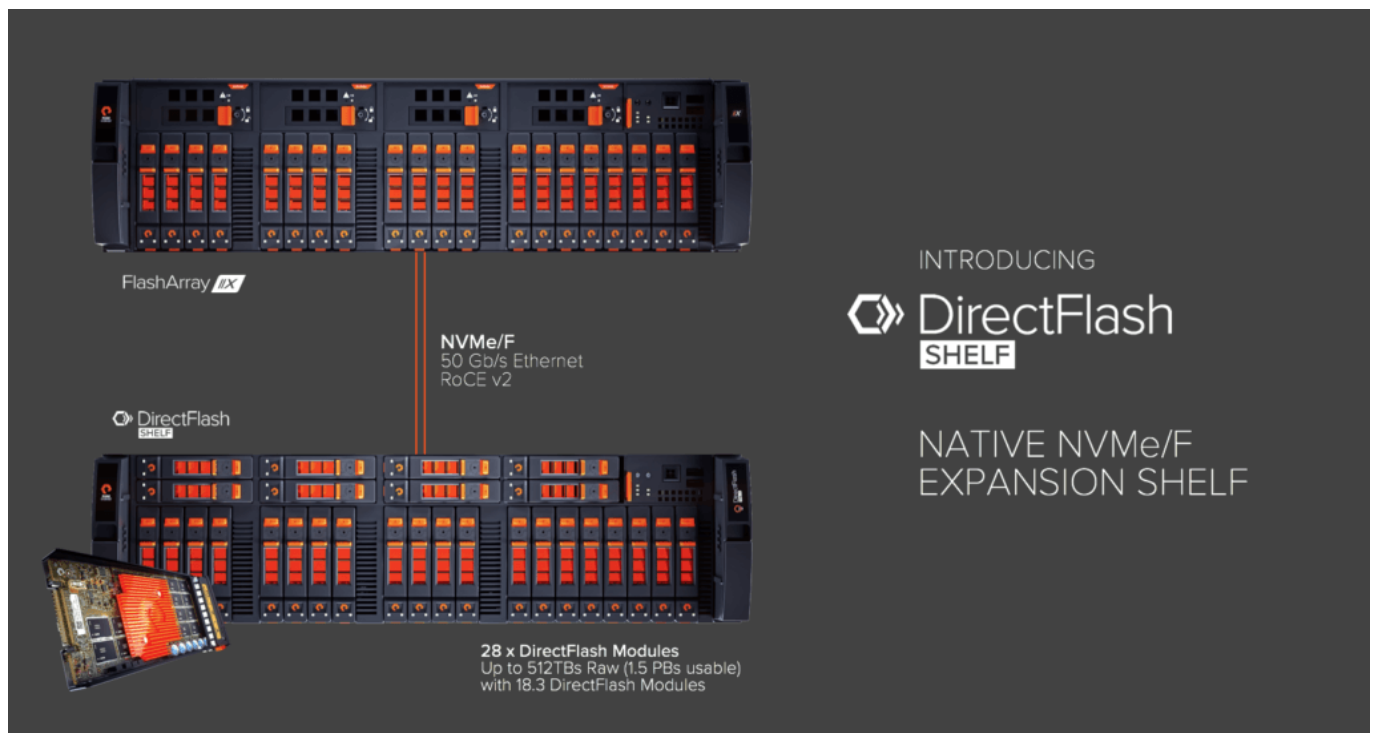


# Extending NVMe Leadership: DirectFlash Shelf and Previewing NVMe/F



Last month we launched the industry's first mainstream 100% NVMe all-flash array, FlashArray//X. Excitement has been high, both for our leadership in making NVMe affordable, as well as the unique software-defined architecture of our DirectFlash™ Modules. DirectFlash Modules bring new levels of performance density by exposing the full parallelism of flash to software through the NVMe interface. But at launch, customers naturally had questions about how we will extend our NVMe technology outside the //X chassis — today we answer those questions.



