WHITE PAPER
September 2024

The Impact of Samsung Gen-5 NVMe on VMware Tanzu® Greenplum Optimization

Table of Contents

VMware Tanzu Greenplum

Learn more about the full capabilities and components of VMware Tanzu® Greenplum: tanzu.vmware.com/greenplum

Executive Summary

We stand at the precipice of a transformative era. The ever-growing volume, velocity, and variety of data, known as big data, is fueling a revolution in Artificial Intelligence (AI). The growing sophistication of artificial intelligence (AI) necessitates robust data platforms like VMware Tanzu Greenplum. Advanced AI models require massive datasets for training and inference, making efficient big data management critical for their success.

VMware Tanzu Greenplum is a massively parallel processing (MPP) database, It delivers exceptional scalability, seamlessly managing multi-petabyte data workloads. VMware Tanzu Greenplum leverages a cluster of powerful servers to execute complex queries in parallel across the entire dataset. However, traditional hardware infrastructures often hinder VMware Tanzu Greenplum's ability to reach its full performance potential.

Samsung PM1743 NVMe SSD is an ideal solution, significantly enhancing data processing speeds and overall system performance. This enhancement enables legacy systems to align with modern data analytics demands and ensures compatibility with VMware Tanzu Greenplum.

This white paper explores the transformative potential of New Samsung Gen-5 Non-Volatile Memory Express (NVMe) devices in optimizing VMware Tanzu Greenplum's performance which include:

- 3.6x faster overall performance compared to traditional storage
- 12x dramatic reduction in simple query execution times
- Transform query processing efficiency gains 3 times
- 70% reduction in footprint while maintaining identical workload performance
- Combined initial hardware and ongoing operational costs can be reduced by 80% over a defined five-year period (49.6% reduction in hardware cost and 70% reduction in electricity cost)

Background

Optimizing the Big Data Journey: New Approaches to Storage and Management

The exponential growth of big data is transforming the way organizations operate and make decisions. Speed and agility is now the ultimate double-edged sword for organizations' future growth or extinction.

While big data holds immense potential for unlocking valuable insights, it also presents significant challenges for storing data at scale and sustaining performance at the agile business level. Moreover, storing and maintaining storage infrastructure at the massive scale is eminent challenge that organization face financially and physically.

Evolution of Solid-State Drive (SSD)

The evolution of Solid-State Drives (SSDs) has been a crucial factor in big data analysis. The initial phase of this evolution involved the transition from Hard Disk Drives (HDDs) to Serial Advanced Technology Attachment (SATA) SSDs. This migration significantly improved the performance of big data analysis as the SATA SSDs offered faster read write speeds and reduced latency.

The second phase of SSD evolution was the introduction of NVMe (Non-Volatile Memory Express) SSDs that operated on the PCIe (Peripheral Component Interconnect Express) interface. NVMe SSDs provided even higher performance compared to SATA SSDs, resulting in quicker data processing and analysis. These SSDs also offered lower latency, making them highly suitable for real-time data analytics.

Current generation of NVMe SSDs with PCIe Gen5 have impressive sequential and random read speeds of 14,000 MB/s and 2,500 KIOPS respectively, as opposed to SATA SSDs that have sequential and random speeds of 550 MB/s and 98 KIOPS. The physical interface limit of SATA SSDs at 600 MB/s restricts further performance improvements. On the other hand, NVMe has the potential for further performance improvements in future generations. The current generation of NVMe uses a PCIe Gen5 interface with four lanes. It is expected that PCIe Gen6, which will double the bandwidth, will be launched in the coming years. Furthermore, the number of PCIe lanes can be increased to enhance the bandwidth. Another performance-related difference between SATA and NVMe is the queue depth. SATA SSDs have a limited queue depth of one, while NVMe can theoretically support up to 64,000 queue depth, making it more efficient for handling random tasks.

