

NVMe 现代 SAN 之入门

作者: NetApp 技术营销工程师 Mike Kieran
NetApp 技术营销工程师 Michael Peppers

2018 年 5 月



引言

在企业数据存储领域，新技术再次引发变革。

这次的新兴技术是 NVMe（非易失性内存标准 [Nonvolatile Memory Express] 的英文缩写），它是一种新的存储访问与传输协议，可为业务关键型企业级应用程序提供最快速的响应。

与闪存存储出现时一样，NVMe 也将为企业数据存储系统带来重大速度提升。但这一次，变革的影响范围可能更大，因为 NVMe 不只是一种存储规范；基于网络结构的 NVMe (NVMe over Fabrics, NVMe-oF) 协议覆盖范围更广，重新设计了从服务器到存储系统的整个数据路径。

在这篇入门简介文章中，我们将介绍 NVMe 技术如何帮助企业构建具有运行核心企业级应用程序所需的性能、可靠性、可用性和可扩展性的存储区域网络 (SAN)。

您将了解 NVMe 技术如何：

- 为固态驱动器 (SSD) 定义全新高速数据传输协议，以便与企业或客户端系统进行通信
- 通过网络结构，提供从主机软件堆栈到存储阵列的更高 IOPS 和更低延迟
- 利用速度日益加快的光纤通道传输及其他网络技术
- 为您的企业提供运行业务关键型应用程序所需的一流的 SAN 性能

全新企业要务

如今，很难找出有哪家大型企业不进行“数字化转型”，他们通过审视业务的各个方面，了解可以在哪些地方利用数据来加快上市速度、降低错误率和改善客户体验。

在某些情况下，企业正通过更新客户关系管理或企业资源规划等现有应用程序来达到此目的。也有些企业正在部署全新工作负载，其中包括移动、视频、自动机器人、实时分析、人工智能和物联网等。

无论哪种使用情形，创建、聚合、共享和存储的数据量都在快速增加。更为重要的是，这些数据的价值随着时间的推移呈指数级增长。如今，世界上最宝贵的资源不再是石油，而是数据。

这正是 NVMe 在这个最佳时机出现在数据中心的原因所在：

- 它可以助力 IT 组织为内部客户提供对关键业务工作负载的实时响应。
- 它支持企业发现并利用会对业务产生重大影响的机会，从而加快实现价值。

目前，IT 主管负责升级现有基础架构、以服务的形式提供一切，以及启用新工作负载，而且通常几乎没有额外的预算。这正是 NVMe 技术在现代数据中心内至关重要的原因。它可以将延迟（访问 SAN 上的应用程序数据所需的时间）从毫秒级缩短至微秒级，从而加快关键应用程序的响应速度。

利用 NVMe 技术降低延迟还可以在诸多使用情形中增加业务价值，其中包括：

- **实时分析。**零售商和信用卡公司每天要处理数十亿的借记卡和信用卡交易，因此他们迫切希望防止欺诈行为的发生。NVMe SAN 是一种支持接近瞬时响应的技术，可避免在店内和线上造成巨大的潜在损失。
- **软件代理。**智能软件机器人正越来越多地提供高质量的客户交流、技术支持和其他交互式服务，而且速度往往比人工服务更快。未来，NVMe 技术将使系统能够快速、准确地做出响应，让用户分辨不出他们是在与人交流还是与机器人交流。
- **物联网。**基于 NVMe 的存储系统可以快速载入海量数据集，而且支持同时连接大量客户端。因此，它们既可作为边缘数据采集点，也可用作集中式存储池（可以容纳通过聚合所有传感器数据所形成的数据湖）。

“与 SCSI FCP 相比，基于光纤通道的 NVMe 将 IOPS 提高了 58%，将延迟降低了 34%。”

