Reimagining Scalable, Low Latency Storage Using NVMe™ Over TCP

利用 NVMe™ over TCP 技术重新构建可扩展的低延迟存储

Recently Sagi Grimberg, co-founder and CTO at Lightbits Labs, engaged in a Q&A discussion on the future of data centers and delivering greater performance, flexibility, and lower TCO for NVMe over Fabrics (NVMe-oF™)-based disaggregated storage. Below is a summary of the conversation:

最近，Lightbits Labs 联合创始人兼首席技术官 Sagi Grimberg 针对数据中心的未来，以及如何为基于 NVMe over Fabrics（NVMe-oF™）技术的解耦合存储实现更高的性能、灵活性和更低的总体拥有成本（TCO）等话题发表了自己的观点。以下是他对一些问题的具体回答。

How important is NVMe-over-TCP?

NVMe-over-TCP 有多重要？

NVMe™/TCP is very important in my mind. Folks have wanted an alternative to Fibre Channel Storage Area Network (FC SAN) that is performant, but much more affordable and allows consolidation of data center networks and practices. The Internet Small Computer Systems Interface (iSCSI) offers SAN with Ethernet TCP/IP networks but has been disappointing in its overall performance characteristics. NVMe/TCP on the other hand, offers the speed and low latency of NVMe, but more importantly preserves the ubiquity of Ethernet and capitalizes on the well-established base of TCP/IP networking knowledge and practices. In short, it is an industry standard, becoming widely supported, ubiquitous, performs better than the legacy iSCSI and Fibre Channel SAN, and leverages standard Ethernet, which costs less and offers

higher bandwidth and flexibility compared to Fibre Channel.

在我看来，NVMe™/TCP 非常重要。人们想要一种光纤通道存储区域网络（FC SAN）的替代方案，这种方案要具备高性能，但价格要实惠得多，同时能够支持数据中心网络和实际操作的整合。Internet 小型计算机系统接口（iSCSI）技术可以通过以太网 TCP/IP 网络实现存储区域网络（SAN），但其总体性能令人失望。另一方面，NVMe/TCP 可以提供 NVMe 的快速和低延迟特性，但更重要的是其保留了以太网的普遍性优势，并可以利用完善的 TCP/IP 网络知识和实践基础。简而言之，它是一个行业标准，具有很强的普适性且正在得到广泛的支持，其性能优于传统的 iSCSI 和 FC SAN，同时可以利用标准以太网，与光纤通道技术相比，以太网成本更低，带宽和灵活性则更高。

Commented [CD1]: 建议像这种存储人都知道的名称就不要翻译了，反而看的很累。包括后面的 iSCSI， SAN 这些词汇

Commented [CD2R1]:

How is this technology used?

如何使用这项技术？

NVMe-over-TCP can be used anywhere iSCSI is used but provides greatly improved latency and much higher levels of IOPs on the very same Ethernet/TCP networks. Moreover, it has a good fit for highly transactional workloads (databases, analytics and message streaming) as well as high bandwidth (real-time analytics, video processing, AI/ML) workloads with emphasis on large-scale deployments and higher network speeds and feeds.

可以在任何使用 iSCSI 的场景中使用 NVMe-over-TCP，其可以在完全相同的以太网/TCP 网络上提供显著改善的延迟特性和高得多的 IOPs 性能。此外，它非常适合于高度事务性的工作负载（数据库、分析和消息流）以及高带宽工作负载（实时分析、视频处理、人工智能/机器学习），同时重要的是，其可以进行大规模部署，且拥有更高的网络速度和传输能力。

Who currently offers, or is planning to offer, the technology?

谁目前提供或计划提供这项技术？

Lightbits Labs was the first to offer a production NVMe/TCP based solution, but other smaller and larger companies also have products available, or have announced products that will support NVMe/TCP. Vendors include Pure Storage (future), Dell/EMC (future), NetApp have discussed NVMe/TCP in blogs, Infinidat, Fungible, Pavilion Data and more. In addition, Network vendors such as Intel, Mellanox (Now NVIDIA), Marvell, SolarFlare (Now Xilinx), Kazan-Networks (Now WD) have also announced support, offloads and enhancements for NVMe/TCP.

