

BUSINESS WHITE PAPER

The Pure//E Family of All-flash Storage

Why It Spells the End of HDD in the Enterprise

August 2023

Introduction

In 2012, enterprise storage vendor Pure Storage® shipped the first enterprise-class all-flash array (AFAs), called the FlashArray™. By 2016, all the other enterprise storage vendors had introduced their own AFAs, and over the last decade AFAs have come to dominate storage shipments for high-performance and high-availability workloads. Compared to hard disk drive (HDD)-based storage systems, AFAs delivered better and more consistent performance, higher availability, better reliability, enabled the use of critical new technologies like compression and deduplication for latency-sensitive workloads, were far easier to manage than HDD-based systems, and lowered both energy and floorspace consumption. As a result, the market for “high-performance” 10K and 15K RPM HDDs has almost disappeared.

Today, nearline HDDs are still in broad use with more capacity- and cost-conscious workloads, representing roughly 90% of overall worldwide capacity. Because of the lower cost of raw storage capacity with HDDs, many enterprises mistakenly perceive systems based around them as more cost-effective for capacity- and cost-conscious workloads. These workloads include use cases like data protection, artificial intelligence (AI) and machine learning (ML)-driven data analytics, log analytics, content repositories, electronic design automation (EDA), and high performance computing across verticals like financial services, video surveillance, media and entertainment, healthcare, and others. Even the staunchest proponents of HDDs would admit, however, that if flash cost the same as disk, the market for HDDs would disappear almost overnight. With innovations and efficiencies introduced by Pure Storage, the industry is now at that inflection point.

The Right Metric for Cost-effective Storage Infrastructure: Cost per Effective Gigabyte

When enterprises size a storage system, they typically focus on purchasing a desired level of performance at a certain amount of target data storage capacity. Since nearly the beginning of magnetic spinning disk media, HDD vendors have represented the cost of their products using a raw capacity metric. The problem with that approach is that raw capacity does not determine how much data can be stored—it is effective capacity which determines that. And due to a variety of variables—formatting, on-disk data protection and capacity utilization, compression, deduplication, thin provisioning, and performance at the storage device level—effective capacity can be anywhere from 30% to 70% of raw capacity with HDDs. To get to the desired storage capacity to store a target amount of data, a lot more raw capacity has to be purchased. What enterprises should look at is how different media technologies and storage systems compare in terms of the cost of effective, not raw, capacity.

With our Pure//E™ family of enterprise storage systems, Pure Storage can meet the requirements of all HDD-based enterprise block, file, or object workloads with all-flash platforms, including storage-as-a-service. And we can do that with an acquisition cost that is at or below that of HDD-based arrays and ongoing costs (energy, floor space, management, maintenance, technology refresh, etc.) that are 60–80% lower. If customers purchase through our storage-as-a-service offering Evergreen//One™, there can be even greater savings in storage and environmental costs. This white paper will explain how we can do that, based on certain product strategies and innovations that are unique to Pure Storage.



Comparing Enterprise HDD Storage to Pure Storage Flash

Let's face it: the only reason customers still buy HDDs is because of their low cost of raw capacity. In every other way, flash from Pure Storage far outpaces HDDs for cost- and capacity-conscious workloads. It has higher performance, which means you need a lot fewer storage devices to meet performance requirements. It has much higher density, which also means you need a lot fewer storage devices and less floor space. Together, these two characteristics of flash from Pure Storage mean that you need up to 75% fewer storage devices¹ to meet a given performance and capacity requirement. This reduction in storage device count also affects how much supporting infrastructure—controllers, enclosures, fans, power supplies, cabling, and even storage administrators—is required for a storage system, additionally impacting both acquisition cost and energy and floor space consumption.

In addition, flash systems from Pure Storage has higher reliability—up to 20x better than HDDs—which means you spend less time replacing and rebuilding failed devices and enjoy a longer media life cycle. It consumes far less energy—commonly 80-85% less, driving significantly lower operational costs. And it can be non-disruptively refreshed across media generations to improve performance and density over time without having to perform forklift upgrades.

Simply put, HDD-based arrays require too much space, too much energy, and too many management resources. With data growing at 30% to 40% per year for many enterprises and looming power grid and data center floor space limitations², enterprises can't just keep adding HDDs to expand their storage capacity. They have to move to a more efficient kind of enterprise storage infrastructure.

