

面向云网络设计的交换架构

概述

当今的网络要求具备可预测的性能，而且与支持设备静态编址的传统网络相比，新型网络对应用流的感知程度大为提高。过去的企业网络仅支持某些特定的应用，而数据中心的新型云网络则可支持多种应用。这与目前超额配置的网络相比显然是一种飞跃，因为延迟和传输时延较高始终是困扰现有网络的一大难题。

基于应用的可预测网络性能：

与基于传统Web (256KB传输)、电子邮件(1 MB)或文件传输(10Meg)的传统客户端—服务器设计不同，数据中心的新型云网络必须达到某个既定的性能指标。当今的应用程序在功能方面针对性较强，而且定义明确，比如需要超低延迟的高频算法交易或地震勘探分析应用。其他类似的例子还包括海量存储内容或虚拟机映像的迁移或web 2.0应用的大规模数据分析。这些数据中心需要无阻塞的可预测性能。对于交换架构来说，一个关键要求是保证物理和虚拟机之间应用性能的一致性。所有的节点都必须具备同等数量的无阻塞带宽和可预测的延迟。新型多核处理器对于网络带宽也提出了更高的要求。因此，一致的性能、太比特可扩展性、可预测的低延迟、无阻塞吞吐率，以及面向多1/10GbE和未来的40/100GbE的高速互连构成了云网络架构的基本特性。

数据中心交换架构的基础

目前云交换数据中心出现了两种交换架构：直通式交换架构在处理高性能计算集群(HPC)应用时提供超低延迟，而配有深度存储结构和虚拟输出队列(VOQ)机制的存储与转发交换架构则为基于Web的数据中心应用提供了必要的缓冲（图1）。Arista 7000系列交换机适用于低延迟的双层“主干-分支式”(leaf-and-spine)HPC网络设计。对于采用非对称千兆以太网和万兆以太网连接支持存储与Web应用的下一代高负荷数据中心，Arista 7048是最佳的解决方案。



图1a: 低延迟HPC网络应用

大型数据中心

- 非对称计算与存储链路速度
- 混合的1 GbE和10 GbE节点
- 高可用性、高扩展性
 - 冗余电源
 - 采用MLAG的冗余拓扑
 - 可靠的EOS交换机
- Arista 7500系列主干交换机
 - 容量高达10太比特
 - 低延迟
 - 深度缓冲
 - VOQ架构

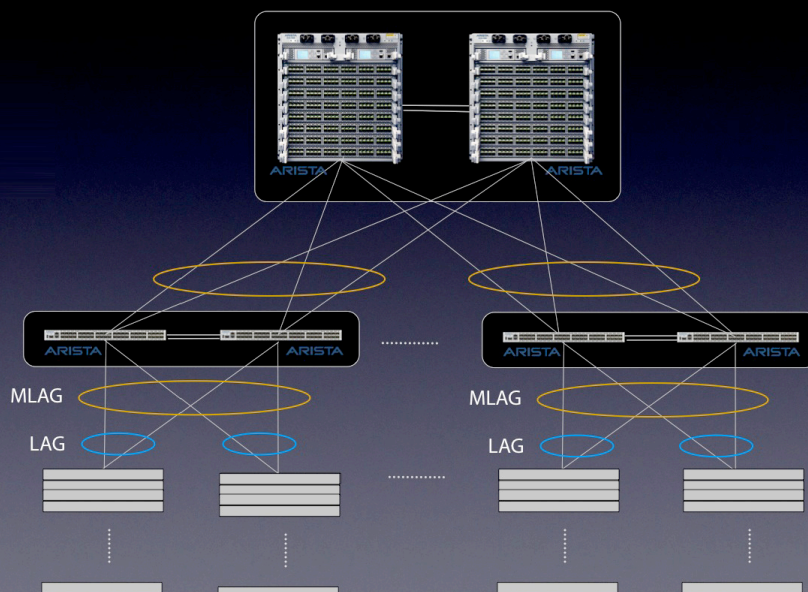


图1b:大型非对称数据中心设计

低延迟、高性能计算(HPC)集群:

当今金融服务领域的高频交易应用普遍采用实时市场数据馈送,这些应用需要直通式和共享存储交换技术来最大限度缩短延迟,使延迟保持在几微秒,甚至几百纳秒。万兆以太网架构的优势就在于能够提供最短的延迟,并最大限度减少端口级的缓冲。这意味着信息几乎即刻就能穿越网络到达数据馈送处理程序、客户端和算法交易应用。如果流量模式“中规中矩”而且比较对称,直通式交换是分支服务器的理想架构:例如HPC、地震分析和高频交易应用。这种架构假定网络的负载率不到50%,因此不会发生拥塞,保持低延迟至关重要。在直通式交换中,特别是在传输大型数据包和巨型帧数据包时,延迟低至几微秒。在“时间即金钱”的商业环境中,这种出色的低延迟特性可帮助企业节省数百万美元的成本。Arista 7100系列交换机可实现超低的延迟,因为数据包在到达之后将被立即转发,而不会在内存中进行缓冲。此外,该交换机还支持快速组播,同时能最大限度缩短排队和序列化延迟(图2)。