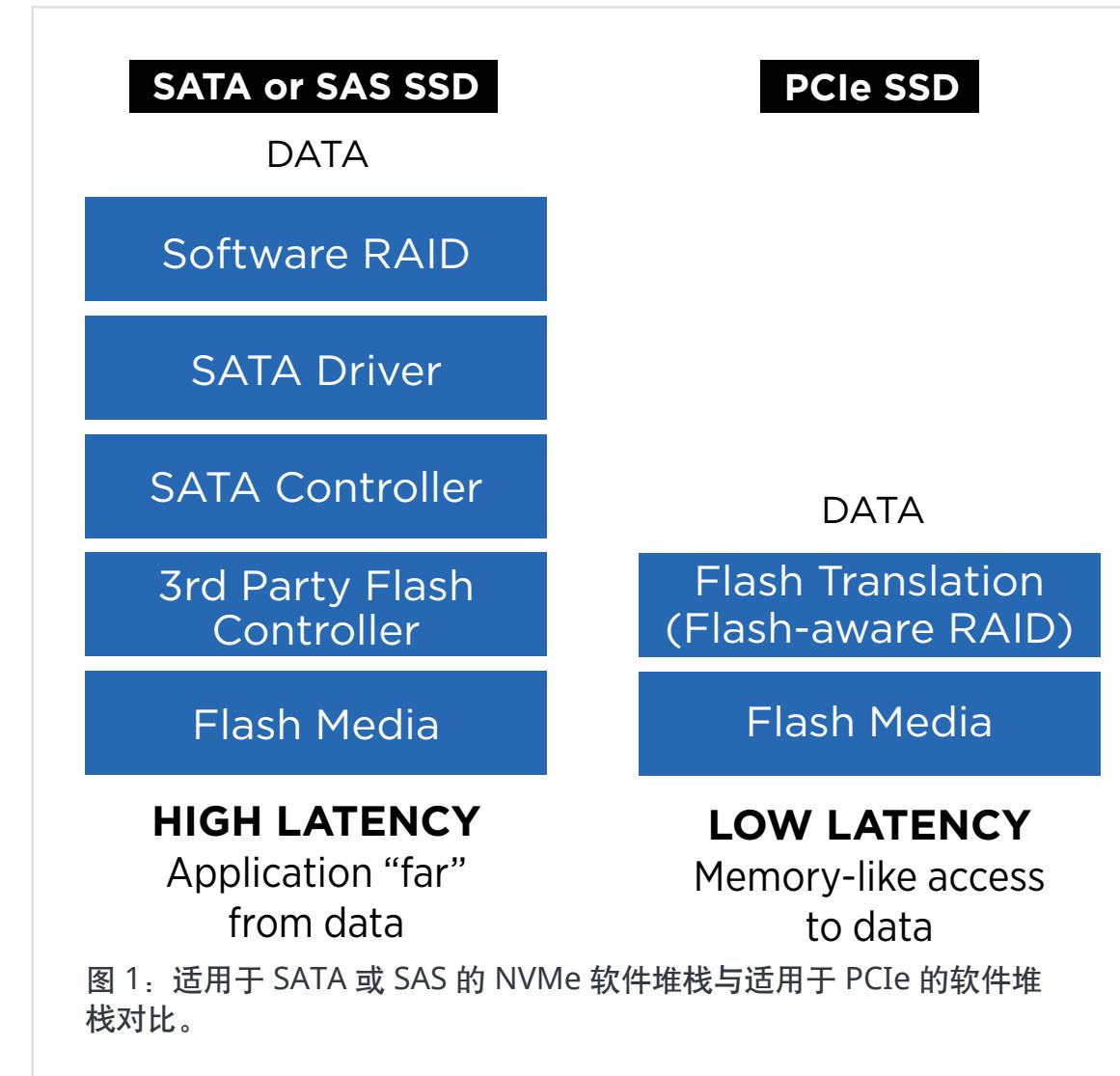
— Demartek 分析师报告《基于光纤通道的 NVMe 的性能优势 — 一种全新的并行高效率协议》，2018 年 5 月

NVMe 与固态驱动器

将闪存技术引入数据中心时，首先采用的是混合磁盘闪存阵列形式，然后是全闪存阵列形式，两种形式都对企业级应用程序的性能产生了巨大影响。闪存存储还改变了最终用户对性能的预期。使用支持闪存的 PC 的用户已经习惯了更快的响应速度，即使这些应用程序变得越来越复杂，而且从数据中心扩展到了云中，他们仍期望在使用所有业务应用程序时获得相同体验。

不过，因为数据路径的其他部分存在瓶颈，所以 SSD 可能带来的速度提升正接近极限。目前的两种主要网络协议（光纤通道和以太网）均使用 SCSI 命令集来实现存储协议。SCSI 是将近 40 年前为机械介质开发的，虽然它能处理进出旋转磁盘的数据流，但已经无法满足全新闪存存储介质的需要。