In terms of form factors and capacities, SATA SSDs are available in the U.2 form factor with a maximum capacity of 8 TB, while NVMe SSDs come in various EDSFF (Enterprise and Datacenter Storage Form Factor) with a capacity of up to 16 TB. This is crucial in limited server space as NVMe SSDs can fit into a 1U server, while SATA SSDs require a 2U server.

VMware Tanzu Greenplum in the Big Data Landscape

The big data landscape is brimming with diverse database solutions, each catering to specific needs. VMware Tanzu Greenplum stands out as a compelling choice for organizations navigating the complexities of big data due to its unique set of benefits.

- VMware Tanzu Greenplum leverages an MPP (Massively Parallel Processing) architecture, distributing data and processing tasks across multiple nodes. This enables parallel processing, significantly improving query performance and handling massive datasets efficiently. This is particularly beneficial for big data analytics involving large data volumes.

- VMware Tanzu Greenplum boasts exceptional scalability. New nodes can be seamlessly added to the cluster as data volumes grow, ensuring the system can keep pace with big data demands. Additionally, it offers elastic scaling, allowing organizations to scale resources up or down based on workload requirements. This flexibility optimizes resource utilization and cost efficiency.

• It is built on familiar SQL standards, making it accessible to existing database users and analysts. This reduces the learning curve for new users and simplifies integration with existing data pipelines and tools commonly used in the big data ecosystem.

• VMware Tanzu Greenplum goes beyond basic data storage and retrieval. It empowers users with advanced analytics functionalities, including built-in support for machine learning algorithms and data mining tools. This allows organizations to extract deeper insights from their big data for data-driven decision making.

Samsung PCIe Gen-5 NVMe (PM1743) Drives

We have chosen the Samsung PCIe Gen5 NVMe drives, specifically the PM1743 models, as we expected to leverage our architecture with their superior performance and impressive capacity specifications. These drives excel at balancing exceptional speed with substantial storage, making them ideal for data-intensive workloads.

Within this architecture, individual servers we have used in the test scenario can accommodate up to eight Samsung PM1743 drives, each offering a capacious 16TB of storage. This configuration yields a formidable raw capacity of up to 128TB per server, creating a substantial storage pool capable of effortlessly managing expansive datasets.

The table below provides a performance comparison of various storage device types for reference.

Read/write speed comparison

	PM893 SATA SSD	PCIe Gen4 SSD*	PM1743 Gen5 SSD
Sequential Read	550 MB/s	6,700 MB/s	14,000 MB/s
Sequential Write	520 MB/s	4,100 MB/s	6,000 MB/s
Random Read	98K IOPS	1,000 K IOPS	2500 K IOPS
Random Write	30K IOPS	180 K IOPS	280 K IOPS

Table 1: *PM9A3 as PCIe Gen4 drive NOT in-use in this test scenarios, only for reference purpose.

With the performance comparison above, The Samsung PM1743 drives exceptional read speeds hold significant advantages for VMware Tanzu Greenplum deployments. Their ability to rapidly retrieve large datasets from disk into memory liberates the CPU from the data retrieval bottleneck, enhancing overall performance.

By offloading data retrieval to the high-performance storage, the system operates with greater efficiency. The CPU, freed from this intensive task, can dedicate its resources to core computational tasks and complex analytics. This significantly enhances the system's ability to handle massive data loads and deliver faster results for demanding analytical workloads.

VMware Tanzu Greenplum Architecture

The VMware Tanzu Greenplum architecture we have used for this white paper is centered on two base server models. (Dell R750 and Dell R7625). Each architecture consists of four servers. These servers house 16 PM893 SATA drives, 4 PM1743 NVMe drives and 8 PM1743 NVMe drives respectively, with configurations varying based on server capacity. For this test, a shared master node is additionally included in the configuration for accepting client connections and SQL queries, and distributing work to the segment instances. As this architecture is designed for horizontal scaling, additional servers can be seamlessly integrated as needed to maintain desired performance and efficiency levels. Each server within this architecture utilizes dual 100 Gigabit network interface cards (NICs) bonded together to achieve increased bandwidth and mitigate potential network bottlenecks.

