



Microsoft 存储空间

借助支持 Mellanox RDMA 的网络解决方案加速应用程序

执行摘要

企业中业务关键型信息、非结构化数据和丰富内容的扩张不断改变着存储动态。这种趋势在推动对较高连接速度的需求，以便满足进行数据访问的应用程序要求。为了应对这些市场状况，Microsoft 与 Mellanox 通过消除 CPU 开销以及为企业应用程序提供更快且更高效的数据访问，来携手解决存储访问问题。

Microsoft SMB

该解决方案以 Windows Server 2012 R2 中对 Microsoft 服务器消息块 (SMB) 3.0 协议的增强为基础。SMB 协议进行了扩展，包括两个新功能，即 SMB 直通和 SMB 多通道。SMB 直通使用各种高速 RDMA 方法，在几乎无需 CPU 干预的情况下传输大量数据。SMB 多通道使文件服务器可以同时使用多个网络连接，并通过用于根据需要动态添加连接的网络路径自动发现来提供容错。这两个功能的添加使支持 Mellanox RDMA 的 IO 适配器可以通过优化标准以太网上服务器与存储之间的数据传输来提供线路速率性能并提高可用性。通过提供对 Mellanox RDMA 的内置支持，Microsoft 使客户可以在低成本的文件服务器上部署存储，同时提供较高性能。这样可实现胜过昂贵低效的光纤通道 SAN，并且具有较低延迟和运营成本的存储。

提高服务器效率

Mellanox ConnectX®-4 Lx EN 以太网适配器支持采用 10/25/40/50 GbE 速度的 RDMA，而 ConnectX®-4 网络适配器提供 10/25/40/50 和 100Gb/s 以太网连接，可为高性能 RDMA 实施提供性能最高且最灵活的解决方案。RDMA 使操作系统无需处理用于数据传输的协议堆栈，使应用程序可以直接读取和写入远程虚拟内存以及直接交换消息，从而提供了低延迟。这样可显著降低 CPU 在数据 I/O 请求中的参与程度，并减少内存带宽瓶颈。Mellanox RDMA 技术可在适配器硬件中处理所有传输协议，从而完全绕过主机操作系统。这样便使适配器可以在主机 CPU 参与程度最低的情况下在服务器与存储之间传输数据。因此，SMB 直通在客户端到文件服务器性能方面的速度极快，几乎相当于使用本地连接存储的解决方案。CPU 在这种处理中的参与程度降低可腾出更多周期用于服务器应用程序。因此每台物理服务器上承载的 VM 数量更大，同时运行的 VDI 实例更多，并且 SQL 环境可实现高性能文件共享以更快地完成查询。数据中心效率和可扩展性在所有这些情况中都得到了提高。

提高网络效率

Mellanox 交换机能够在不丢弃数据包的情况下以完全线路速率处理虚拟化和容器化的数据包，从而确保网络 and 应用程序性能不受影响。交换机上的卓越缓冲区和所有端口间可预测的带宽分配使它们可以更好地处理争用问题。这使交换机可以采用可预测的方式，公平合理地划分缓冲区和带宽资源。例如，当微型猝发或 incast（多对一）广播发生时，网络不能允许一个应用程序或客户端使用大多数网络容量，以及意外地允许其他对象缺乏资源。而且，在虚拟化环境中，几百台服务器可能会承载数万个 VM。这会对需要访问转发表来完成传递的交换机施加巨大

Mellanox 互连通过以下方面提供高效率：



将 CPU 效率提高 33%



将 IOPS 提升到超过 100 万



将延迟减少 50%

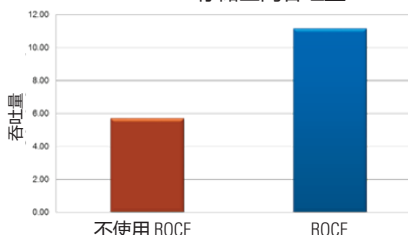


提供持续的近乎线路速率带宽



通过半宽交换机最大程度实现灵活性整合

Microsoft 存储空间吞吐量



- 使用 RoCE 时性能可提高 2 倍
 - 带宽提高 2 倍并且 CPU 效率提高 2 倍
- RoCE 可实现完整闪存存储带宽
 - 无损的远程存储

压力。这在网络在高负载下运行时，可能会导致服务器网络性能下降。Mellanox 交换机提供完整的 L2/L3 交换、路由和数据中心桥接功能，这些功能可消除网络虚拟化问题并提高整体网络效率。