原因在于 SCSI 将 I/O 请求放入一个队列，其中最多包含 256 个命令。当来自应用程序的 I/O 请求到达时，它们只能排队等待其他请求处理完毕。在旋转磁盘时代，这种数据访问延迟的确无关紧要，因为驱动器的读写磁头一直都



引言

全新企业要务

NVMe 与固态驱动器

基于网络结构的 NVMe

数据库、SAN 和 NVMe

业务成果

在旋转，忙于在快速旋转的盘片上寻找微小的磁块。闪存及其他固态介质没有移动部件，因而固有延迟非常低，所以数据路径中的延迟成为新的瓶颈，导致数据访问速度变慢。

过去十年间，随着快捷外设互联标准 (Peripheral Component Interconnect Express, PCIe) 总线的出现，这种情况有所改观。该标准在速度上远超 SATA 和其他传统协议。例如，PCIe 总线支持多达 65,535 个队列，每个队列的队列深度可达 65,535 个命令，而且可以将存储直接连接到 CPU，从而提供内存式访问。此外，PCIe 所需的软件堆栈远远小于传统协议（见图 1）。

如今，基于 PCIe 的设备广泛应用于各种计算环境，这可能会给 SSD 带来问题，因为每个供应商的硬件都需要使用专有驱动程序。为了消除兼容性问题，业内所有重要厂商一致同意对运行于 PCIe 总线上的闪存以及其他形式的固态内存采用新的现代数据访问标准：NVMe 规范。

