



利用高效硬件横向扩展开源云环境

具有无状态卸载的高速以太网可在超大规模云服务中提供较高网络和服务效率

执行摘要

OpenStack（通常称为“云的 Linux”）使公司可以利用开源计划以及透明的协作方法来实施公共云和私有云解决方案，以便实现业务灵活性、基础架构弹性

和运营简单性。尽管 OpenStack 可提供灵活的框架，不过云基础架构（由计算、网络和存储资源组成）需要以最高性能和效率运行以保证整体应用程序性能，这点尤其重要。本文讨论云网络基础架构正确支持将 OpenStack 作为云管理平台的 Web 级 IT 的要求，以及对应的 Mellanox 解决方案。

部署效率挑战

为了实现多租户和自动化目标，云部署通常利用虚拟化技术，这类技术从计算虚拟化开始，然后扩展到网络和存储虚拟化。虚拟化可显著提高硬件利用率以及资源编排简单性，但是如果部署不正确，则可能会导致性能和效率下降。下降表现为数据通信和存储访问性能降低，以及 CPU 利用率提高或最终增加硬件占用空间和资本支出。

计算虚拟化问题

在虚拟机监控程序管理的虚拟化环境（与裸机相反）中，多个虚拟机 (VM) 实例会在物理服务器硬件上同时运行。

这便要求虚拟交换机软件（通常与虚拟机监控程序一起驻留在操作系统内核中）处理与 VM 之间的双向网络 I/O 流量。虚拟交换机确实可带来增强的灵活性，不过也会导致 I/O 性能下降（因为软件中的处理层增加）。随着服务器 I/O 过渡到并超过 10Gb/s，vanilla 虚拟交换机达到线路速率是一种挑战。在某些应用

程序需要小数据包处理能力（在理论上可以支持 15 Mpps 的 10 Gbps 链路上实达到的甚至会低于 1Mpps（百万个数据包 / 秒））。

网络虚拟化问题

软件定义网络 (SDN) 技术正在成为 OpenStack 云部署通过网络虚拟化改善网络可编程性、弹性和自动化的关键组件。主流 SDN 供应商采用了覆盖样式 SDN 来支持快速虚拟网络设置并最终支持云灵活性，而无需严重依赖底层物理交换机结构。覆盖 SDN 引入了新的隧道协议（如 VxLAN、NVGRE 或 GENEVE）。并非所有服务器网卡都可以识别并处理这些新的数据包格式，在不对覆盖隧道协议处理进行网卡硬件卸载的情况下，大多数虚拟网络 I/O 处理需要由 CPU 中的虚拟交换机来进行，从而导致不确定的低 I/O 性能，同时增加 CPU 负载。

存储虚拟化问题

新一代的云和大数据技术推动了对分布式软件定义、横向扩展和基于对象的存储的需求，并且促使组织更倾向于以太网速度和融合基础架构而不是原有基于光纤通道的存储网络。但是在以太网上通常使用的 TCP/IP 协议堆栈对于支持存储网络而言并不是最高效的。TCP/IP 源自广域网，在这种网络中，带宽曾经十分稀缺，拥塞是一种常见现象。因此，这些协议设计为包含端点之间的大量握手，几乎无法将所有协议处理操作卸载到网卡硬件中，因为复杂的协议软件需要在 CPU 中运行。这些情况可能会导致低存储访问带宽、低 IOPS 和高 CPU 开销。

为了克服这些挑战并实现最高云效率和应用程序性能，云运营商希望实施可提高出色的虚拟化、加速和自动化支持的高效虚拟网络解决方案。

主要优势：

Mellanox 与 OpenStack 集成，使客户可以充满信心地进行部署



借助 RDMA 使吞吐量提高 6 倍



借助 SR-IOV 将连接提高 20 倍



借助卸载将 CPU 利用率减少 80%

Mellanox OpenStack 云网络解决方案

通过一套端到端互连产品（如网卡、交换机、电缆/光纤以及关联网络驱动程序和管理软件），Mellanox 高效虚拟网络 (EVN) 解决方案使云数据中心可以通过具有丰富的网络卸载、加速和自动化功能的高性能、低延迟云网络，来实现最高效率。EVN 可以缓解上述虚拟化问题，提供的云网络可以按 10、25、40、50 和 100Gb/s 速度进行线路速率处理，以最少 CPU 开销支持高吞吐量、高 IOPS 存储操作，以便可以最大化专用于实际应用程序工作负载的基础架构资源。

Mellanox EVN 通过将在以下各节中讨论的虚拟化效率、加速技术和自动化技术来实现更高的云效率目标。



图 1. Mellanox EVN – 高效云基础架构的基础

虚拟化

Mellanox 虚拟化解决方案可缓解与计算、网络 and 存储虚拟化关联的性能问题，使云应用程序可以按最高性能和效率来运行。

克服计算虚拟化与 SR-IOV

单根 I/O 虚拟化 (SR-IOV) 有关的问题使设备（如网络适配器）可以在各种 PCIe 硬件功能间分离对其资源的访问。这使流量可以在虚拟机与其关联 PCIe 分区之间直接传递，从而使应用程序可以直接访问 I/O 硬件。因此，软件模拟层中的 I/O 开销会削减。SR-IOV 可以使 VM 实现几乎与非虚拟化环境中相同的网络性能。