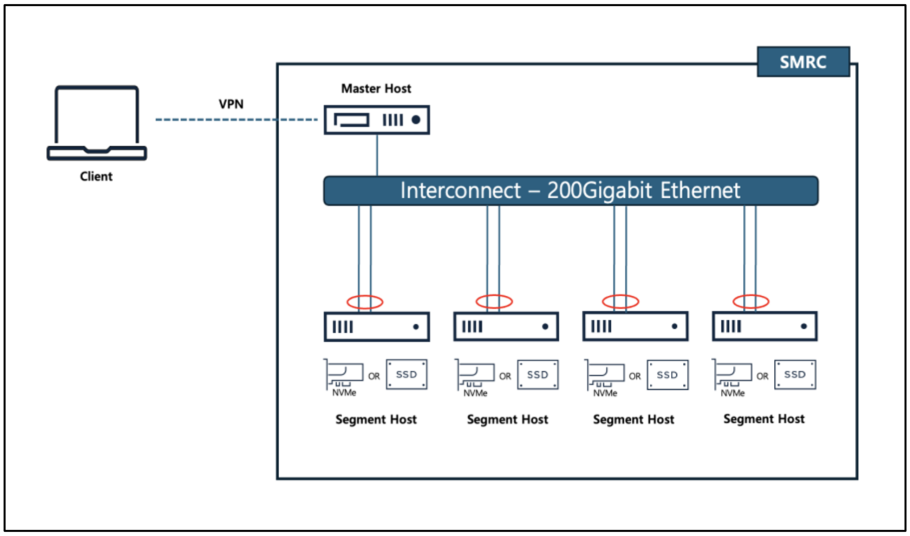


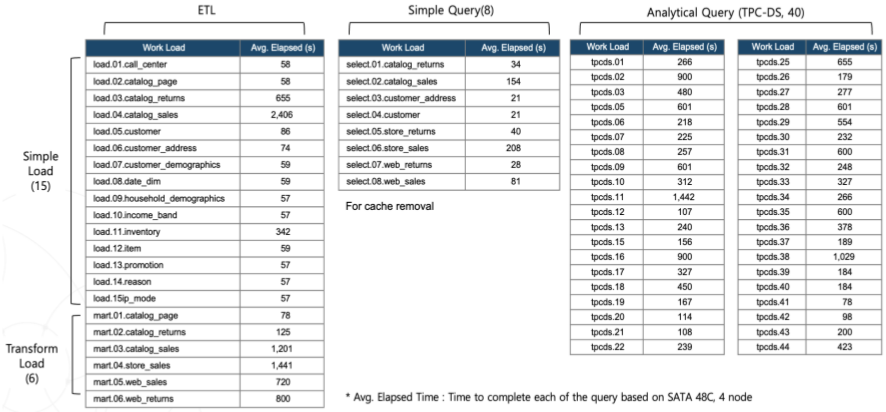
Figure 1: Infrastructure Architecture for VMware Tanzu Greenplum Testing

Each server in this architecture leverages VMware Tanzu Greenplum. The MPP capabilities make it an ideal platform for handling large and complex datasets and workloads. Finally, each server in this architecture executes the open-source Rocky Linux version 8.7, adhering to the recommendations outlined in the latest VMware Tanzu Greenplum updates for supported operating systems.

Benchmark Test Suites: TPC-DS

The TPC-DS Benchmark (Transaction Processing Performance Council Decision Support) is an industry-standard tool designed to measure the performance of decision support systems on modern data warehouse platforms. It simulates the processing of complex queries against a large dataset that reflects real-world business scenarios in the retail industry. The TPC-DS Benchmark contains 99 distinct SQL queries. These queries encompass a variety of complexity levels and target different aspects of data stored within the TPC-DS schema.

