NVMe-oF Storage for Spark

The most performant, dense, scalable, and flexible data storage platform for Spark

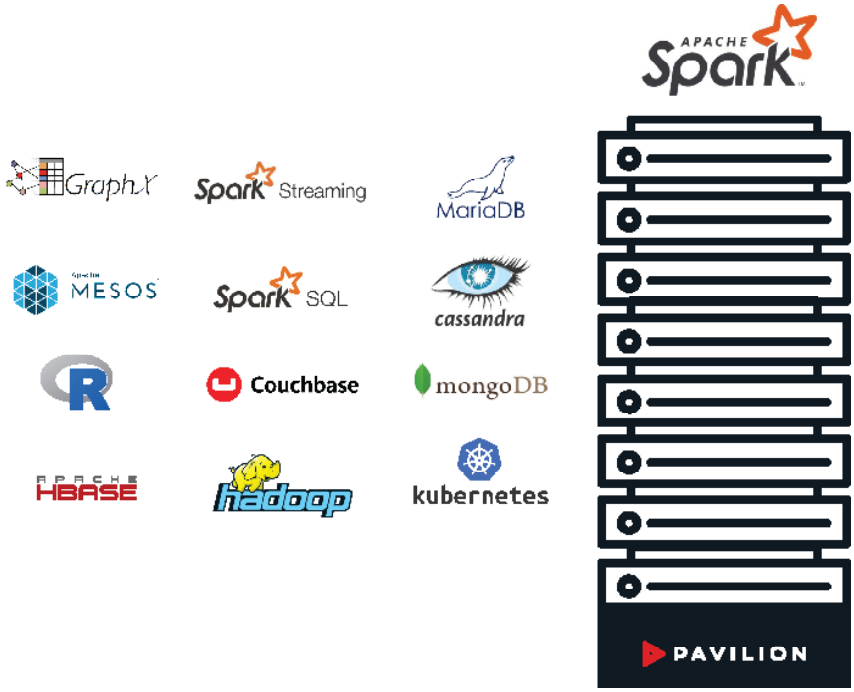
Modern applications that use Spark to analyze structured and unstructured data need to be future proofed. Reduce storage over provisioning, accommodate storage growth, consolidate applications and their data, speed time to results, and reduce TCO by disaggregating Spark storage with the Pavilion HyperParallel Data Platform™. It is the industry's first NVMe oF storage array. It scales to over a petabyte of data, provides ultra-high performance, ultra-low latency, linear scalability, and uses NVMe SSDs.

A DAS-based Spark infrastructure results in 2-3X overprovisioning and islands of storage—while using software-defined storage takes CPU resources away from Spark, impacting time to results and increasing TCO. Pavilion changes all that with the Pavilion HyperParallel Data Platform.

The Pavilion HyperParallel Data Platform

The Pavilion HyperParallel Data Platform delivers never before seen NVMe performance (120 GB/s throughput, 25µs of latency, and 20M IOPS) & density (2.2 Petabyte) in a compact 4U storage array. It gives Spark applications the performance of locally attached NVMe SSDs enabling organizations to move to a Composable, Disaggregated Infrastructure (CDI) infrastructure, where application resources are readily available.

The Pavilion HyperParallel Data Platform requires no proprietary software to be installed on a server farm and uses standard NVMe, NVMe-oF, Ethernet, and InfiniBand drivers, freeing up host resources for processing, reducing deployment



Benefits

- Increases Spark density 2X
- Petabyte scalability, high-performance, low-latency, and linear scaling maximizes data center efficiency
- Protection for your data and your business. Meet evolving requirements for data security and compliance
- Search 20X faster than DAS
- Deployment flexibility using concurrent protocols
- Enterprise design and data integrity validation ensure reliable access to data
- Rack scale/CDI management via Web GUI, vCenter, Kubernetes, RESTful API, OpenStack, DTMF/ Redfish, and Swordfish

Scalable & Flexible

Provide up to 2.2 Petabytes to Spark deployments and simultaneously use the NVMe /RDMA (Ethernet, IB), NVMe/TCP, iSCSI, NFS (v3, v4, pNFS, RDMA) or S3 interfaces. Grow performance and capacity linearly without impacting on-going operations. Pavilion's thin provisioning feature provides Spark with significantly better utilization than DAS, since the Pavilion array will only allocate physical space when the Spark application needs it. This results in physical space savings vs. DAS of up to 75% per server. No wasted capacity, no time spent messing with volume managers or file systems. Just set and forget.

Fast & Dense

The ultra-high performance, extreme low latency, and multiple storage controllers of the Pavilion HyperParallel Data Platform accelerates Spark workflows and boosts time to results. The Pavilion HyperParallel Data Platform lets Spark applications search terabytes of structured and unstructured data over 20 times faster than a DAS NVMe SSD.

Get 2X more density of a Spark cluster by using storage-less server nodes and moving all storage and storage management to the Pavilion HyperParallel Data Platform. No custom software needs to be installed on the Spark cluster, enabling it to take full advantage of host processing resources as well as simplifying deployment complexity.

Economical

Spark performs parallel computing across nodes and uses local SSDs and RAM to reduce the I/O and execution times of tasks. SSDs and RAM are some of the most expensive components of the Spark cluster, why not reduce them both? The Pavilion HyperParallel Data Platform disaggregates storage while providing Spark applications with ultra-high performance and ultra-low latency. This enables DevOps to eliminate DAS NVMe SSDs and remove RAM in every Spark server, reducing TCO.

With Pavilion, you no longer are constrained by the size of the SSDs. Thin provisioning allows the application to use the required amount of storage at any given time, regardless of how much capacity has been advertised to that specific Spark server. No more extra copies of data, the Pavilion HyperParallel Data Platform reduces the amount of raw flash storage deployed in a Spark environment, by up to 3X, reducing TCO. At 50% less expensive than a DAS SSD in terms of GB/s, the Pavilion HyperParallel Data Platform is designed for all your Spark environments.

Safe & Secure

Protect the SSDs with RAID-6 and Pavilion SwarmController™, which rebuilds a failed SSD at the rate of less than 5 minutes per TB. The Pavilion HyperParallel Data Platform ensures that the failure of an SSD does not impact Spark operations.

Keep data and snapshots safe with FIPS-compliant data at rest encryption that ensures its always-on encryption keeps data secure without impacting Spark performance.

The Pavilion data assurance features work with the RAID feature to provide self-healing bit-rot support for data, assuring every Spark workflow gets uncorrupted data. Take consistent, instant, zero-footprint, and uncorrupted snapshots, encrypt them and provide them to backup and disaster recovery processes to speed operations and ensure consistency.

Enterprise Strength & Support

Get 24/7 proactive support, end-to-end data integrity, a modular chassis, and redundancy throughout the storage array to protect your Spark applications as infrastructures grow. With the Pavilion HyperParallel Data Platform, all the features come in-the-box, including thin provisioning, snapshots, clones, data at rest encryption and more.

Manage the Pavilion HyperParallel Flash Array via its Web GUI or use the management framework of your choice, including: vCenter, Kubernetes, RESTful API, OpenStack, DTMF/Redfish, Swordfish, and more.

Find Out More

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 twitter at https://twitter.com/PavilionData