IS JUST THE **BEGINNING** OF A JOURNEY



FlashArray//X



DirectFlash  
MODULE



DirectFlash  
SOFTWARE

NVMe  
EXPANSION?

NVMe SERVER  
FABRIC?

This blog is part of our 11-part launch series, **Cloud-Era Flash: “The Year of Software” Launch Blog Series**. Click-around to all the blogs to get deep dives on all the new features!

### Cloud-Era Flash: “The Year of Software” Launch Blog Series

1. The Biggest Software Launch in Pure’s History
2. Purity ActiveCluster – Simple Stretch Clustering for All
3. FlashBlade – Now 5X Bigger, 5X Faster
4. Fast Object Storage on FlashBlade
5. Purity CloudSnap – Delivering Native Public Cloud Integration for Purity
6. Simplifying VMware VVols – Storage Designed for Your Cloud
7. Purity Run – Opening FlashArray to run YOUR VMs & Containers
8. Extending our NVMe Leadership: Introducing **DirectFlash Shelf** and Previewing NVMe/F (**THIS BLOG**)
9. Windows File Services for FlashArray: Adding Full SMB & NFS to FlashArray
10. Introducing Pure1 META: Pure’s AI Platform to enable Self-Driving Storage

Adding capacity to FlashArray systems by connecting expansion shelves allows systems to start small and grow as needed. Expansion shelves also allow systems to scale efficiently to high capacities without the cost of unneeded controllers and without requiring exotic ultra-dense flash modules. But our existing expansion shelves use SAS. How can we bring the advantages of expansion shelves to the all-NVMe FlashArray//X, without giving up the advantages of NVMe and DirectFlash modules?

How can we build an expansion shelf for DirectFlash modules? Clearly, we can’t use our existing SAS shelves – even if we could translate between NVMe and SAS, we wouldn’t want to give up on native NVMe. So we know we need to build new hardware – a new shelf that holds DirectFlash modules and talks native NVMe.

One simple way to build this shelf would be to use PCIe switches, cables and retimer cards to extend the FlashArray//X PCIe fabric into the shelf. However, our experience is that while PCIe is efficient and well-suited to connecting components within a chassis, PCIe is not ideal for linking multiple chassis in a fabric. It certainly can be made to work, but events like cable hot-plugs push PCIe error handling too far – like any other exotic technology, multi-chassis PCIe systems are fragile and high-maintenance.