Not Just Any Flash, DirectFlash® Architecture

Instead of using commodity off-the-shelf (COTS) storage devices like our competitors, we are using a very different product strategy. We build our own storage devices, called DirectFlash Modules (DFMs), that are far denser already than HDDs, and manage the flash media both globally and directly using our Purity operating environment. We ship 48TB DFMs today, will be shipping 75TB DFMs later this year, and will introduce 150TB DFMs by the end of 2024. Our ability to manage the flash media globally and directly allows us to achieve capacity utilization in the 95%+ range, which enables us to convert a much higher percentage of raw capacity to effective capacity, reducing cost, energy, and floor space consumption. Because of performance limitations with spinning disk, most HDD vendors suggest that enterprises not fill their devices more than 60-80% full, significantly lowering effective capacity.

The Capacity Utilization Dilemma with HDDs

While larger capacity HDDs do exist, few enterprises use them. There are several reasons for this. The number of IOPS per device does not increase as the device gets larger, which means that the IOPS/TB ratio decreases with larger-capacity devices. This tends to limit how much of the larger devices' capacity can effectively be used. Rebuild times also increase with larger devices, and storage managers already concerned with data risk during rebuilds are hesitant to increase that exposure by moving to larger devices. For these reasons, most enterprises do not widely use the highest capacity COTS storage devices. They tend to use smaller devices that enable higher capacity utilization (i.e. less wasted space within each device) and shorter rebuild times. This use of smaller capacity devices has an impact, though, on both energy and floor space consumption (and ultimately cost) since it means that more devices must be purchased to hit a given performance and effective capacity target.

The impact of storage density and capacity utilization is clear if we look at what this means for the number of devices that must be purchased. If an enterprise is using 12TB HDDs to achieve their desired blend of performance and capacity to build a 3PB (3072TB) system, they will need 256 devices. Pure Storage FlashArray//E™ systems will need only 41 75TB DFMs (which by the way can fit into a 6U system). And that is just a comparison based on raw storage capacity. If we calculate the comparison using effective capacity, the number of HDDs required will go up much more than the number of DFMs will.



While it's true that a single DFM costs more than a single HDD, the fact that for most capacity- and cost-conscious workloads enterprises need to buy three to six times more HDDs quickly erases any cost advantage. And our cost, energy, and floor space advantages in this area will only increase as we release denser media. But it is not just that we produce much larger storage devices—it is that we combine those larger devices with our innovative Purity software that allows extremely high-capacity utilization and much higher parallelism for faster rebuilds. It is that differentiated product strategy—the DirectFlash architecture—and its execution that vendors using COTS storage devices will not be able to copy.

Many of the capacity- and cost-conscious workloads running on HDDs today use 7200 RPM nearline SAS HDDs that in volume typically cost about \$0.015/GB for raw capacity, using a traditional purchase motion. Given the capacity overhead inherent in HDD-based systems, along with all the additional infrastructure that has to be purchased to handle the high device count, the acquisition cost for effective capacity is in the \$0.30/GB range for multi-petabyte (PB) scale systems. This includes three years of 24/7 maintenance. To continue the traditional purchase comparison, Pure Storage has priced their Pure//E family of storage systems at \$0.20/GB for raw capacity, including three years of 24/7 maintenance, either using the Evergreen//Forever™ up-front purchase and term-based subscription, or the Evergreen//Flex™ model where the up-front purchase price is lower and followed by a usage-based subscription. This translates to an effective cost of capacity of \$0.25-\$0.30/GB, depending on the size of the configuration. However, with Pure Storage's Storage-as-a-Service offering Evergreen//One, this cost can be reduced even further. We compare very favorably to all-HDD systems at these initial acquisition price points, but when you take into account the 60-80% lower operating costs of Pure Storage systems (proven out over more than a decade of enterprise usage)³, HDDs do not compare favorably at all. It is this reality that drives our claim that the demise of HDDs is inevitable. As more enterprises understand this reality, HDDs will be rapidly replaced in enterprise data centers.