NVMe 实际上包含两个截然不同但又相互交织的层面：

- 一个针对固态存储设备进行了优化的协议和指令集
- 一套适用于非易失性内存系统的前瞻性开源架构标准

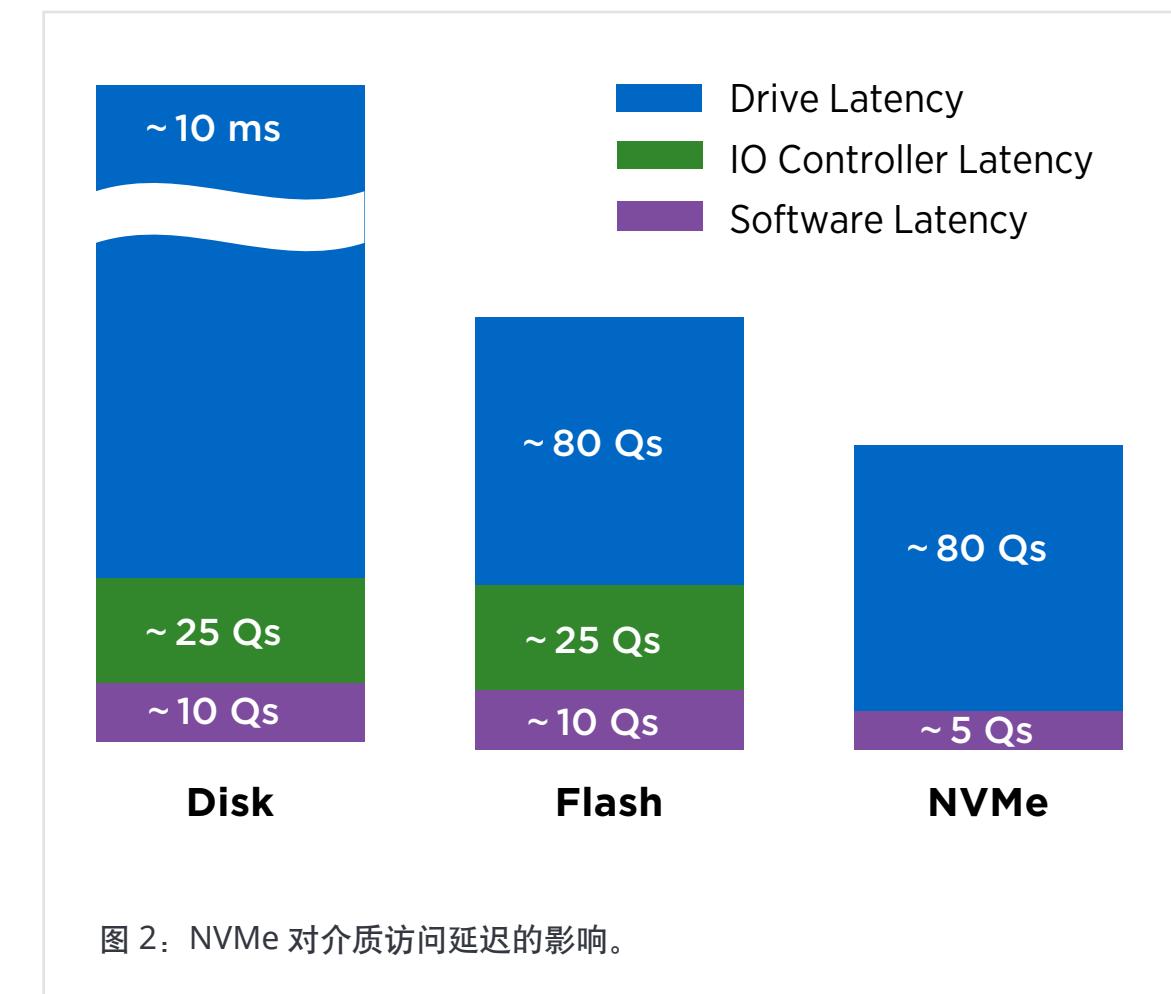


图 2: NVMe 对介质访问延迟的影响。

引言

全新企业要务

NVMe 与固态驱动器

基于网络结构的 NVMe

数据库、SAN 和 NVMe

业务成果

NVMe 规范是从头开始设计的，目的是在从移动电话到网络级服务提供商的各种计算环境中利用非易失性内存。此外，由于 NVMe 可实现更大吞吐量和更低延迟，因此正快速地在数据中心内得到应用（见图 2）。

在 NVMe 上投入了大量投资的公司包括 NetApp、Broadcom、Intel、Samsung、Micron、Seagate、Cisco、Western Digital、Microsoft 和 Toshiba。NetApp 目前在其大多数混合闪存和全闪存存储系统中使用 NVMe 存储，可显著提高这些系统的性能。

总之，NVMe 正在推动着一场架构变革，实现与存储系统的大规模并行通信。最终结果是在服务器和存储设备之间提供带宽更高、延迟更低的连接。

此外，其目的在于满足未来需求：NVMe 不仅支持当前的 NAND 闪存和 3D XPoint 技术，而且随着目前正在开发以及尚未发现的持久内存技术的兴起，也将能与它们协同工作。

当然，优势不止这些！

在许多其他因素的共同作用下，NVMe 得以在数据中心提供无与伦比的性能，其中包括：

- 中断处理
- 序列化 I/O 请求所需的内部锁定
- 命令简化
- 上下文切换减少
- 无锁定设计
- 轮询模式

基于网络结构的 NVMe

到目前为止，我们着重讨论了 NVMe 协议如何加快数据存储速度。但 NVMe 同时也是一种数据传输协议，而且是最新一代高速数据网络结构的关键组成部分。

NVMe 规范在不断发展演变；随着基于网络结构的 NVMe (NVMe-oF) 的问世，NVMe 的性能和延迟优势现已扩展到包括以太网、光纤通道和 InfiniBand 在内的整个网络结构中。NVMe-oF 增加了对前端存储接口的支持，数据中心的 NVMe 设备和子系统之间可以进行更远距离的通信，并且可以横向扩展到大量 NVMe 设备。

随着网络本身变得越来越快，这些增强功能将发挥至关重要的作用。例如，光纤通道的速度在不断加快，一开始，第 5 代达到了 16 Gbps（每秒兆位数），目前已发展到第 6 代 (32 Gbps)，而且第 7 代也即将问世。与此同时，以太网的速度也增加到了 100 Gbps 及以上。