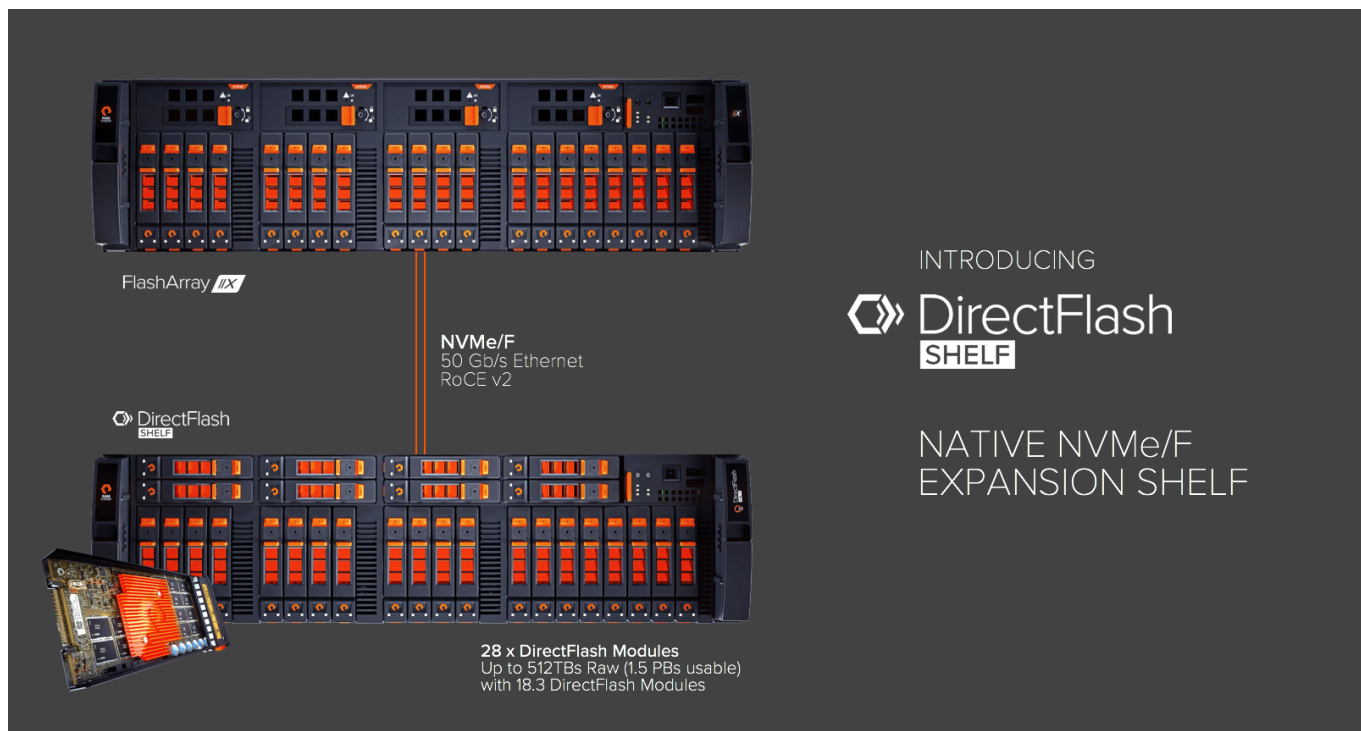
## Enter NVMe over Fabrics

Fortunately, there is a standard perfectly suited for our needs: NVMe over Fabrics. NVMe over Fabrics is recently-released extension of the original NVMe standard, released by the same standards body. NVMe over Fabrics keeps the same lightweight NVMe command set and highly parallel queueing model, but generalizes the low-level transport to support not just PCIe but also networked fabrics – Ethernet or Fibre Channel.

We already have networks that are great at connecting multiple boxes together in a fabric. NVMe is a great way to talk to highly parallel flash. NVMe over Fabrics leaves the network communication to existing networking standards, and uses NVMe for talking to flash, and combines them in a straightforward way – the [whole standard](#) is less than 50 pages long.



NVMe over Fabrics is the basis for our new DirectFlash Shelf. Exactly the same DirectFlash modules used in the FlashArray//X chassis are used in the DirectFlash shelf. The shelf connects to FlashArray//X via RoCE running over 50 Gb/sec Ethernet. Our new DirectFlash Shelf Controller translates between NVMe over Fabrics and NVMe (over PCIe). The translation is simple, because the commands and responses stay NVMe throughout. The shelf controller offloads data movement to RoCE hardware. Our software is highly efficient – built with an advanced architecture that is lockless, multicore, and runs in polling mode with no context switches.

