CASE STUDY

DATA LAKE CENTRALIZED BACK UP

pavilion data

Make Better Decisions, Faster.

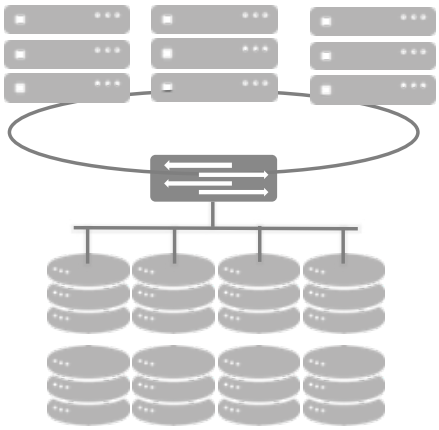
INTRODUCTION

For one of the world's largest financial institutions, backup and compliance of a Greenplum data lake are of paramount concern. With a centralized cluster that has grown to more than 1PB over a short period of time and a strict backup window, a new approach to addressing high-speed backup was required.

CHALLENGE

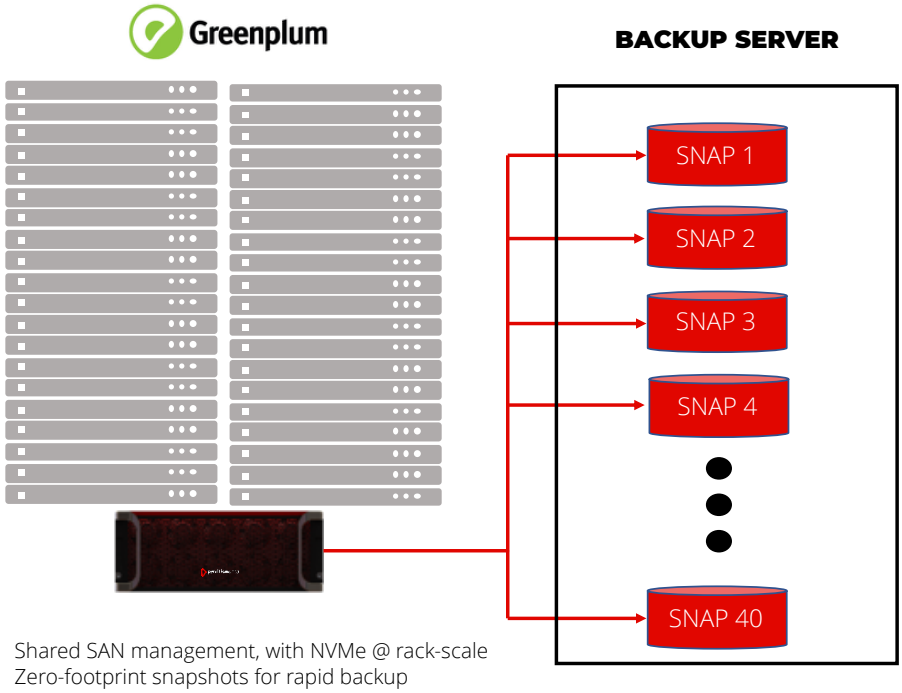
For this financial institution, centralized and shared storage provided great tools for data management. However, as the data lake increased in size, network traffic for backups quickly became a bottleneck, and the sprawl of SAN-based storage consumed more space and power than the team had forecast.

AS BUILT
(WITH TRADITIONAL SAN)



- Simple management of backups from shared SAN
- Unable to meet backup windows @ 1PB scale
- Network traffic impacting other workloads

DISAGGREGATED STORAGE
(WITH SNAPSHOTS FROM PAVILION)



- Shared SAN management, with NVMe @ rack-scale
- Zero-footprint snapshots for rapid backup
- Direct copies of snapshots to backup targets minimizing network traffic

COMPOSED BY PAVILION

By deploying Pavilion NVMe-Over-Fabrics and leveraging zero-footprint snapshots, the financial institution was able to immediately meet their backup windows. 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 footprint for storage from multiple racks to a single 4U appliance.

SUMMARY

Pavilion technology was implemented seamlessly into the environment with all of the benefits the team previously had with centralized SAN storage. However, the impact was astounding. Smaller footprint, instant backups, minimal network traffic to impact other workloads. This organization even returned to performing weekly full backups, a capability that had long been stopped due to the size of complete volume copies on the SAN. In sum, as data grew, Pavilion was able to shrink physical and logical footprints along with the backup window.