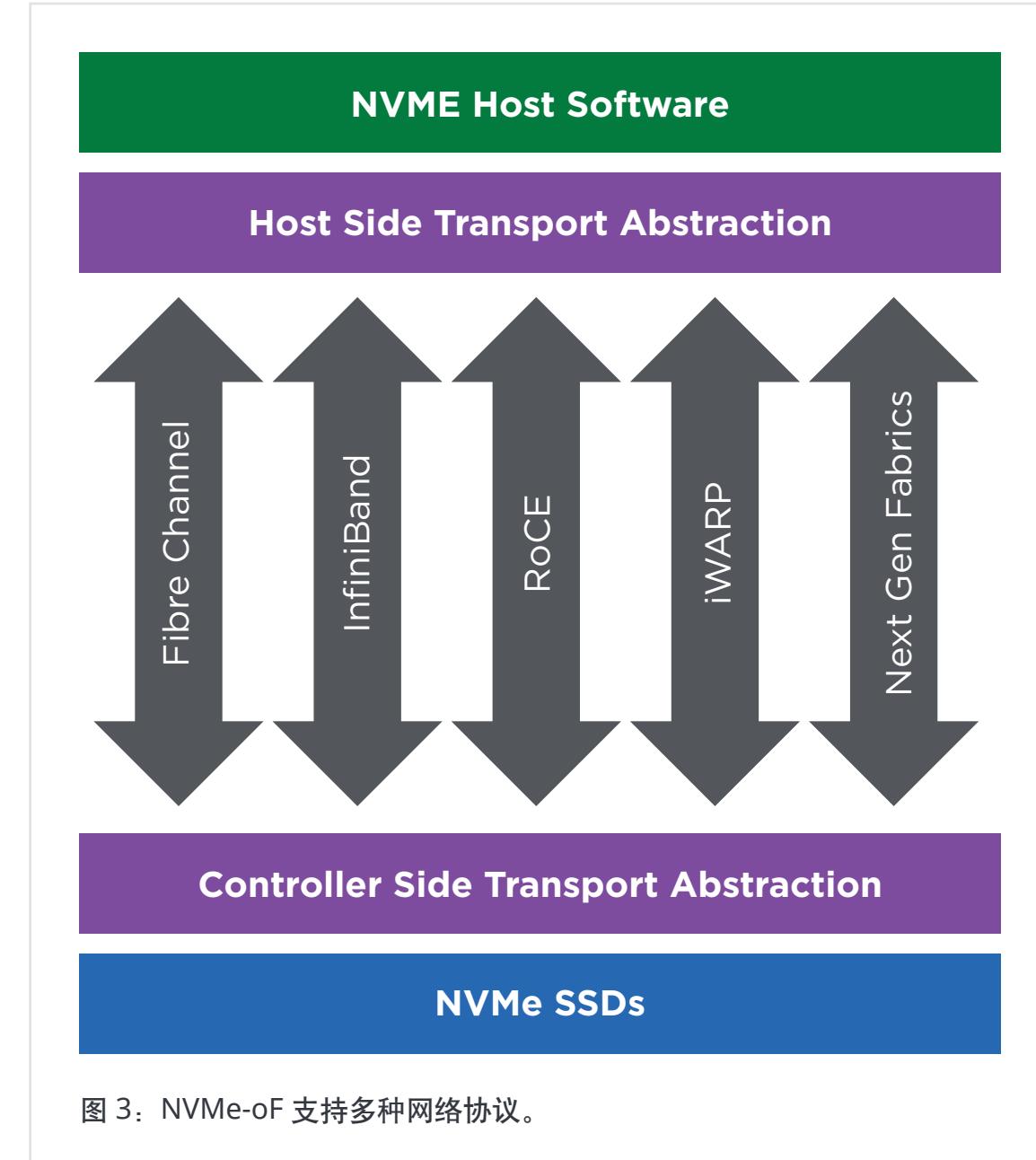


图 3：NVMe-oF 支持多种网络协议。

引言

全新企业要务

NVMe 与固态驱动器

基于网络结构的 NVMe

数据库、SAN 和 NVMe

业务成果

消除接口瓶颈

NVMe-oF 虽是一项相对较新的技术，但其底层组件在企业数据中心已非常成熟。一个重要因素在于它不受底层传输机制的限制。尽管许多企业 IT 供应商专注于使用光纤通道 (NVMe/FC)，但是 NVMe-oF 标准也支持远程直接内存访问 (RDMA) 传输机制，例如 InfiniBand、RoCE 和 iWARP（图 3）。

过渡到 NVMe/FC

对于转向 NVMe-oF 的许多 IT 架构师来说，光纤通道仍是首选网络技术，因为它不仅提供高性能和可靠性，而且支持基于网络结构的分区和名称服务。更大的优势在于，光纤通道能同时传输多个更高级别的协议，例如 NVMe/FC 和 FCP（基于光纤通道的 SCSI）。

对一些企业而言，业务优势是判断是否采用端到端 NVMe 系统的主要依据。端到端 NVMe 系统一般采用将 NVMe-oF 与光纤通道或 RDMA 相结合的多个 NVMe 存储设备。无论是从 IOPS 还是从延迟来看，系统都将拥有闪电般的速度。