Lightbits Labs 是第一家提供量产的 NVMe/TCP 解决方案的公司，但其他大型或小型的公司也有可用的产品，或者已经宣布其产品将支持 NVMe/TCP。包括 Pure Storage、Dell/EMC、NetApp 等供应商已经在博客中讨论了 NVMe/TCP，其他供应商还有 Infinidat、Fungible、Pavilion Data 等。此外，英特尔（Intel）、迈络思（Mellanox，现已被英伟达 NVIDIA 收购）、美满电子（Marvell）、SolarFlare（现已被赛灵思 Xilinx 收购）、Kazan-Networks（现已被西部数据 WD 收购）等网络供应商也宣布了针对 NVMe/TCP 的支持、卸载（offload）特性和增强功能。

What is the target market?

有哪些目标市场？

The target market is exceedingly broad. First of all, the modern data centers look more and more like a cloud as we've come to know it, with enterprises adopting cloud practices as well as cloud native deployments in environments such as Kubernetes or Openstack®. It's hard to imagine the cloud built on top of dedicated Fiber Channel networks, so NVMe/TCP becomes much more applicable due to its superior performance and latency compared to iSCSI (as well as other TCP based storage protocols). Secondly, as the maturity and ecosystem evolves for NVMe-oF and NVMe/TCP specifically, also traditional bare-metal deployments currently built on iSCSI or Fiber Channel SANs will benefit from adopting NVMe/TCP.

NVMe/TCP 技术的目标市场非常广阔。首先，随着企业在 Kubernetes 或 Openstack®等环境中采用云实践和云原生部署，现代数据中心看起来越来越像我们所了解的云。很难想象在专用光纤通道网络之上构建云，因此 NVMe/TCP 变得更加适用，因为相比 iSCSI（以及其他基于 TCP 的存储协议），其具有更卓越的性能和延迟特性。其次，随着 NVMe-oF 尤其是 NVMe/TCP 技术的逐渐成熟及生态系统的不断发展，目前构建在 iSCSI 或 FC SAN 上的传统裸金属部署也将因采用 NVMe/TCP 而受益。

What are the potential advantages compared to other NVMe-over-fabric options like Fibre Channel or Infiniband?

相比光纤通道或 Infiniband 等其他 NVMe-over-fabric 技术，NVMe/TCP 有哪些潜在优势？

NVMe/TCP's biggest advantage over NVMe Remote Direct Memory Access (NVMe/RDMA) is simplicity. NVMe/TCP runs on every NIC under the sun. It lowers the barrier for users to evaluate different products without requiring non-standard practices or specific HW. Another advantage is scalability. Infiniband has been traditionally deployed as a backend fabric when it comes to storage systems where the scale is more limited and the freedom for specialization is acceptable. Front-end fabrics however, require strict ubiquity and typically higher scalability. That is why TCP/IP and FC are used much more broadly there. The advantage compared to Fiber Channel-NVME is mainly cost reduction, consolidation and increased bandwidth.

相比 NVMe/RDMA（RDMA 即远程直接数据存取），NVMe/TCP 最大的优势是简单。NVMe/TCP 可以在任何一个网络接口卡（NIC）上运行。它为用户降低了评估不同产品的门槛，且无须进行非标准实践，也无需特定硬件。NVMe/TCP 的另一项优势是可扩展性。传统上，当涉及规模更有限且专业方面的自由度可以接受的存储系统时，Infiniband 会被部署为后端结构。然而，前端结构需要更全面的普适性和更高的可扩展性。这就是 TCP/IP 和光线通道技术在前端得到更广泛

应用的原因。相比光纤通道 NVMe 技术，NVMe/TCP 的优势主要是成本更低、带宽更高且具备整合能力。

What are the potential roadblocks to adoption?

采用 NVMe/TCP 的潜在障碍是什么？

Unlike some of the storage vendors, VMware and Microsoft do not currently support NVMe/TCP yet. Thus today, NVMe/TCP is really limited to Linux environments. Additionally, you need a fairly modern kernel/distribution (such as recent versions of RHEL, SUSE, Oracle Linux, etc.) to have the full NVMe/TCP driver and multipath capabilities. These “roadblocks” should evaporate as organizations upgrade to newer releases and when Microsoft and/or VMware support NVMe/TCP.

