ARTIFICIAL INTELLIGENCE
SOLUTION BRIEF

PAVILION

Benefits

- Unrivaled performance, density, scalability, and flexibility
- Ultra-low latency that meets or exceeds that of direct attached NVMe SSDs
- Unlimited scale out capacity to support any size dataset
- Enable AI model training and inference on entire datasets that are otherwise too large to use

Features

- Up to 20 controllers per system
- Simultaneous high performance for block, file, and object data
- Extreme read performance of up to 120GB/s
- Unrivaled write performance of up to 90GB/s
- Linear scale out across systems
- Simultaneous block, file & object protocols

Pavilion HyperParallel Data Platform for AI

Unrivaled High Performance Storage for artificial intelligence, machine learning, deep learning, and big data analytics

As organizations move to a data centric model, one where they are increasingly reliant on processing massive volumes of data to make decisions quickly, the ability to move and analyze data quickly becomes critical. GPU-based systems have revolutionized how data can be processed, enabling AI based initiatives to achieve more.

Critical to these initiatives is the data that the AI models are trained on. The more data they have access to, the better the model. The need for high performance, low latency data access has driven the use of direct attached NVMe SSDs in GPU-based systems. These solutions enable high speed access to the data on the drives.

While this solution has enabled organizations to quickly analyze the local data, the capacity limitations of internal storage have restricted the size of the datasets that can be used for AI model training and inference. Legacy external storage solutions are unable to move the volumes of data required for meaningful analysis in the timeframe required. The result is that organizations are applying their AI initiatives to limited datasets that yield questionable results.

Current Limitations

Retail organizations run queries on months' worth of sales, when they need to compare several years' worth of data. Law enforcement agencies perform facial recognition on thousands of images, when they need to study millions. Financial institutions ingest and analyze data on select companies, when they need to view the whole market. Research organizations are delayed in developing new treatments for diseases. The list goes on.

A new solution is needed. One that resets expectations for what's possible.

Pavilion HyperParallel Data Platform

The Pavilion HyperParallel Data Platform revolutionizes how GPU-based systems can ingest and process data. With a unique switch-based architecture, the Pavilion HyperParallel Data Platform delivers the performance of internal NVMe SSDs with unlimited capacity.

With up to 20 independent controllers per system and an end-to-end NVMe design, organizations can take advantage of unmatched performance, density, scalability, and flexibility.

Unlimited Scale, Unmatched Performance

One of the things that makes AI so powerful is the ability to perform broad analysis across large datasets and to recognize patterns within the data. Once a pattern is identified, the data becomes information, which can then be used to solve a problem, gain a market advantage, or further the goals of the organization. The larger the dataset being analyzed, the more reliable the pattern recognition and the better the outcome.

The challenge is getting enough data to the GPUs running the AI model fast enough for training. AI training can be a continuous process, as more data becomes available. The AI then needs to apply a trained model against unseen data to make a prediction.

AI solutions require a storage system that is big enough to hold all the data that needs to be analyzed, and fast enough to provide that data in the available window. Internal, direct attached NVMe SSDs are fast enough to provide the data to the GPUs running the AI, but they cannot hold enough data. Traditional external storage systems are large enough to hold the data, but too slow to deliver it in the needed time.

The Pavilion HyperParallel Data Platform easily solves the capacity challenge, delivering up to 2PB of usable capacity in a compact 4RU footprint. That capacity scales linearly across any number of systems. All that capacity is then delivered using NVMe-oF and RoCE for high throughput and ultra-low latency. The Pavilion HyperParallel Data Platform provides GPU based systems with unlimited, scalable capacity along with throughput and latency comparable to that of internal NVMe SSDs, giving AI applications both the performance and capacity that they need.

Unrivaled Data Ingest

Data analysis is being performed on larger datasets than ever before and those datasets are continuously growing. That data is being created by a wide range of sources, including IoT data, cameras, sensors, log data, and more. This data is generated across different data types, including block, file, and object. To be useful, any storage system must be able to support high data ingest (write) performance across each data type, simultaneously.

The Pavilion HyperParallel Data Platform delivers universally unmatched data ingest performance with write speeds of up to 90GB/s, which is faster than the read performance of many other solutions.

Uniquely capable of native support for block, file, and object data types across any of 20 independent controllers, the Pavilion HyperParallel Data Platform delivers industry leading performance for each data type, simultaneously.

The alternative is for organizations to use a range of products from different manufacturers to meet the performance requirements of each data type, driving up cost, complexity, and creating needless management challenges.

While most storage solutions are designed to deliver high read performance, Pavilion understands that before the AI application can read the data, it must first be ingested from the source creating it. The Pavilion HyperParallel Data Platform delivers unmatched performance for both read and write operations, across data types.

Best-in-Class Read and Write Performance for Each Data Type

Block		File		Object	
Read Up to 120GB/s Up to 20M IOPS 100µs Latency	Write Up to 90GB/s Up to 5M IOPS 25µs Latency	Read Up to 90GB/s	Write Up to 56GB/s	Read Up to 80GB/s	Write Up to 35GB/s

Ultra-Low Latency

PGPU-based computing has revolutionized the rate at which AI models can be trained by parallelizing processing. To efficiently utilize GPUs requires that data be delivered with the lowest latency possible.

The Pavilion HyperParallel Data Platform takes advantage of NVMe-oF and RoCE technologies to deliver high performance and ultra-low latency IO. The Pavilion solution is capable of latencies as low as 100µs for reads and 25µs for writes.

These latency numbers are measured at the host and include network latency.

Latencies this low means that the Pavilion HyperParallel Data Platform delivers external, scalable storage with performance comparable to direct attached NVMe SSDs.

Analyze Data at Scale

The Pavilion HyperParallel Data Platform provides linear scale out across systems. Each Pavilion system supports up to 2PB of usable capacity and can deliver up to 120GB/s of performance. Two systems can deliver 4PB and up to 240GB/s, and ten systems offer 20PB and up to 1200GB/s. Every incremental system provides linear increases in performance and capacity.

The Pavilion HyperParallel Data Platform

Pavilion delivers universally unmatched storage with the same or better performance than direct attached NVMe SSDs and unlimited scale to support any size data set. The most performant, dense, scalable, and flexible data storage platform in the universe, the Pavilion HyperParallel Data Platform uniquely accelerates AI to train faster and analyze more data in less time, delivering results that were previously thought impossible.

About Pavilion

Pavilion shatters customer expectations and resulting organizational outcomes by revolutionizing data processing for modern AI/ML, HPC, Analytics, Enterprise Edge and other data-driven applications. The Pavilion HyperParallel Data Platform, powered by Pavilion HyperOS, delivers unmatched performance and density, ultra-low latency, unlimited scalability and flexibility, providing customers unprecedented choice and control. Learn why Fortune 500 companies and federal government agencies choose Pavilion. Visit www.pavilion.io or follow the company on LinkedIn