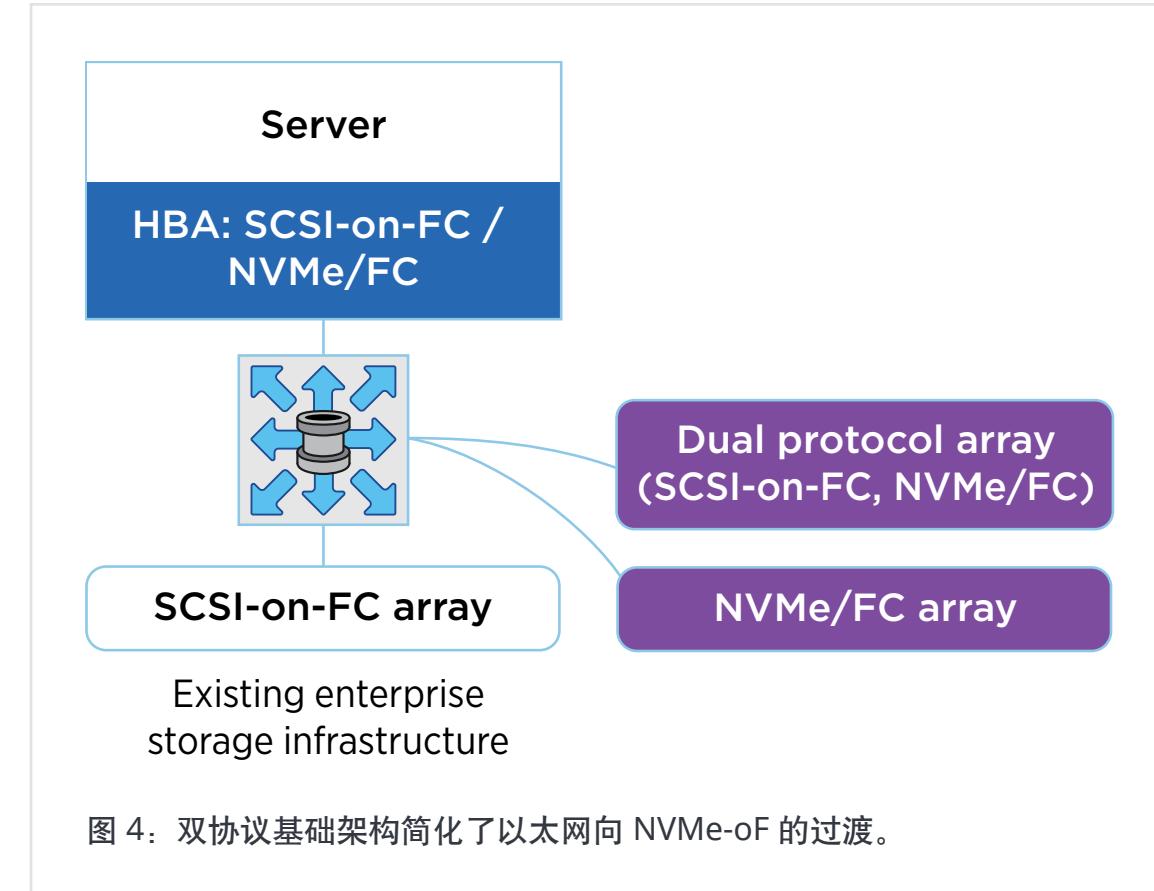


图 4：双协议基础架构简化了以太网向 NVMe-oF 的过渡。

NVMe 的多协议支持简化了您在进行如此重要的迁移时将面临的购买决策。关键在于能否利用企业的现有基础架构。对于许多企业而言，过渡到 NVMe-oF 需要几年的时间，但这不是问题。

