Technical White Paper

Everpure

# FlashBlade//EXA SPECstorage Solution 2020_ai_image

Benchmark test results

# Contents

## Introduction

AI has matured from experimental research to a mission-critical foundation of enterprise production systems. As this transformation accelerates, organizations face increasing performance, scalability, and cost pressures across their AI infrastructure. Despite significant investment in GPUs and high-performance computing (HPC) clusters, many enterprises find their accelerators underutilized during training and inference cycles. The culprit is rarely a lack of compute power—it is the inability of legacy storage systems to deliver data fast enough. Bottlenecks in data throughput and I/O latency leave GPUs idle, resulting in wasted capacity, delayed outcomes, and a rising total cost of ownership.

Everpure™ FlashBlade//EXA™ is a scale-out, disaggregated file and object platform specifically engineered for very large-scale AI, HPC, and unstructured data pipelines, combining extremely high throughput with exabyte-class scalability.

## Executive summary

AI infrastructure has entered a new performance era where storage—not GPUs—defines the pace of innovation, efficiency, and return on investment. As ML models and data sets continue to expand across petabyte-scale environments, the ability to deliver continuous, high-throughput data to accelerators is now the primary factor driving training and inference time, system utilization, and total cost of ownership.

The SPECstorage Solution 2020_ai_image test within the SPEC-FS AI benchmark suite provides a rigorous, industry-standard framework for evaluating how well storage systems sustain real-world AI workloads. Designed to emulate the I/O demands of large-scale image classification and vision-based deep learning workflows, this benchmark measures key performance dimensions such as aggregate throughput, metadata efficiency, and latency under concurrent access conditions typical of enterprise and research pipelines. Its results expose a crucial operational insight: storage I/O behavior—not GPU count—dictates whether AI deployments reach full performance potential.

Everpure FlashBlade//EXA stands out as the leader in this new performance frontier. It consistently exceeds published SPECstorage Solution 2020_ai_image results, delivering superior concurrency, low latency, and sustained throughput to keep GPU utilization high. This directly translates into achieving equivalent or greater AI performance with fewer GPUs and servers—driving down capital costs, reducing energy consumption, and simplifying data center operations.

By standardizing on FlashBlade//EXA, organizations gain more than raw speed—they secure a scalable, predictable, and cost-efficient foundation purpose-built for enterprise AI. This enables faster deployment of AI initiatives, stronger ROI from infrastructure investments, and the agility to adapt and innovate as business demands evolve.

## Understanding the SPECstorage Solution 2020_ai_image benchmark

### Overview

The SPECstorage Solution 2020_ai_image benchmark, part of the SPEC-FS AI suite, is an industry-standard performance evaluation designed to measure how storage systems handle AI and ML workloads dominated by image data. It represents a major evolution from synthetic performance testing toward workload-realistic benchmarking, capturing how actual AI pipelines interact with underlying file systems during training and inference processes.

### Purpose

Modern AI workflows depend on the continuous delivery of large volumes of image data to GPUs. When storage I/O cannot keep pace, compute resources stall and operational costs rise. The SPECstorage Solution 2020_ai_image benchmark was developed to provide an objective, reproducible standard for evaluating a system's ability to sustain these demanding I/O patterns. Its results give IT leaders, system architects, and storage vendors insight into how efficiently their infrastructure supports real AI workloads.

### What the benchmark measures

- **Throughput and bandwidth**: quantifies how much image data can be delivered to compute nodes per second under sustained load
- **Latency and responsiveness**: evaluates how quickly the storage system responds to data requests, particularly during parallel GPU access
- **Concurrency and scaling**: tests performance under multiple simultaneous read/write operations representative of distributed training scenarios
- **Metadata operations**: measures how efficiently the file system handles metadata, which is critical in high-file-count image data sets