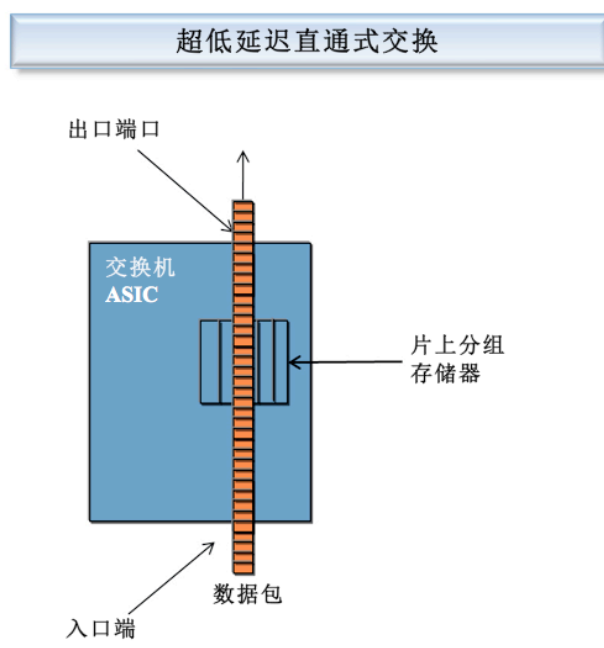


图2: 面向HPC集群的低延迟直通式交换架构

为下一代数据中心提供确定的性能：

对于高负载网络，例如“主干”应用、对存储、计算与应用“分支”资源的无缝访问，数据中心骨干网，为交换网络横向扩展提供可预测的统一性能是一项关键要求。最典型需要大型数据块的应用是：地图简化集群、分布式搜索以及数据库查询系统。一致性的性能是必须满足的要求。3-6微秒稍高的延迟是可以接受，因为传统的交换机有20-100微秒延迟，提供很差的性能。新的交换机架构为每个端口提供几兆的缓存，这就是为存贮转发系统提供最好设计，并进一步优化应用。当今的存储与转发交换架构采用虚拟输出队列(VOQ)，能够更好地协调“任意端口到任意端口(any to any)”流量。VOQ能够避免交换矩阵拥塞，这是传统交换机经常遇到的“头端阻塞”问题。VOQ方法与大型缓冲区相结合，为提高应用、整体网络行为的灵活性。在发生突发流量，或设备上高负载同时汇聚在通用服务器上时，大型缓冲区可有效地缓解拥塞。设备上高负载汇聚的一个例子是，当应用服务器同时收到来自一串存储服务器阵列的数据响应时，这些所有的响应措施将同时发起。在这种情况下，交换机必须拥有足够强大的缓冲功能，才能保存数据，避免数据丢失。深度缓冲对于万兆和千兆网络之间的非对称传输也非常重要，可容许链路速度的不匹配。

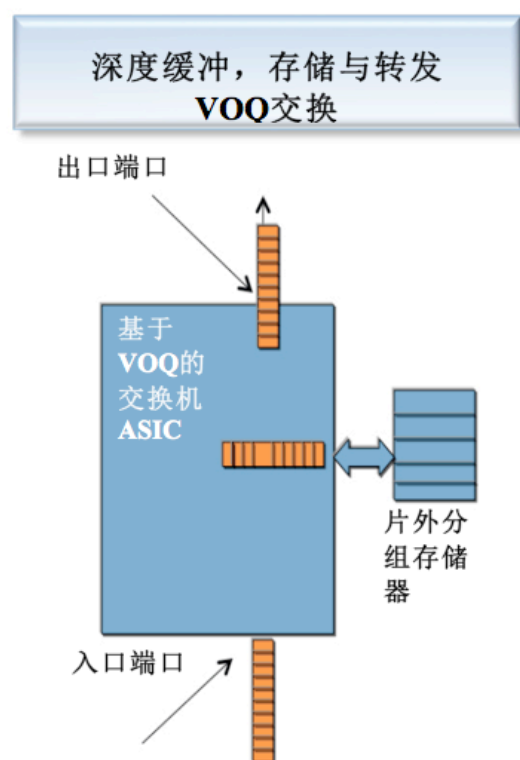


图3:大型数据中心要求采用大型缓冲区和VOQ/存储与转发交换，以支持非对称流量

云网络案例：

用户可以利用Arista的主干、分支式交换机来构建二层云网络(图4)。以大家颇为熟悉的一种重要社交网络应用Facebook为例：资料显示，他们建立了一个由3万台服务器构成的云网络，每个“memcache”（“内存缓存”）集群包含800台服务器，可生成5000万到1亿个请求，并访问28万亿字节的存储！Facebook没有使用访问时间超过5毫秒的传统数据库检索方案，而是采用了一个能将访问时间降至0.5毫秒的“memcache”架构！此外，减少重发延迟、采用永久连接也能改进性能。在这种环境中，应重点考虑配置可保证成功访问的大型缓冲区。对于大量存储、搜索、数据库请求和Web流量的应用，具有大型缓冲区、高级拥塞控制机制和VOQ架构的Arista 7048交换机和7508系列交换机是最理想的选择。