引言

全新企业要务

NVMe 与固态驱动器

基于网络结构的 NVMe

数据库、SAN 和 NVMe

业务成果

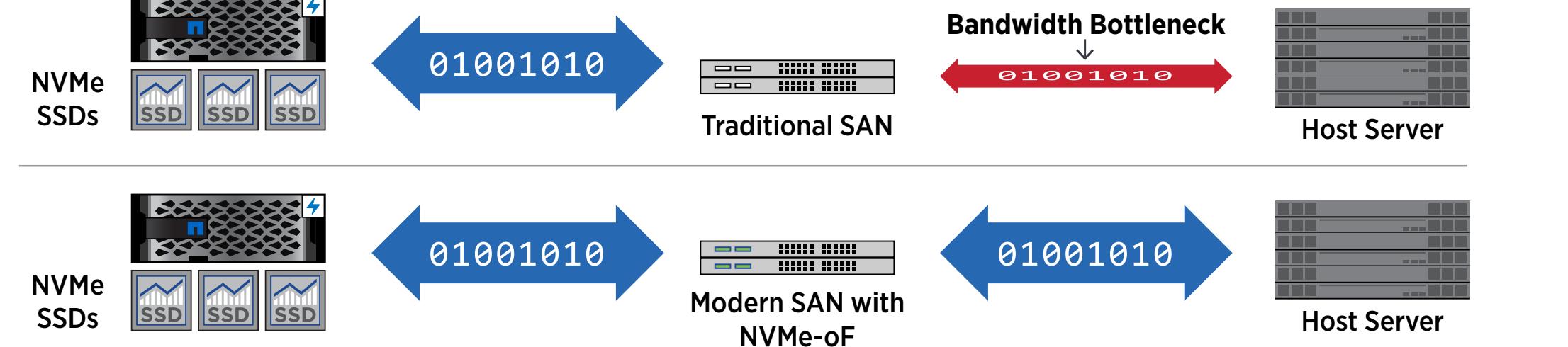


图 5：完整的 NVMe/FC 解决方案提供从服务器直接到存储系统的低延迟数据路径。

与用一种技术代替另一种技术相比，更好的解决方案是采用双协议光纤通道网络结构，运行并发的 FCP 和 NVMe/FC 流量。这种方式提供了稳定且经济高效的升级路径，消除了关键技术过渡期间的不确定性和风险。

NVMe/FC 解决方案堆栈

虽然 NVMe 是一项颠覆性技术，但不会颠覆您的持续运营，您可以无中断地将其引入到您的企业中。

如图 5 所示，端到端 NVMe/FC 解决方案堆栈有三个主要组件：服务器、存储系统以及将它们连接起来的网络结构基础架构。虽然可以用 16 Gb 光纤通道交换机和控制器实施这种类型的 SAN 基础架构，但为了取得最佳效果，建议使用 32 Gb 端到端光纤通道基础架构。

引言

全新企业要务

NVMe 与固态驱动器

基于网络结构的 NVMe

数据库、SAN 和 NVMe

业务成果

目前，NetApp 解决方案融合了 Brocade 和 Broadcom 等技术合作伙伴的一流网络基础架构，可提供实现最大性能提升所需的完整 NVMe-oF 数据路径。此外，一些拥有最新存储系统和光纤通道硬件总线适配器的客户，将能够通过更新驱动程序和其他软件无缝地升级。

图 6 所示示例配置列出了服务器硬件和软件、网络结构基础架构以及存储系统。

NetApp 与 Brocade 合作编写了一本简明扼要、可免费下载的简要电子书《[基于光纤通道的 NVMe for Dummies](#)》。

有关详细信息，请下载此[技术报告](#)，了解如何设计和实施端到端 NVMe/FC 解决方案。

	ONTAP 9.4 堆栈	来源
服务器		
服务器 OS	SLES 12 SP3	SUSE Linux
服务器端 HBA	Emulex LPe32004 32 G FC HBA	Broadcom
服务器端 FC/NVMe 驱动程序	Brocade 最新驱动程序	Broadcom
	SUSE Linux Enterprise 12 机载驱动程序	SLES
网络结构基础架构		
交换机	第 5 代：Brocade 6505/6510/6520/8500	Broadcom
	第 6 代：Brocade G610/G620/G630/X6	Broadcom
	Cisco MDS 9132T	Cisco
交换机固件级别	固件 8.1.1–8.2.x (G610/G620)	Broadcom
	固件 8.2.x (G630)	Broadcom
	Broadcom 固件下载	Broadcom
	固件 NX-OS 8.2(1)、8.2.(2)	Cisco
交换机管理软件	Brocade Network Advisor (BNA) 14.4.0	Broadcom
存储和 ONTAP		
控制器	A300、A700、A700s、A800	NetApp
FC/NVMe 协议许可证	ONTAP 9.4	NetApp

