

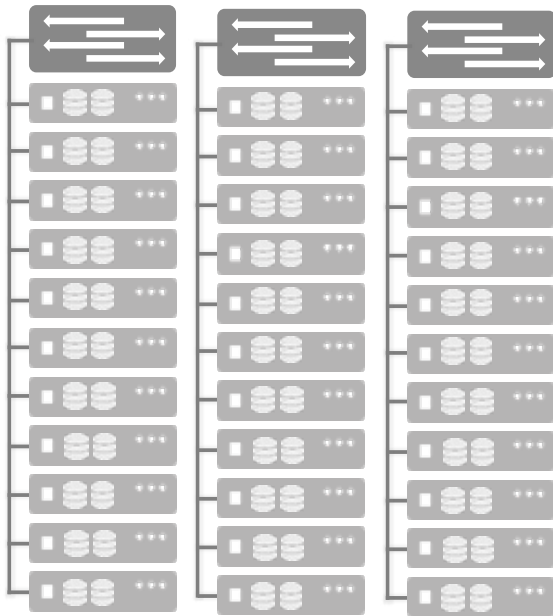
### INTRODUCTION

Big data analytics can present big problems. The larger the data set, the more expense is allocated to storage. This leads to compromises in the performance of SSDs deployed, slowing analytic insights. Larger data sets usually imply large footprints of space and power for storage systems. In many cases, these systems are legacy SANs with associated vendor lock-in management tools and exorbitant support contracts.

### CHALLENGE

A 120 node parallel data processing cluster using Apache Hadoop® and Spark® for analytics needed future-proofing with NVMe storage super-fast Infiniband NVMe-Over-Fabrics network transport. Existing analytics storage was based on a proprietary architecture with limited support and an end-of-life notification forthcoming from the supplier.

#### AS BUILT (TRADITIONAL DAS)



- Proprietary Architecture
- Limited Support Capability
- Pending EOL notification

#### DISAGGREGATED STORAGE (OPENCHOICE FROM PAVILION)



- Standards-based NVMe-Over-Fabrics
- 24x7 Support
- Future-proof Design

### COMPOSED BY PAVILION

Using disaggregated, rack-scale flash, the customer was able to achieve their goal of more than 120 GB/sec of read bandwidth with a pair of single 4U appliances totaling more than 1.4PB of available NVMe capacity. The full RAID-6 architecture with rapid swarm-based SSD rebuilds provided organizational leaders with confidence that downtime could be minimized. Simultaneously, operations staff found the benefits of SAN-like management for snapshots, clones and thin provisioning aligned to the future of their applications which require massive parallelism.

### SUMMARY

Team Pavilion brought expertise to the design, implementation, and support of a mission-critical system shared by research teams 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.