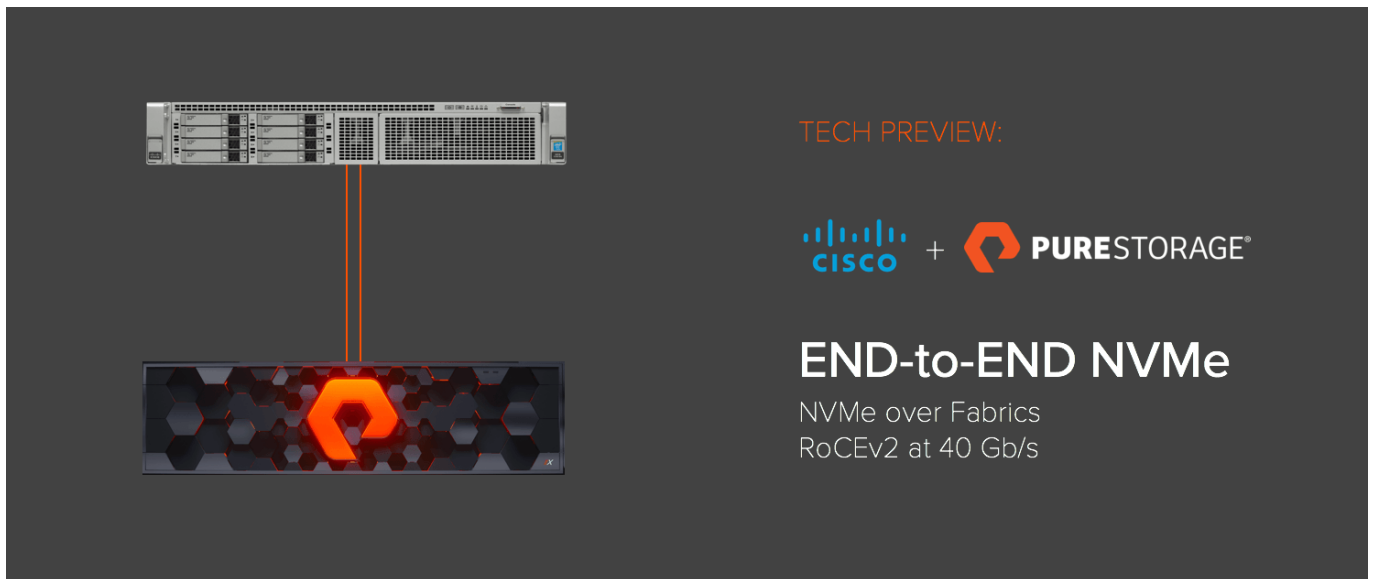


The result is that accessing DirectFlash Modules in the DirectFlash Shelf adds single-digit microseconds of incremental latency, even when running at hundreds of thousands of IOPS and multiple GB/sec of throughput. I can't help but want to talk about the cool technology inside the DirectFlash Shelf – but key is that we achieved a very simple aim: we can expand FlashArray//X without compromising performance, efficiency or resiliency.

## The Future

When we explain the advantages of NVMe and talk about how it can be extended over the network with NVMe over Fabrics, there's an obvious follow-up question: "How can we move away from SCSI and start having our servers use NVMe over Fabrics to talk to our all-flash arrays?"

Finally, I can give our answer to that question. As I described earlier, inside the DirectFlash Shelf is fast and scalable software that implements the target side of it over Fabrics. At Pure//Accelerate, we are previewing a version of Purity software running on FlashArray//X that incorporates the same implementation of it over Fabrics. Our implementation of it over Fabrics follows the official 1.0 standard so servers can use standard adapter cards and in-box host drivers.



We're demonstrating a server with a standard RoCE adapter mounting an NVMe over Fabrics device from the FlashArray. The server is using an unmodified Linux kernel (NVMe over Fabrics host drivers have been available in the upstream Linux kernel since version 4.8, released in October 2016) and the open source [NVMe management package](#). The NVMe device appears on the server as a normal `"/dev/nvme0n1"` block device. On the array side, it is a Purity volume, so the data is deduplicated and compressed, protected with RAID-HA, and can be snapshotted and replicated as usual.

NVMe over Fabrics is exciting in particular for cloud provider and SaaS architectures, as it enables ultra-dense configurations by using top-of-rack flash, instead of PCIe flash or SSDs inside each server. Can you imagine your business running on a rack like this?

## NVMe Summary

We expect using NVMe everywhere will improve performance and reduce CPU cycles wasted on locking and SCSI stacks. We can combine the best of both local flash and shared storage in future architectures. One of the rewards for the hard work of putting together this NVMe over Fabrics demo is the fun and excitement of finally sharing it with the world. I hope to share this in person with as many of you as I can at Pure//Accelerate - see you there!

Note that FlashArray//X and DirectFlash Modules are already available. We plan on shipping DirectFlash shelf in Q4 of this year. We have not yet announced a timeline for shipping front-end NVMe over Fabrics yet — stay tuned. If you are interested in being an early adopter beta customer for NVMe over Fabrics, reach out to us through your sales team.