？这就是后来被称之为 Scale-wide 的按需自由扩展架构。

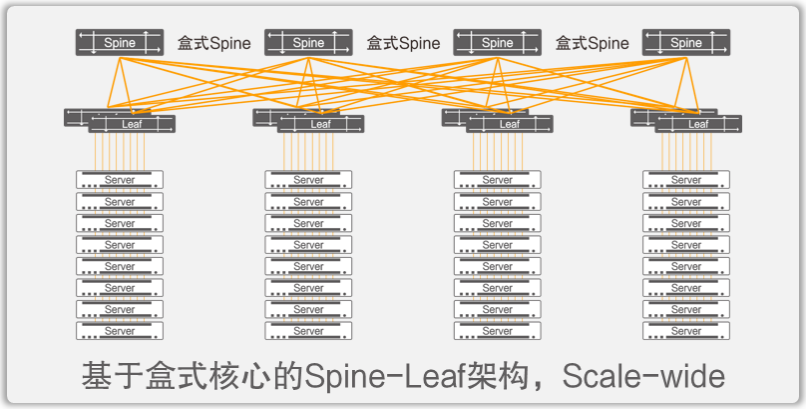

图 4：CLOS 被释放到网络中的 Scale-wide 架构

在 Scale-wide 架构的 Spine-Leaf 网络中，以框式交换机（多为两台）作为 Spine 的思路被彻底抛弃，取而代之的是以一组具备高速接口的盒式交换机作为 Spine 与底层的 Leaf 交换机构成 CLOS 交换架构。在一个能够提供同样的接口密度和交换性能的 Scale-wide 架构中，其 Spine 部分的总成本仅仅是框式设备的二分之一

到三分之一，但其总体功耗却大大降低，维护成本也因备件的简单性和设备的一致性大幅降低。更为重要的是，Scale-wide 架构为网络提供了极强的可扩展性，使得网络在极简盒式设备的支撑下，能够从很小的规模（例如 2 台 Spine+16 台 Leaf，连接 384 台双 10G 上行服务器）按需、动态地增长到很大的规模（例如 64 台 Spine+128 台 Leaf，连接 4096 台双 25G 上行服务器），而且，通过多级 CLOS 架构，Scale-wide 能够轻松地为超大规模的云提供模块化的网络支撑。

仅仅用了很短的时间，Scale-wide 架构就完成了其从理论到实验室原型、再到真实生产环境的发展过程。今天，在全球领先的公有云运营商的机房中，数十万、上百万租户所享用的云服务就运行在这一个个 Scale-wide 的网络模块之上。

基于 Scale-wide 架构设计的 Asterfusion 云网络

Asterfusion CX 系列云交换机采用最新的设计理念，将传统厂商禁锢在机箱内的 CLOS 交换架构彻底分散开来，采用按需自由扩展（Scale-wide）的架构帮助用户构建扁平化、大规模、按需弹性扩展的云网络。在 Scale-wide 的 Asterfusion 云网络中，用户不需要再为价格高昂的框式 Spine 交换机支付超额的成本，并且将网络的规划、部署、调整、优化、扩展的主动权牢牢掌握在自己手中。

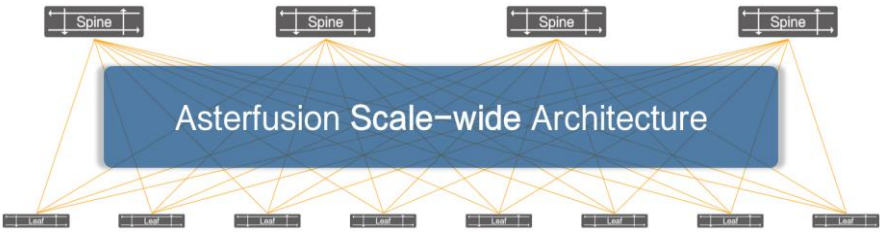

图 5：基于 Scale-wide 架构设计的 Asterfusion 云网络

CX 系列云交换机构建的扁平云网络极大地简化了云网络基础架构的复杂性，同时具备超强的按需自由扩展能力，大幅度降低云网络的 TCO，使云计算 Pay-as-you-grow 的基本理念在云网络中得以体现。

==========

关于星融元数据

星融元数据技术有限公司（Asterfusion Data Technologies Co., Ltd.）为云计算提供领先的、真正意义上的软件定义网络解决方案。凭借所拥有的专利技术，基于高度软件定义的整体架构、完全开放透明的操作系统、突破传统限制的硬件平台，星融元数据帮助用户重新定义云计算的网络基础设施，并且为云计算的使用者赋予真正弹性与超高性能的虚拟网络、为云计算的管理者提供纯粹的开放接口和自动部署调度、为云计算的开发者构建可编程和可视化的业务环境。

星融元数据的核心团队来自于中美两国多家知名的通信和 IT 企业，共执对客户需求的透彻理解、对产品技术的笃定热爱、对下一代云计算网络的美好梦想，源于中国、面向全球，致力于成为中国最优秀的云计算公司。星融元数据坚信云计算将为用户、产业与我们带来多方共赢的美好未来。

联系方式

苏州（总部）
苏州市工业园区星湖街 328 号
2 栋 B401
0512-62982976

北京
北京市海淀区宝盛南路 1 号
奥北科技园 20 号楼 207
010-62672668

西安
西安市曲江新区旺座曲江
C 座 2605
029-89834058

希望获取更多有关星融元数据公司、产品及解决方案的信息，请登录 www.asterfusion.com、或发送邮件至 sales@asterfusion.com、或关注我们的官方微信、微博。