Hyper-V 使用案例

该使用案例的用途是展示在 Hyper-V 中运行工作负载时的 CPU 效率。该演示利用两台服务器（客户端和文件服务器），它们使用运行 RoCE 上的 SMB 直通（融合以太网上的 RDMA）的 ConnectX®-4 100Gb/s 以

在启用以及未启用 RoCE 的情况下运行概念证明。结果证明，在 RoCE 上的 Hyper-V VM 上运行的工作负载可以实现超过 100 万 IOPS 的吞吐量并维持近乎线路速率的 100Gbps 总带宽而不会对 CPU 造成沉重压力。与未启用 RoCE 时的结果相比，延迟减少了几乎 50%，CPU 效率提高了 33% 以上。CPU 由于需要处理数据传输协议而受到束缚；这些处理包括序列检查、数据包丢失恢复、CRC 检查以及将内核传递给用户空间传输。在 CPU 中运行数据传输堆栈占用了本应该用于应用程序的周期，限制了整体工作效率。Mellanox 适配器可将 CPU 从这种处理中解放出来，从而提供更高效率。

RoCE 优势

通过在 Mellanox RDMA 技术的基础上使用 Microsoft 的 SMB 3.0，组织可以提供极高等级的存储性能和弹性。RoCE 需要的 CPU 利用率非常低，可在数据中心桥接以太网中利用优先级流量控制实现无损连接。使用 Mellanox RDMA 配置 Microsoft SMB 使 Hyper-V 中运行的工作负载可以实现 100 万以上的 IOPS。这不是什么新奇的想法，游

戏行业在很长时间以来一直在卸载图形处理。自 21 世纪中期以来，GPU 加速的计算已用于加速科学、分析和工程应用程序。将这些强大的功能添加到您的技术库中可实现随时准备好应对现代数据中心的存储需求的全面平台。

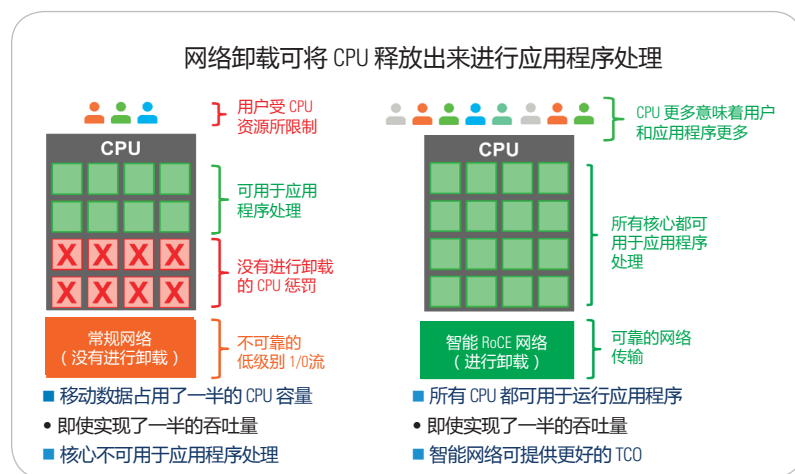


图 1 — Microsoft 和 Mellanox 100Gb/s 云效率演示

以太网适配器进行背靠背连接。三个 Micron NVMe SSD 300GB 6G SAS 适配器用于为主 I/O 生成器和基准测试工具组成全 SSD 存储池和 IOMeter。为了模拟真实环境，使用了不同工作负载来提供具有静态和混合 I/O 的数据流。分配了四个核心来运行数据通信任务，

有关详细信息

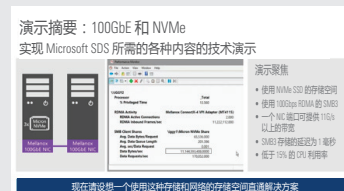
下载 [Microsoft 白皮书](#)

下载 Mellanox 案例分析：http://www.mellanox.com/related-docs/case_studies/CS_MHA_Cloud.pdf
http://www.mellanox.com/related-docs/case_studies/CS_Financial_Institute.pdf

若要了解有关 Microsoft SMB 的详细信息，请访问：[https://technet.microsoft.com/enus/library/hh831795\(v=ws.11\).aspx](https://technet.microsoft.com/enus/library/hh831795(v=ws.11).aspx)

技术演示

在 Microsoft 的首席计划经理 Claus Joergensen 提供的 Microsoft 2015 年科技点亮未来大会演示文稿中可以看到有关该技术演示的详细信息。



“使用具有本地磁盘的服务器实现私有云存储”



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