### Why it matters

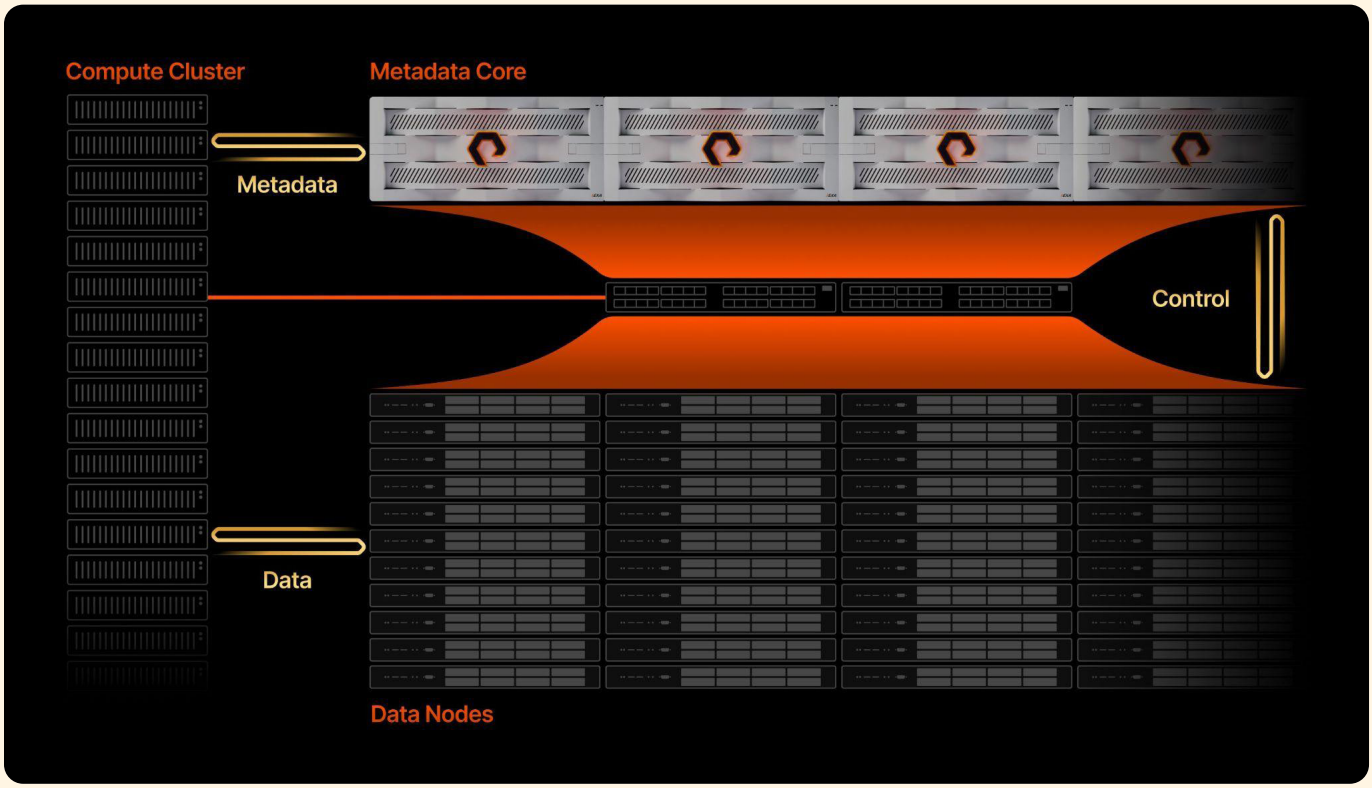
The 2020_ai_image benchmark closely mirrors the operational demands of real-world applications such as computer vision, image classification, and medical imaging analysis. By modeling these workloads, the benchmark exposes the true performance limits of storage systems—far beyond what synthetic metrics like IOPS or raw throughput reveal. High performance in this test directly correlates with increased GPU utilization, faster training and inference cycles, and improved total cost of ownership for AI infrastructure.

### Interpreting results

Strong results on the SPECstorage Solution 2020_ai_image benchmark indicate that a storage platform can sustain the data delivery required for production-grade AI without introducing I/O bottlenecks. Systems optimized for parallelism, low latency, and intelligent caching—such as Everpure FlashBlade//EXA—demonstrate their value by maintaining the highest GPU efficiency, reducing the need for excessive compute scaling and accelerating time to insight across AI initiatives.