Mellanox 网卡支持基本 SR-IOV 以及高级功能（如 SR-IOV 高可用性 (HA) 和服务质量 (QoS)）。SR-IOV HA 使用链路聚合组 (LAG) 将来自相同网卡上两个不同端口的两个 VF 绑定在一起并将捆绑作为一个 VF 公开给 VM，从而为 VF 提供冗余机制。当捆绑中的一个 VF 发生故障时，另一个 VF 会继续转发流量，而不会影响 VM I/O 操作。通过 SR-IOV QoS 功能可以基于 VF 对速率限制和带宽保证进行精细得多的控制。

使用 VxLAN 卸载、VTEP 网关和 ASAP2 克服网络虚拟化问题

自 ConnectX-3 一代的网卡开始，Mellanox 将 VxLAN 协议处理卸载到网卡。VxLAN 卸载功能使网卡可以进行 VxLAN 数据包的无状态处理（如校验和计算、大型分段卸载 (LSO) 等），从而显著提高吞吐量和改善延迟性能并减少与覆盖数据包处理关联的 CPU 开销。除了 VxLAN 之外，Mellanox 网卡还支持其他覆盖封装协议（如 NVGRE 和 GENEVE）的负载。

从 ConnectX-4 一代的网卡开始，Mellanox 以 VxLAN 卸载的成功为基础，引入了加速交换和数据包处理 (ASAP2)。此功能使虚拟交换机执行的大多数数据平面交换和数据包处理可以卸载到 Mellanox 网卡，同时保持虚拟交换机的控制平面操作。ASAP2 是相对于传统 SR-IOV 的显著改进，因为它可确保维持 SDN 和网络可编程性功能，同时网络 I/O 可在计算节点上实现最高性能。ASAP2 也是相对于 VxLAN 卸载的显著改进，因为除了 VxLAN 卸载支持的无状态卸载，它还可以支持其他数据包处理操作卸载（如 VxLAN 封装、解封和数据包分类）。随着在几乎一半的所有生产 OpenStack 部署中使用 OVS，将 ASAP2 与任何 OVS 实施结合使用会十分有意义，这样可在提高数据包吞吐量和降低延迟方面获得巨大的性能提升，并可提高云效率。

VxLAN 网络通常需要与其他网络（如支持裸机服务器的 VLAN 网络，或用于数据中心互连和南北用户流量的广域网）进行通信。这需要 VTEP 网关。Mellanox Spectrum 交换机在硬件中支持 VTEP 网关功能，从而可在云环境中的异构网络相互通信时确保最高性能。



图 2. Mellanox 与 OpenStack 供应商进行广泛合作并紧密集成

使用 RDMA/RoCE 克服存储虚拟化问题

与有状态协议（如 TCP）关联的较大开销导致它不适合于用作软件定义横向扩展存储应用程序的互连协议，尤其是在存储介质从硬盘到固态硬盘 (SSD) 再到非易失性存储器 (NVM) 越来越快时。另一方面，远程直接内存访问 (RDMA) 是为数据中心环境中的高速链路设计的协议，可以克服 TCP 的低效问题。RDMA

可以在 InfiniBand (IB) 或融合以太网 (RoCE) 上运行。借助 RDMA，所有数据传输操作都可以卸载到支持 RDMA 的网卡，从而保证以最低 CPU 开销获得最高可能吞吐量和最低延迟，这使它适合于存储访问。这通常在为无损

运行而配置的增强以太网上实施。但是，Mellanox 最近通过内置的错误恢复机制对 RoCE 进行了增强。虽然无损网络从不是严格要求，不过客户通常会配置其网络以防止数据包丢失并确保最佳性能。借助这一新版本，RoCE 可以部署在普通的标准以太网网络上。通过利用 RDMA 或 RoCE，虚拟服务器可以实现高得多的 I/O 性能，因为大多数数据包转发可以卸载

到网卡。这可进一步提高性能、缩短延迟并显著减少 CPU 负担。实际效果是提高了整体服务器和网络效率。

加速

Mellanox 适配器和交换机可提供运行基于 OpenStack 的高效且可扩展的公共云和私有云解决方案所需的网络加速。OpenStack 的灵活框架可由 Mellanox 高性能以太网解决方案进行补充。它们合在一起组成云基础架构（包括以最高性能和效率运行的计算、网络和存储资源）。