Note that our claim is specific to flash storage the way we implement and manage it. COTS SSDs are better than HDDs in many of the same ways that we are, but because they must emulate HDDs, they are far less efficient. We are two to five times more efficient than COTS SSDs, but up to ten times more efficient than HDDs.⁴

The All-flash Pure//E Product Portfolio

We offer two different system types—scale up and scale out—that leverage our DirectFlash architecture. We use these models of arrays to deliver a unified data repository (UDR) tier of service in the Evergreen//One Storage-as-a-Service offering, including capacity, performance, availability and energy efficiency service level agreements (SLAs). This latter approach yields the highest TCO savings of any of the consumption models, giving customers a mechanism to pay only for what they use.

The FlashArray//E system is a scale-up system that supports both block- and file-based access and is targeted at workloads from 1PB to 4PBs in raw capacity. The FlashBlade//E™ system is a scale-out system that supports both file- and object-based access and is targeted for workloads from 4PB to 10PB+ in raw capacity. The systems leverage common hardware components (like our DFMs), run our Purity operating environment, and feature Pure1®, our AI/ML-driven, cloud-based monitoring and management platform. Our approach is unique in the industry in that we use a common storage operating environment (Purity) across our entire product line that enables much easier and more familiar management, standing in stark contrast to our competitors who offer different storage silos with different operating environments for different workloads.

The scale-up FlashArray//E starts with 1PB of raw storage capacity in 3U and can be expanded to up to 4PB in a 6U configuration using the 75TB DFMs. It supports both block (Fibre Channel, iSCSI) and file (NFS, SMB) access, offers compression and deduplication, and delivers its highly efficient capacity at under 1 watt per TB.



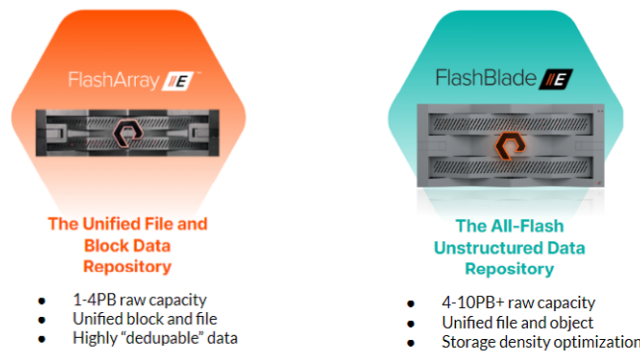


FIGURE 1 The Pure//E family can meet all enterprise data storage needs, regardless of access method, at the cost of HDD-based systems.

The scale-out FlashBlade//E system starts with 2PB of raw storage capacity in 5U and can be expanded to up to 8PB in an 11U configuration using the 48TB DFMs. Once the 75TB DFMs start shipping, FlashBlade//E will support those as well as increasing the number of expansion cabinets, extending its capacity to 30PB raw. It supports both file (NFS, SMB) and object (S3) access, offers in-line deep compression, and delivers its highly efficient capacity at under 1 watt per TB, too.

Within the consumption model options, the scale-up starts as low as 300 TiB with Evergreen//Flex™ (combined with an up-front traditional purchase at deep discounts) and 500 TiB with Evergreen//One (with no other up-front purchase or any other cost line items). Unit rates are as low as \$.016 per GiB per month, comparable to public cloud providers.

The Pure//E systems use quad-level cell (QLC) NAND flash-based storage devices and a new DFM design (the DFMe) that optimizes for high storage density and low cost. The DFMe first becomes available in the 75TB capacity and it is the only DFM shipping on the FlashArray//E. The FlashBlade//E system, which was introduced earlier this year, supports the smaller 48TB DFM but will support the 75TB DFMe as soon as it is available. The DFMe is eight times denser than typical HDDs deployed for the target workloads, uses one-third the watts per TB of HDDs, and exhibits up to a twenty times lower failure rate than HDDs (and twice the useful life). Customers can trade in the smaller DFMs with full trade-in credit to consolidate capacity onto larger DFMs on-demand, paying only for the increase in capacity.