## Everpure scalable architecture advantages

Everpure FlashBlade//EXA is a scale-out, disaggregated file and object platform specifically engineered for very large-scale AI, HPC, and unstructured data pipelines, combining extremely high throughput with exabyte-class scalability. Architecturally, FlashBlade//EXA separates metadata and data services so they can scale independently, eliminating the overprovisioning and contention that typically occur in monolithic, tightly coupled designs. This separation, combined with the best-in-class capacity of DirectFlash® Modules, allows the system to deliver more than 10TB/s of read bandwidth in a single namespace, while also optimizing power and cooling at high densities.




## SPECstorage Solution 2020_ai_image details

SPECstorage Solution 2020_ai_image is designed to emulate a complete, production-grade AI image processing pipeline built on TensorFlow, so the testing focuses on how an end-to-end storage solution behaves under that workload.

### What SPECstorage 2020_ai_image is

SPECstorage 2020_ai_image is one of several workloads in the SPECstorage Solution 2020 benchmark suite, alongside genomics, video data acquisition, software build, and electronic design automation (EDA). The 2020_ai_image profile specifically models TensorFlow image processing using traces captured from real systems running COCO, ResNet-50, and Cityscapes data sets, so it represents a realistic mixed-I/O pattern rather than synthetic reads/writes.

### What is actually tested

The benchmark measures how a full storage solution (hardware, software, network, and configuration) supports concurrent AI image jobs at defined service levels.

## Key test dimensions

The SPECstorage 2020_ai_image benchmark evaluates storage systems across several key test dimensions, including:

- **Throughput and concurrency**: The benchmark measures file server throughput and the number of simultaneous "AI jobs" (complete workflows) that can be sustained while meeting strict criteria. Each job is a bundle of subcomponents (for example, data ingest, shuffling, checkpointing, and logging) derived from TensorFlow image training traces.
- **Latency and responsiveness**: The benchmark enforces pass/fail rules on average latency, maximum latency, and balance of operations across workload subcomponents. If latency or imbalance exceeds thresholds, additional jobs cannot be counted as valid, which prevents gaming the benchmark with "best-effort" performance.
- **End-to-end solution behavior**: The benchmark evaluates the entire solution under test (SUT), including clients, storage servers, network, protocols (NFS/SMB), and any data reduction features like deduplication and compression. Results published for vendors describe client counts, virtual machine or server types, network bandwidth, jumbo frame settings, and file system configuration, underscoring that this is solution-level testing.
- **Scalability and distribution**: The benchmark uses a highly scalable load generation framework with up to about four million processes across geographically distributed clients, synchronized at submillisecond resolution. This allows validation of large-scale and cloud-based AI storage architectures (for example, many clients in public cloud instances).
- **Operational analytics and visualization**: The benchmark integrates detailed logging and runtime counters that can be streamed to external databases and visualization tools for deeper diagnostics on bottlenecks and behavior.

## Why this testing matters

The importance can be framed in three business-oriented angles:

- **Real-world AI relevance**: The workload is not a synthetic micro-benchmark; it is based on real TensorFlow image training traces using mainstream data sets (COCO, ResNet-50, and Cityscapes). That makes the results meaningful for organizations running computer vision training or large-scale image analytics because the tested I/O pattern closely mirrors production.
- **Objective, comparable KPI**: The main metric, SPECstorage Solution 2020_ai_image, expresses how many AI image jobs can be run concurrently while staying within strict latency and balance limits. Because the benchmark is standardized and governed by SPEC run and reporting rules, customers can compare different vendors' published results on a level field.
- **Proof of architectural strength**: High 2020_ai_image scores with low latency demonstrate that the storage architecture can handle mixed small and large file I/O, intense metadata activity, and parallel access patterns typical of modern AI pipelines. Vendors achieving strong results often emphasize that they did so with a single, general-purpose configuration across all SPECstorage workloads, highlighting architectural consistency rather than fragile tuning.