和一些存储供应商不同，VMware 和 Microsoft 目前尚不支持 NVMe/TCP。因此，现在 NVMe/TCP 实际上仅限于在 Linux 环境中使用。此外，你需要一个很新的核心/发行版本（例如最新版本的 RHEL、SUSE、Oracle Linux 等），才能拥有完整的 NVMe/TCP 驱动程序和多路径功能。随着用户机构将 Linux 升级到更新的版本，以及 Microsoft 和/或 VMware 开始支持 NVMe/TCP，这些“障碍”应该会消失。

Commented [CA3]: 这篇文章因为实在和 VMware 的合作之前发的，所以这里的描述并不准确，请看看如何修改这段话。

Commented [CD4R3]: 你就把 VMware 删除吧，后面写可以在 Linux 及 Vmware 上使用，但是要比较新的版本

At the same time, NVMe-oF still has some gaps compared to the mature and complete SCSI and FC standards mainly around in-band authentication as well as automated discovery and enumeration that are important mainly in enterprise environments. These are areas that are actively being worked on as we speak. Having these gaps addressed, in combination with the evolving ecosystem support, and capable products becoming predominant in the market, will make NVMe-oF and NVMe/TCP viable options for almost every deployment out there.

与此同时，相比成熟、完整的 SCSI 和光线通道标准，NVMe-oF 仍存在一些差距，主要在于带内身份验证以及在企业环境中很重要的自动发现和枚举功能。这些都是我们目前正在积极进行改善的领域。弥合这些差距，再加上生态系统支持不断提升，以及功能强大的产品在市场上占据主导地位，将使 NVMe-oF 和 NVMe/TCP 成为几乎所有部署的可行选择。

Are the major storage vendors planning to adopt this technology?

主要的存储供应商是否计划采用这项技术？

Commented [CA5]: 这个问答很简单，建议删去

Yes – see above.

是的，请看上文的介绍。

What does the future look like for NVMe-over-TCP?

NVMe-over-TCP 在未来会如何发展？

I believe that the ubiquity of Ethernet & TCP/IP will naturally drive people toward NVMe/TCP. If you could have the simplicity of iSCSI but with substantially more IOPs and much lower latency on the same Ethernet fabric, why wouldn't you switch?

我相信无处不在的以太网和 TCP/IP 将自然而然地推动人们走向 NVMe/TCP。如果你可以拥有 iSCSI 的简易性，而且在同样的以太网结构上可以拥有更高的 IOPs 和更低的延迟特性，你为什么会不替换呢？

We're seeing the entire ecosystem spectrum, hardware, system and platform vendors, as well as customers making substantial investments in NVMe-oF and specifically in NVMe/TCP. I believe it is no longer a question of "if" but rather a matter of time until it will be the de-facto standard for block storage over the network.

我们看到整个生态系统，硬件、系统和平台供应商，以及客户都对 NVMe-oF，特别是 NVMe/TCP 进行了大量投资。我相信这不再是一个“是否”的问题，而是时间问题，它将会成为网络块存储的事实标准。

As trailblazers in this field, Lightbits Labs NVMe/TCP solution has been successfully tested and deployed at industry leading cloud data centers. Lightbits Labs NVMe architecture provides efficient and robust disaggregation with low latency, delivering data faster to applications and unlocking system performance capability.

作为 NVMe/TCP 领域的开拓者，Lightbits Labs 的 NVMe/TCP 解决方案已经在行业领先的云数据中心进行了成功测试和部署。Lightbits Labs 的 NVMe 架构可以提供高效、强大的解耦合功能和低延迟特性，从而能更快地向应用传输数据并释放系统性能。

Commented [CA6]: 这句建议删去，因为在文中没有出现过 PCI

英文原文链接：

https://www.lightbitslabs.com/blog/reimagining-scalable-low-latency-storage-using-nvme-over-tcp/?utm_content=177545881&utm_medium=social&utm_source=linkedin&hss_channel=lcp-10588524