The Pure//E family of systems are more cost-efficient, more reliable, and more sustainable than HDD-based alternatives. The higher density of DFMs translates into revolutionary energy and floor space savings with operating costs 60-80% lower than HDD-based arrays configured to meet the same performance and capacity requirements. Their greater reliability supports high availability and ensures that administrators spend far less time replacing failed devices. Their significantly higher throughput, along with Purity's optimized data rebuild algorithms, mean that administrators also spend far less time rebuilding data as well. The infrastructure efficiency of the DirectFlash architecture and Pure//E systems support a more sustainable enterprise storage option that requires far fewer drives and less supporting infrastructure for a lower manufacturing impact with 85% less e-waste, 80% less energy usage, and up to 95% less floor space consumption.

Pure Storage's Evergreen Architecture Delivers IT Agility

All Pure Storage systems feature our Evergreen Architecture, which enables non-disruptive, in-place upgrades across technology generations that insures against technical debt and avoids disruptive refreshes every five years. The Evergreen//Forever™ and Evergreen//Flex subscriptions give our customers who have purchased Pure Storage arrays a subscription to innovation that



means their systems will always reflect the latest in storage technology without ever having to perform disruptive upgrades. We have many customers that purchased our original arrays back in 2012 whose systems today include the latest in controller, backplane, media and software technology, and over 97% of the arrays we have ever sold are still in production use today.

An interesting implication of the Evergreen architecture is that it more than doubles the useful life of a storage system without incurring any performance, availability, reliability, or efficiency risk. In addition to all of the other operating cost savings of the Pure//E family of systems, the “ten year-plus” life cycle that our customers enjoy provides additional advantages. As application growth drives performance and capacity requirements, most HDD-based systems need to be upgraded every three to five years. These “forklift” upgrades require that customers re-buy all of their capacity, relicense the software, reprovision the storage, and migrate the data and applications, imposing considerable cost, risk, and manual labor commitments. Pure Storage customers avoid that while staying with a familiar storage operating environment.

All of the Pure//E family of systems can be purchased outright in the traditional way, or they can be deployed using Evergreen//One, our storage-as-a-service offering where Pure Storage owns the assets and the customers pay for usage at an all-inclusive unit rate. Evergreen//One is backed by a number of SLAs that are unmatched in the industry, such as at least five-nines of availability, 25% buffer capacity (relative to usage, not the asset), performance (measured in IOPs and GiB/sec), and energy efficiency (measured in watts per TiB). Pure Storage also recently announced an optional ransomware recovery add-on, where we will ship clean arrays next business day to a customer whose infrastructure has been embargoed following a ransomware attack and provide onsite engineering support to assist with restoring business data to those clean arrays, not just a temporary fix.

All of this gives organizations an excellent option for dealing with uncertainty and mitigating risk, and control over an on-premises deployment of non-disruptively scaling and evolving infrastructure with pay-as-you-go cloud economics.

For customers making traditional purchases of hardware, Evergreen//Forever is an aftercare subscription powered by Evergreen Architecture that delivers seamless, rapid software and hardware upgrades and expansions, without disruption. These upgrades are included for all Evergreen//One and Evergreen//Flex customers.

Evergreen//Flex is similar to Evergreen//Forever but with two major differences. The cost of the up-front hardware purchase is drastically reduced, and the subscription charges are based on the utilization of the hardware, providing customers the ability to own their storage arrays while still taking advantage of consumption economics based on asset utilization.

Learn More

If you are already power- or floor space-constrained or just looking to move to more efficient storage infrastructure as you implement technology refreshes on your HDD-based arrays, we'd like to meet to discuss the business value we can deliver. As the enterprise storage sustainability leader, we are better positioned than any other storage provider to maximize energy and data center floor space resources to meet your ownership cost and sustainability goals over the long term while providing the best-in-class enterprise storage capabilities that will set you apart from your competition.

1 Assumes competitive systems are configured with 18TB HDDs and Pure Storage systems are configured with 75TB DFMs. A Pure Storage system requiring 1024TB of raw capacity would require 75% fewer storage devices (14 DFMs vs 57 HDDs).

2 Evaluating Power and Cost Efficiency Considerations in Enterprise Storage Infrastructure (IDC #US49354722, July 2022) and IDC US Datacenter Operational Survey, December 2021

3 Based on internal competitive analyses of Pure Storage customers over the last five years

4 Efficient IT Infrastructure Saves More Than Just Costs, June 2023, Pure Storage

purestorage.com

800.379.PURE