## Which AI functions it stresses

Within the AI life cycle, the 2020_ai_image benchmark primarily exercises the storage side of:

- **Data ingestion and preprocessing**: The benchmark exercises rapid streaming of large image data sets into GPU/accelerator hosts, including sequential reads, random access to many small files, and directory/metadata traversals.
- **Training I/O patterns**: The benchmark exercises checkpointing and model snapshot writes, log file updates, and shuffled mini-batch reads typical of TensorFlow image training loops. This captures the I/O characteristics of running popular vision models at scale rather than just static data set scans.
- **Scalable multi-job and multi-tenant AI**: Concurrent execution of many jobs models a shared infrastructure where multiple teams or pipelines train, retrain, and experiment in parallel. The benchmark validates whether the storage layer can maintain consistent performance under that contention, which is critical for MLOps productivity.

## Everpure FlashBlade//EXA SPECstorage 2020_ai_image test results

Everpure FlashBlade//EXA achieved the number one result on the SPECstorage Solution 2020_ai_image benchmark with support for 6,300 concurrent AI_Jobs at 616,129MB/s of throughput, demonstrating extraordinary end-to-end architectural strength and real-world AI readiness. This level of performance shows that the storage solution can reliably feed massive, parallel TensorFlow-based image workloads without becoming a bottleneck, even as thousands of training and inference pipelines run simultaneously. For enterprises scaling computer vision and generative AI initiatives, these results validate that their data infrastructure can keep up with GPU investments, maintain low latency under extreme concurrency, and deliver consistent throughput for faster model development, higher infrastructure utilization, and accelerated time to insight.

### SPECstorage Solution 2020_ai_image

All published results

| Tested By | Solution Name | Results | | | System Configuration | | | Published |
|---|---|---|---|---|---|---|---|---|
| | | AI_Jobs | ORT | MB/s | Workload Name | Memory (GiB) | Total Capacity | |
| DELL Technologies | PowerEdge R6715 | 75 | 0.26 | 7335 | AI_IMAGE | 1536 | 60T | Oct. 10, 2024 |
| HPE and WEKAIO, Inc | HPE WEKA - Physical server Reference | 5000 | 1.06 | 488945 | AI_IMAGE | 24576 | 502.87 TiB | Jan. 28, 2025 |
| Nettrix | R640 G50 Storage Server | 220 | 0.64 | 21520 | AI_IMAGE | 1024 | 25.6 TB | Sep. 4, 2024 |
| New H3C Technologies Co., Ltd. | H3C UniServer R3900 G7 | 75 | 0.36 | 7335 | AI_IMAGE | 768 | 35 TiB | Mar. 18, 2025 |
| New H3C Technologies Co., Ltd. | H3C UniServer R3950 G7 | 75 | 0.24 | 7337 | AI_IMAGE | 768 | 70 TiB | Mar. 18, 2025 |
| New H3C Technologies Co., Ltd. | H3C UniServer R4900G6 Ultra | 50 | 0.47 | 4891 | AI_IMAGE | 1024 | 14 TiB | Feb. 20, 2024 |
| New H3C Technologies Co., Ltd. | H3C UniServer R4900G6 Ultra | 50 | 0.47 | 4890 | AI_IMAGE | 1024 | 11.64 TiB | Mar. 19, 2024 |
| Nutanix | Nutanix Files | 300 | 1.94 | 29339 | AI_IMAGE | 5120 | 43.67 TiB | Aug. 31, 2022 |
| **Pure Storage** | **Pure Storage FlashBlade//EXA** | **6300** | **0.97** | **616129** | **AI_IMAGE** | **20844** | **1843.5 TB Raw, 866.4 TB Usable** | **Feb 17, 2026** |
| Qumulo, Inc. | Azure Native Qumulo - Public Cloud Reference | 704 | 0.84 | 68849 | AI_IMAGE | 8448 | 100 TiB | Jun. 13, 2024 |
| Qumulo, Inc. | Qumulo - Public Cloud Reference | 400 | 1.22 | 39112 | AI_IMAGE | 6144 | 78.54 TB | Dec. 8, 2023 |
| Samsung Datacenter Technology and Cloud Solutions Lab | Samsung PM9A3 NVMe and WekaFS | 1400 | 0.84 | 136899 | AI_IMAGE | 11264 | 188 TiB | Jan 26, 2022 |

