



Big Data Analytics

Use Case

- Parallel processing cluster for analytics using Apache Hadoop® and Apache Spark®

Challenge

- Upcoming EOL notification, slow SSD performance, and support issues for multiple servers with DAS SSDs
- Data growth resulted in excessive resource utilization, performance decline, and poor TCO
- Data migration required doubling DAS SSDs

Solution

- Disaggregated storage from 120 servers
- Deployed two Pavilion Hyperparallel Flash Arrays with over 1.4PB of capacity
- Moved from proprietary 2U servers to commodity 1U servers
- Leveraged OpenChoice Storage™ to re-use the SSDs in each server

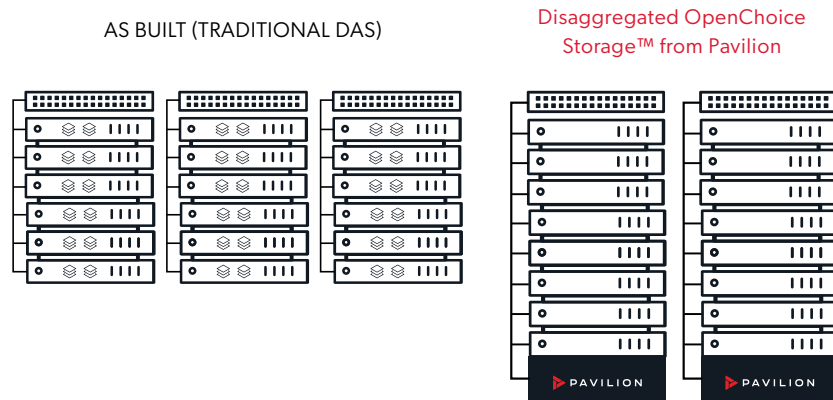
Results

- Reduced storage TCO 2-3X
- Deployed 12X faster and 2X smaller clusters
- Exported data at 120GB/s, meeting SLAs
- Decreased storage management overhead by 3X

This client provides one of the largest cloud-based customer support platforms to thousands of worldwide enterprise customers. They have stringent SLAs for system uptime and are known for providing high performance and low-cost cloud services.

In keeping with their massively parallel design goals, the client implemented a proprietary 120-node rack-scale clustered infrastructure with Hadoop, Spark, and 240 DAS SSDs for high-speed data analytics. The customer's data more than doubled in a few years, and their SSDs could not keep up, increasing their time to results. They also needed to export the data to a different system at 120GB/s, which would increase east-west traffic and double their storage expenses. The SSDs in each server were 50%-75% larger than necessary, in anticipation of the application growing. Finally, their supplier was going to issue an EOL notification for their proprietary servers, so they needed a future proofed compute and storage architecture.

Composed By Pavilion



The client disaggregated storage and moved from a proprietary 2U server to a 1U off-the-shelf server, reducing space by 2X. They deployed a pair of 4U Pavilion Hyperparallel Flash Array's (HFA's) with over 1.4PB of available NVMe capacity. They used Pavilion's NVMe-over-Fabric support to provide the ultra-high performance and ultra-low latency that ingested data 2X faster than their DAS SSDs and provided the performance their application's needed. Also, the Pavilion HFA's OpenChoice enabled the client to re-use the NVMe SSDs that were in each 2U server.

The customer used Pavilion's instant, space saving, and transparent snapshot and achieved their goal of more than 120 GB/sec of read bandwidth for data export, without impacting ongoing analysis. Its RAID-6 architecture with rapid SWARM rebuilds, and data assurance technology provided organizational leaders with confidence that downtime could be minimized.

By using shared storage and thin-provisioning the client reduced their SSD needs by almost 75% and were able to perform compliance, backup, and audit operations without impacting analytics.

Benefits

- Export of data does not impact analytics
- Thin-provisioning maximized storage utilization
- Export can be run without impact to production
- Future-proofed storage infrastructure
- 24/7 proactive support

We cannot afford to double the space in our rack-scale design for data migration. Pavilion's zero-impact snapshots with 120GB/sec export is the only solution we have found that meets our strict cost and performance SLA requirements.

Hardware Engineering Architect

Simultaneously, operations staff found the benefits of the Pavilion HFA's SAN-like management for snapshots, clones, and thin provisioning aligned to the future of their applications which require massive parallelism. Additionally, they were able to manage the storage with 1/3 of the resources previously required by using the Pavilion HFA's SAN-like management.

Summary

Team Pavilion brought expertise to the design, implementation, and support of a mission-critical system shared by this client's customers across the globe. Using standards-based hardware, protocols, and orchestration, this customer reduced risk, improved analytics time-to-insights and defined a rack-scale architecture that can evolve with their requirements.