图 6：您目前可以实施的完整 NVMe/FC 解决方案堆栈。

数据库、SAN 和 NVMe

正是当下对企业至关重要的工作负载造就了 NVMe 越来越高的受欢迎度。这些工作负载范围广泛，从早期建立的工作效率标准到尖端新应用程序均在其中。

在这些业务关键型工作负载中，许多只在 SAN 系统上运行或优先运行：

- SAN 基于目前速度最快的协议（即，光纤通道和 iSCSI）构建。
- SAN 是许多数据库工作负载的最佳选择，而这些工作负载也是关键的业务需求。
- SAN 属于块级系统，它像外部硬盘驱动器一样与主机相连，这是许多环境中的一个重要架构优势。

事实上，大多数业务关键型应用程序均基于数据库构建，特别是 Oracle 数据库、Microsoft SQL Server 和 SAP HANA。

SAN 非常适合数据库工作负载，因为它们的设计标准超越了企业最严格的关键性能指标，而这些指标通常围绕以下因素设置：

- 性能
- 可用性
- 易于使用
- 适应未来需求
- 投资回报率

企业数据库工作负载

为了优化性能，数据库管理员会持续监控系统，以确保充分利用每个硬件组件。这对于数据库应用程序尤其重要，因为服务器无序增长或 I/O 瓶颈会大幅降低响应时间。

引言

全新企业要务

NVMe 与固态驱动器

基于网络结构的 NVMe

数据库、SAN 和 NVMe

业务成果

例如，当 I/O 请求必须等待处理时，连接到基于 SATA 的存储系统的单个数据库服务器必然会遇到性能问题。为了解决此问题，IT 架构师有时会启用另一台服务器，在两台服务器之间拆分工作负载，或者将一台服务器专用于用户交互，另一台服务器用于后端处理。但是这种替代方案成本高昂，一个企业数据库许可证每年就要花费数万美元，更不用说运营费用了。再加上电耗和散热成本，NVMe 显然能让您以更少的硬件和更低的成本完成更多任务。

内存数据库

NVMe 非常适合 SAP HANA、Apache Spark 以及依靠主内存进行数据存储的其他内存数据库应用程序。这些工作负载会使用海量数据集运行，其大小有可能超过组合的集群内存，这会给存储系统造成巨大压力。

在 SATA 环境中，即便使用 RAID 协议，也会存在硬件发生故障的重大风险，导致节点临时降级甚至停止运行，令情况雪上加霜。然而，如果采用 NVMe，即使是使用当前速度惊人的多核处理器，总线速度也足以跟上 CPU 的速度。

业务成果

技术研发不断发展，永无止步。因此，随着 NVMe 及其他颠覆性技术创新进入市场，IT 主管面临的一大挑战是如何面向未来做出合理的投资。

对于已经踏上数字化转型之旅的企业来说，关键要务包括：

- 利用闪存和混合云迁移到新兴的现代数据中心架构
- 通过高级分析和人工智能挖掘数据的战略价值
- 通过在高速的现代化 SAN 架构上运行核心工作负载来满足它们在未来的需求

现在，NetApp 拥有相关的技术、产品和专业知识，可为全球企业提供无与伦比的 SAN 价值主张。从某种程度上说，这个价值主张是指产品，具有闪电般速度的存储阵列和数据网络结构提供了企业所需的持续可用性。还有生态系统，即系统集成商和增值经销商，他们围绕这些产品构建完整的解决方案，根据其客户的行业特定需求进行配置。

除此之外，还包括客户青睐的平台：NetApp ONTAP® 可为现代数据中心提供一流的数据管理工具，为核心企业工作负载提供自定义配置以及独有的 NetApp Active IQ® 自动支持系统。

几十年来，NetApp 作为基于文件的数据存储领域的主导品牌，过去几年也在 SAN 技术领域占据了领先地位，AFF SAN 系统的销量同比增长一倍以上。事实上，大多数 NetApp AFF 客户都在实施 SAN。市场研究机构 IDC 最近的报告证实，NetApp 以强劲的势头迈入 NVMe 时代，目前是五大 SAN 供应商中增长速度最快的。

了解更多信息

netapp.com/cn/products/storage-systems/storage-area-network.aspx

*来源：2018年3月出版的IDC《WW Quarterly Enterprise Storage Systems Tracker》2017年第4季度季刊。

© 2018 NetApp, Inc.。保留所有权利。NetApp、NetApp 标识和 [netapp.com/TM](http://netapp.com) 上所列的商标是 NetApp, Inc. 的商标。其他公司和产品名称可能是其各自所有者的商标。2018年6月