| Tested By | Solution Name | Results | | | System Configuration | | | Published |
|---|---|---|---|---|---|---|---|---|
| | | AI_Jobs | ORT | MB/s | Workload Name | Memory (GiB) | Total Capacity | |
| SPECstorage™ Subcommittee | Reference Submission in the Public Cloud | 190 | 1.26 | 18560 | AI_IMAGE | 3840 | 98 TiB | Dec. 15, 2020 |
| SPEC Storage™ Subcommittee | SPEC Storage™ Solution 2020 Reference Solution | 12 | 6.77 | 1130 | AI_IMAGE | 320 | 8 TiB | Dec. 15, 2020 |
| Super Micro Computer | AS-1115CS-TNR | 75 | 0.24 | 7337 | AI_IMAGE | 768 | 35T | Oct. 10, 2024 |
| Super Micro Computer | AS-1115SV-WTNR | 50 | 0.31 | 4891 | AI_IMAGE | 384 | 35T | Sep 19, 2023 |
| Super Micro Computer | SYS-121H-TNR | 100 | 0.48 | 9781 | AI_IMAGE | 1024 | 20.96 TiB | Dec. 15, 2023 |
| Super Micro Computer | SYS-220U-TNR with 22 NVMe Storage Node | 40 | 0.58 | 3912 | AI_IMAGE | 1536 | 35TiB | Apr 21, 2020 |
| Super Micro Computer | SYS-221H-TN24R Hyper Storage Server | 40 | 0.57 | 3912 | AI_IMAGE | 3072 | 83.82TiB | Jan. 10, 2023 |
| WEKAIO, Inc | WEKA - Public Cloud Reference | 700 | 0.85 | 68318 | AI_IMAGE | 8192 | 88 TB | Mar. 13, 2024 |
| WEKAIO, Inc | WEKA - Public Cloud Reference | 2400 | 1.38 | 234711 | AI_IMAGE | 38400 | 1.66 PiB | Mar. 13, 2024 |

See the Appendix for the bill of materials for the full configuration used in the benchmark.

## Conclusion

Everpure FlashBlade//EXA dominated the SPECstorage Solution 2020_ai_image benchmark because its architecture is purpose-built for exactly the kind of massively parallel, mixed-I/O workloads that define modern AI image pipelines. By delivering industry-leading performance in both aggregate throughput and the number of concurrent AI_Jobs supported, FlashBlade//EXA proved it can keep GPUs and AI frameworks continuously fed with data instead of forcing them to idle while waiting on storage. This is the difference between an infrastructure that merely "supports AI" and one that actively accelerates AI outcomes. In the context of SPECstorage 2020_ai_image—a test grounded in real TensorFlow image training traces—FlashBlade//EXA results show that it can sustain production-grade AI at scale, not just win in a lab tuning exercise.

## Unified architecture designed for scale

At the heart of this leadership is the FlashBlade//EXA unified scale-out design, combining fast all-flash media, a massively parallel file system, and high-bandwidth networking into a single, tightly integrated platform. Unlike architectures that bolt together separate performance tiers, filers, and gateways, FlashBlade//EXA presents a simple, high-performance namespace that scales linearly as capacity and blades are added. For AI image workloads, this means consistent latency and predictable throughput as projects, data volumes, and user counts grow. It is precisely this ability to maintain service levels at high concurrency that SPECstorage 2020_ai_image rewards, and it is where FlashBlade//EXA clearly excelled.

Another crucial factor in its dominance is the way FlashBlade//EXA handles the diverse and often punishing I/O patterns of AI. Training and iterating on computer vision models is not just about streaming large image files; it also involves intensive metadata operations, small file access, random reads, checkpoints, and logging. FlashBlade//EXA is engineered to handle both high metadata rates and large-block throughput without forcing customers into complex tuning or workload segregation. As a result, the same platform that ingests and prepares training data can also support active experimentation, retraining, and validation—exactly the kind of multidimensional stress that the 2020_ai_image workload models.