以下是一个关于低延迟、高频交易（HFT）应用的成功案例。该案例采用了能够根据实时标准（如时间、价格或订单数量）自动执行资金交易的方案。这些应用被对冲基金、养老基金、共同基金和其他交易机构所广泛采用。在运行过程中，应用会响应任何输入的内容或信息，并能在几分之一微秒内（几乎是在眨眼之间）处理一个交易。在证券和市场交易中，金融协议被广泛应用于实时国际信息交换。预计未来将有更多的应用会成为多线程应用，因此未来的云架构和交换架构必须要在计算集群节点之间实现低延迟互连。针对这些超低延迟HPC设计，Arista 7000系列交换机提供了完美的解决方案。

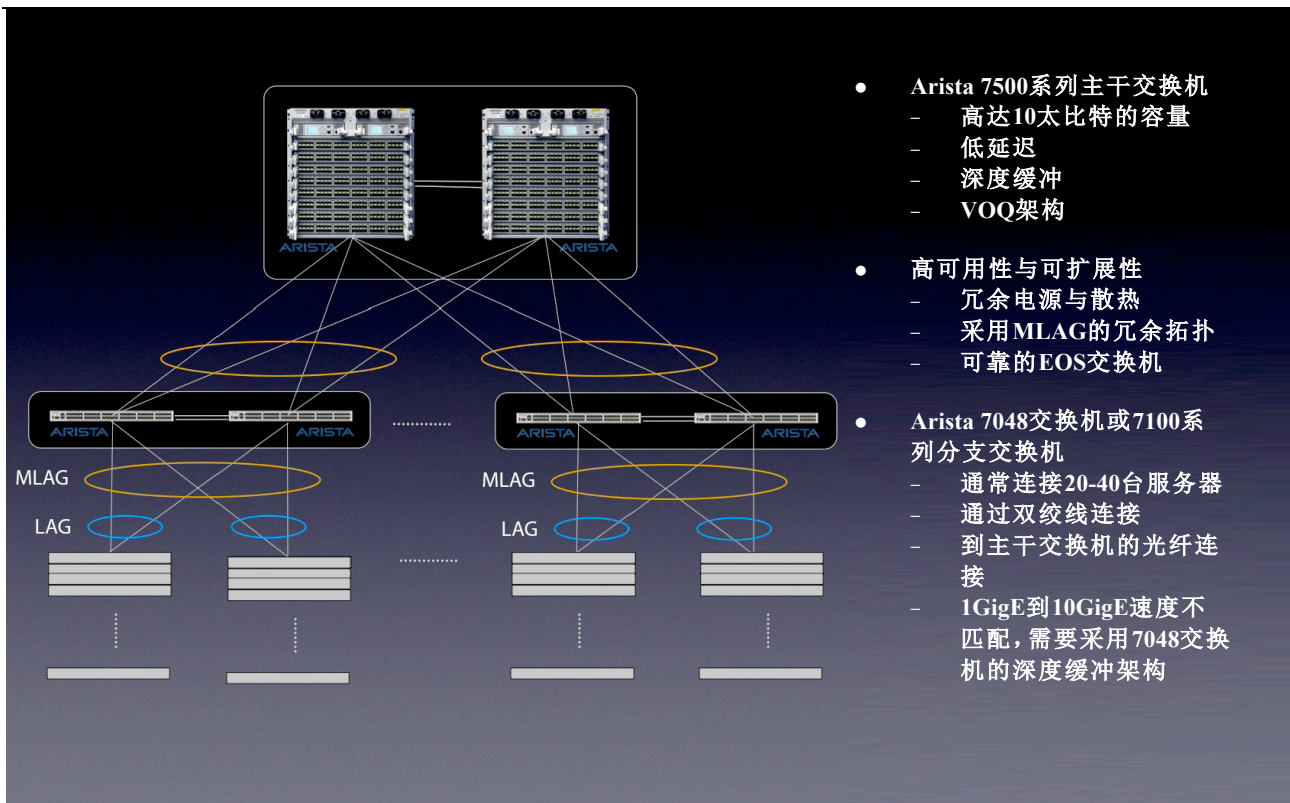


图4：基于Arista 7000系列交换机的云网络设计

总结

不断增长的“杀手级”云应用充分利用Arista面向数据中心的新型交换架构。这些应用包括：

- 高频金融交易应用；
- 高性能计算 (HPC)
- 集群计算应用
- 视频点播
- 网络存储访问
- Web分析、地图化简、数据库或搜索查询
- 虚拟化

上世纪90年代末设计的网络主要用于支持静态应用和电子邮件。当今的新应用和流量模式则是动态多变的，需要新的交换方法来支持实时应用访问。未来的云网络可优化以下变量：有保障的性能、低延迟和任意端口到任意端口的通信模式。Arista的交换架构和扩展的7000系列交换机致力于提供优化的云网络解决方案。

总部

5470 Great America
Parkway
Santa Clara, California
USA 95054
408-547-5500

支持

support@aristanetworks.com
408-547-5502
866-476-0000

销售

sales@aristanetworks.com
408-547-5501
866-497-0000

www.aristanetworks.com