PAVILION

Greenplum

Use Case

- Centralized backup of multi-petabyte Greenplum data lake

Challenge

- Server and data growth resulted in excessive resource utilization
- Backup not meeting SLAs & traffic impacting other workloads

Solution

- Disaggregated storage; moved from 72 servers using 8 15K HDDs in each server to 36 servers and one Pavilion Hyperparallel Flash Array
- Instant, zero-space snapshots and clones
- Consistency Groups
- On-Demand reconfiguration

Results

- 40% TCO reduction
- 24X quicker backups
- 2X reduction in data center footprint
- 90% of backup traffic eliminated
- 3X lower storage management Opex

Benefits

- Backup SLAs met
- Backups can be run on demand
- Analytics can be run without impact to production or backups

Our Data Lake is mandatory for compliance. We could not back it up with traditional storage. Pavilion is the only storage platform that provides instant snapshots, nearly zero network traffic and dramatically faster analytics.

**Director of *Infrastructure***

Data Lake Centralized Backup

This customer is the banking arm of one of the world's leading financial services organization. Ensuring that data is in compliance and protected is of paramount concern. This bank operates one of the world's largest Greenplum data lakes for financial systems analytics.

In keeping with their massively parallel design goals, the financial institution implemented the Greenplum data lake using multiple servers with each server node having 8 X 15K RPM HDDs. Their data lake grew to more than 3PB over a short period of time and backup traffic, a strict backup window, and server sprawl all consumed more resources and space than the team had forecast. A new approach to addressing high-speed backup was required.

The customer first looked to Dell EMC for a solution as at that time Greenplum was an Dell EMC company, however the volume and velocity of the bank's data lake growth could not be addressed by any part of the Dell EMC portfolio, so they turned to Pavilion for scale, performance and management.

Composed By Pavilion

40 Node Greenplum Cluster

Pavilion Shared Storage
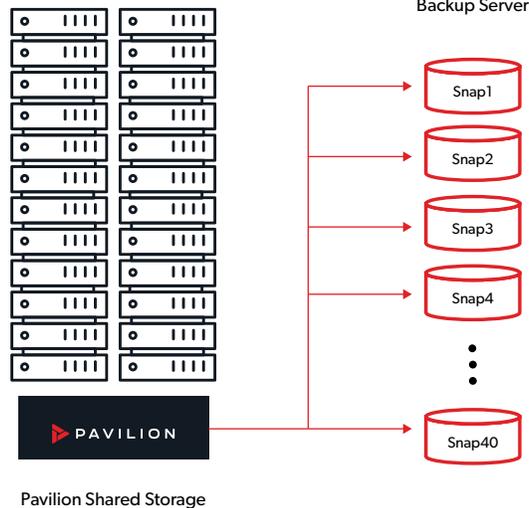
By deploying the 4U Pavilion Hyperparallel Flash Array (HFA) with its NVMe-Over-Fabrics support and leveraging its instant and zero-footprint snapshots, the financial institution was able to reduce their backup time from over 24 hours to less than one, this allowed them to meet their backup window.

With a design that features direct copies to local disk, East-West network traffic was virtually eliminated. The result was secure, compliant backup using familiar shared storage tools, with all of the benefits of scale-out architectures. Simultaneously, the team was able to reduce the number of servers required by half. Additionally, they were able to manage the storage with 1/3 of the resources previously required using the Pavilion HFA's SAN-like management.