## A platform built for modern AI

Ultimately, Everpure FlashBlade//EXA dominated the SPECstorage 2020_ai_image benchmark because it embodies the three qualities that modern AI infrastructures require: uncompromising performance at scale, architectural simplicity, and real-world workload alignment. The benchmark's AI-focused design rewarded a system that can sustain thousands of image-based workflows concurrently, not just spike in narrow scenarios. For Everpure, these results are more than a technical achievement; they speak directly to what customers value most—faster innovation, better GPU utilization, and confidence that their storage foundation will not hold back their AI ambitions.

## Business impact and real-world value

From a customer value standpoint, FlashBlade//EXA benchmark results translate directly into business outcomes. When a single platform can sustain thousands of AI_Jobs at extreme throughput levels, data science and research teams can run more experiments in parallel, operate more GPUs at higher utilization, and iterate on models faster. This reduces time to accuracy and time to market for AI-driven products and services. It also allows organizations to consolidate silos—analytics clusters, AI training environments, and unstructured data lakes—onto a common, high-performance foundation. Dominating SPECstorage 2020_ai_image is therefore not just a performance trophy; it is tangible proof that FlashBlade//EXA aligns with how AI-first organizations actually work.

Operational simplicity is also a differentiator that the benchmark helps highlight. FlashBlade//EXA is designed to deliver this level of performance with far less complexity than traditional scale-out storage, reducing the number of variables that can jeopardize a benchmark run or, more importantly, a production rollout. Automated data services, straightforward expansion, and a consistent management experience lower the operational burden on infrastructure teams, freeing them to partner with AI teams instead of constantly firefighting performance bottlenecks. SPECstorage Solution 2020_ai_image benchmark leadership gives customers independent evidence that "simple" and "high performance" are not competing promises, but two sides of the same product story.

Strategically, FlashBlade//EXA domination on SPECstorage 2020_ai_image positions Everpure as a trusted infrastructure partner for AI at scale. In a market where AI claims are everywhere, customers look for third-party validation that a platform can deliver under standardized, audited conditions. Leading this benchmark gives Everpure a clear, quantifiable proof point to anchor conversations with CIOs, heads of AI, and platform teams: if FlashBlade//EXA can outperform alternatives on a demanding, AI-specific workload, it can be trusted as the backbone for mission-critical AI initiatives. As organizations move from pilots to production, this kind of measurable, standards-based leadership is often the deciding factor in long-term platform decisions.

**Learn More About FlashBlade//EXA**

## Appendix: Bill of materials

The following bill of materials defines the full configuration used for the SUT in this benchmark.