这可保证整体应用程序性能，同时通过高达 100GbE 的端到端产品来支持现在和未来的 Web 级 IT。

自动化

云网络资源的设置和管理是实现业务灵活性和基础架构弹性部署目标的至关重要的组成部分。这意味着能够快速进行行动或反应，需要运营简单性。Mellanox NEO™ 是一种用于数据中心网络编排的强大平台，旨在简化网络设置、监视和操作。这包括通过 REST API 与 OpenStack 组件和核心服务轻松集成。NEO 的网络 REST API 技术使 OpenStack 服务可以访问结构相关数据和设置活动。这样可以实现简化但强大的自动化功能（从网络筹划和启动到日常操作）。

Mellanox 提供
RDMA 功能，
与传统性能相比，
可将存储性能最多
提高 6 倍

开放式可组合网络

OpenStack 部署的目标是提供一组高性能、高可编程性的网络组件，包括交换机、服务器适配器、光学模块和电缆。开放且可组合的关键之处在于支持开放式 API 和标准接口，以及将硬件与软件分离并允许选择网络操作系统。Mellanox 对此方法进行了补充，可完全真正地将组织从供应商锁定策略中解放出来（一直到交换机芯片级别）。Mellanox Spectrum 交换芯片允许选择热门的开源操作系统（如 Cumulus），同时提供同类最佳的硬件（它们可提供可避免的零数据包丢失、公平合理的流量分配，并且与其他竞争对手交换机中的商用芯片相比，可提供可预测性更高的应用程序性能）。

Mellanox NEO™（网络编排和管理软件）是用于管理横向扩展网络的强大平台。Mellanox NEO 使数据中

心操作员可以高效地设置、监视和操作现代的软件定义数据中心结构。NEO 可充当该结构的接口，从而扩展现有工具功能并增强配置和设置交换机和路由器的能力。通过集成到常用高级云网络软件解决方案（如 OpenStack Neutron 和 VMware vSphere）中，NEO 可增强设置并且无需人为干预。

结论

作为高性能网络技术领域的行业领先者，Mellanox 了解使数据中心转型的风险和回报。随着 IT 组织过渡到基于云并且以服务为中心的基础架构，对提高网络和服务效率的需求在超越 10Gb 服务器 I/O 时是至关重要的。通过结合来自适配器和交换机的重要技术，Mellanox 能够通过基于硬件的无状态卸载加速虚拟网络并降低 CPU 利用率，以便在现代的软件定义数据中心中提高可扩展性并提高灵活性。

关于 Mellanox

Mellanox Technologies (NASDAQ: MLNX) 是针对服务器、存储和超融合基础架构的端到端以太网和 InfiniBand 智能互连解决方案和服务的领先提供商。Mellanox 智能互连解决方案可提供最高吞吐量和最低延迟、更快地向应用程序传递数据并充分发挥系统性能，从而提高数据中心效率。

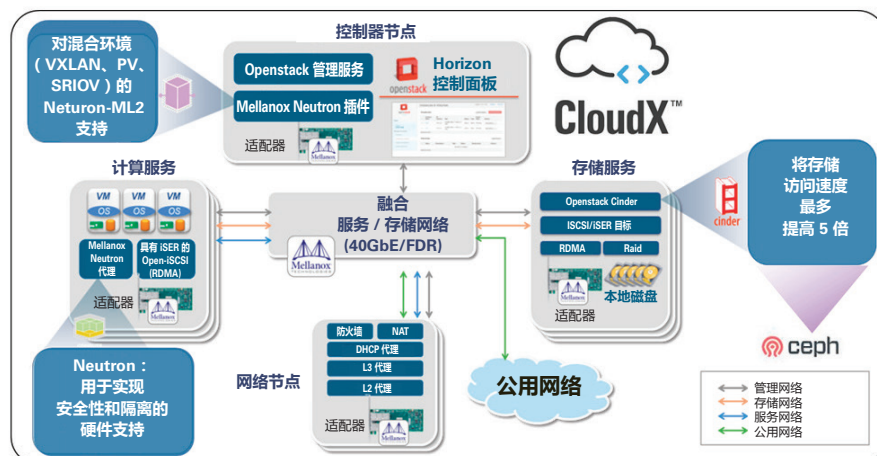


图 3. Mellanox 在其产品与 OpenStack 服务之间提供无缝集成

了解有关 Mellanox 和 OpenStack 的详细信息

Mellanox OpenStack 参考体系结构：

<http://www.mellanox.com/openstack/pdf/mellanox-openstack-solution.pdf>

Mellanox Red Hat OpenStack 参考体系结构：

<http://www.mellanox.com/related-docs/whitepapers/Mellanox-OpenStack-Solution-for-Red-Hat.pdf>

Mellanox 和 OpenStack Ceph 白皮书：

http://www.mellanox.com/related-docs/whitepapers/WP_Deploying_Ceph_over_High_Performance_Networks.pdf



北京迈络思科技有限公司

咨询电话：+86-10-57892000

销售咨询：china_sales@mellanox.com

市场合作：marketing_cn@mellanox.com

*欲了解更多欢迎登陆www.mellanox.com



微信

MLNX-31-426