| Solution Under Test Bill of Materials | | | | | |
|---|---|---|---|---|---|
| **No** | **Qty** | **Type** | **Vendor** | **Model/Name** | **Description** |
| 1 | 1 | Metadata Node system | Pure Storage | FlashBlade //EXA | The Metadata Node system was a 3-chassis multi-chassis configuration with 2 eXternal Fabric Modules (XFMs). Each XFM had 4 x 400Gbps Uplink ports. Each chassis is connected to each XFM with 4x100Gb uplink ports. Each chassis was equipped with 10 x S500R1 FlashBlade//EXA blades. Each blade was equipped with 2 × 37.5TB N58R DirectFlash Modules (DFMs). { Metadata Node system details: [ 2 x eXternal Fabric Modules (XFMs) - Model: XFM-8400 - Part Number: 86-0001-04 ] [ 3 x FlashBlade Chassis - Model: CH-FB-II - Part Number: 83-0383-12 ] [ 10 x FlashBlade S500R1 blades per chassis - Model: FB-S500 - Part Number: 83-0433-08 ] [ 2 x DirectFlash Modules (DFMs) per blade - Raw Capacity: 37.50 TB (34.11 TiB) - Part Number: 83-0489-06 ] [ Pure Storage does not publish publicly accessible specifications for the components of the FlashBlade Metadata Node system. Detailed specifications are available only through customer or partner support documentation. A Technical Deep Dive on FlashBlade//S can be found at: https://www.purestorage.com/video/technical-deep-dive-on-flashblade/6307195175112.html ] } |
| 2 | 30 | Data Node | Supermicro | Supermicro Data Nodes | Supermicro ASG-1115S-NE316R servers {CPU = [single-socket AMD EPYC 9355P processor with 32 physical cores (64 logical CPUs via SMT) on a 64-bit x86 architecture]} {MEMORY = (192 GB of DDR5 ECC memory).} {DATA NETWORK ADAPTERS = (2 x NVIDIA ConnectX-7 EN (MT2910) single-port 400 GbE QSFP112 Ethernet adapters installed, operating over PCIe Gen5 with secure firmware and InfiniBand/VPI functionality disabled.)} {SSDs = [ Each data node has 16 x KIOXIA KCM7DRJE3T84 3.84 TB enterprise NVMe SSDs installed. ] [ Each data node had 61.45 TB of raw capacity and 28.88 TB of usable capacity. ]} {Operating System = [The FlashBlade//EXA Data Node Operating System (Purity//DN) was loaded onto each data node. Security scanning for Purity//DN (the Data Node OS for FlashBlade//EXA) is performed as part of the release process. Purity//DN does not provide mechanisms for non-administrative users to run third-party code, and thus is not affected by common OS vulnerabilities.]} |
| 3 | 60 | Host Initiator | Supermicro | Ubuntu 24.04 Bare-Metal Host Initiators | Supermicro ASG-1115S-NE316R servers {CPU = [single-socket AMD EPYC 9355P processor with 32 physical cores (64 logical CPUs via SMT) on a 64-bit x86 architecture]} {MEMORY = (192 GB of DDR5 ECC memory).} {DATA NETWORK ADAPTERS = (2 x NVIDIA ConnectX-7 EN (MT2910) single-port 400 GbE QSFP112 Ethernet adapters installed, operating over PCIe Gen5 with secure firmware and InfiniBand/VPI functionality disabled.)} {SSD = [1 x Micron 7450-series MTFDKBA480TFR 480 GB enterprise NVMe SSD (NVMe 1.4, PCIe-attached) with full SMART support and 0% media wear, used as a local system disk.]} {Operating System = [Ubuntu 24.04.3 LTS, Kernel Linux 6.14.6clearflag-v1+]} |
| 4 | 20 | Host Initiator | Supermicro | Ubuntu 24.04 Bare-Metal Host Initiators | Supermicro SYS-621C-TN12R servers {CPU = [ dual-socket Intel Xeon Silver 4516Y+ platform with 48 physical cores (96 logical CPUs via SMT) on a 64-bit x86 architecture ]} {MEMORY = (1024 GB of DDR5 ECC memory - reduced to 198752M via GRUB - GRUB_CMDLINE_LINUX_DEFAULT="quiet splash mem=198752M" )} {DATA NETWORK ADAPTERS = (2 x NVIDIA ConnectX-7 EN (MT2910) single-port 400 GbE QSFP112 Ethernet adapters installed, operating over PCIe Gen5 with secure firmware and InfiniBand/VPI functionality disabled.)} {SSD = [1 x Micron 5400-series MTFDDAK240TGA 240 GB enterprise SATA SSD (2.5-inch, SATA 6 Gb/s) with full SMART support and 100% remaining endurance, used as the system disk.]} {Operating System = [Ubuntu 24.04.3 LTS, Kernel Linux 6.14.6clearflag-v1+]} |
| 5 | 8 | Data Network Switch | NVIDIA | NVIDIA SN5600 data network switches | 2 x NVIDIA SN5600 spine data network switches 6 x NVIDIA SN5600 leaf data network switches |



©2026 Everpure, the Everpure P Logo, Pure Storage, DirectFlash, FlashBlade//EXA, and the marks in the Everpure Trademark List are trademarks or registered trademarks of Everpure, Inc. or its licensed subsidiaries in the U.S. and/or other countries. The Trademark List can be found at everpuredata.com/trademarks. Other names may be trademarks of their respective owners.

ID17776-01-en-03/26