In this white paper, we used additional mixed workloads to perform benchmark tests under various circumstances, including simple load, transformational load, and a simple query for removing cached data from memory. The test includes only selected 40 TPC-DS queries to create a more balanced test condition.



ETL

	Work Load	Avg. Elapsed (s)
Simple Load (15)	load.01.call_center	58
	load.02.catalog_page	58
	load.03.catalog_returns	655
	load.04.catalog_sales	2,406
	load.05.customer	86
	load.06.customer_address	74
	load.07.customer_demographics	59
	load.08.date_dim	59
	load.09.household_demographics	57
	load.10.income_band	57
	load.11.inventory	342
	load.12.item	59
	load.13.promotion	57
	load.14.reason	57
	load.15ip_mode	57
Transform Load (6)	mart.01.catalog_page	78
	mart.02.catalog_returns	125
	mart.03.catalog_sales	1,201
	mart.04.store_sales	1,441
	mart.05.web_sales	720
	mart.06.web_returns	800

Simple Query(8)

Work Load	Avg. Elapsed (s)
select.01.catalog_returns	34
select.02.catalog_sales	154
select.03.customer_address	21
select.04.customer	21
select.05.store_returns	40
select.06.store_sales	208
select.07.web_returns	28
select.08.web_sales	81

For cache removal

Analytical Query (TPC-DS, 40)

Work Load	Avg. Elapsed (s)	Work Load	Avg. Elapsed (s)
tpcds.01	266	tpcds.25	655
tpcds.02	900	tpcds.26	179
tpcds.03	480	tpcds.27	277
tpcds.05	601	tpcds.28	601
tpcds.06	218	tpcds.29	554
tpcds.07	225	tpcds.30	232
tpcds.08	257	tpcds.31	600
tpcds.09	601	tpcds.32	248
tpcds.10	312	tpcds.33	327
tpcds.11	1,442	tpcds.34	266
tpcds.12	107	tpcds.35	600
tpcds.13	240	tpcds.36	378
tpcds.15	156	tpcds.37	189
tpcds.16	900	tpcds.38	1,029
tpcds.17	327	tpcds.39	184
tpcds.18	450	tpcds.40	184
tpcds.19	167	tpcds.41	78
tpcds.20	114	tpcds.42	98
tpcds.21	108	tpcds.43	200
tpcds.22	239	tpcds.44	423

* Avg. Elapsed Time : Time to complete each of the query based on SATA 48C, 4 node

Table 2: Mixed workload scenarios**.**

Each test was conducted for a duration of two hours. During this period, the aforementioned queries were executed repeatedly and concurrently. The completion time for each query was observed, and the results were summarized by calculating the ratio compared to the baseline configuration.

VMware Tanzu Greenplum Test Scenarios

Highlighting Hardware Impact on Database Performance: Test Scenario Design

This section of the white paper will explore the impact of hardware configurations on database performance through a series of controlled test scenarios. We aim to demonstrate how different hardware setups can influence query execution times, throughput, and overall database responsiveness. We hypothesized that increasing compute resources on the NVMe configuration would lead to a more significant performance gain compared to the traditional SATA configuration, which bottlenecked by disk I/O. (Specifically, we anticipated that doubling the compute resources would result in at least a twofold improvement in database performance for the NVMe configuration.)

Test Cases and Rationale:

Test Case 1: Baseline Configuration: This test will establish a baseline performance benchmark using SATA disks. This configuration will serve as a reference point for comparison with subsequent tests that introduce hardware variations.

Purpose: This test establishes a foundation for understanding the performance capabilities of the database software, reflecting the experience of a significant portion of our customer base.

Test Case 2: Enhanced Disk Capabilities: This test will involve running the same workload on a system with Samsung Gen-5 NVMe (PM1743) compared to the baseline configuration. Due to limited availability of same hardware configurations optimized for NVMe disks (E3.S form factor), we opted for a system with slightly lower specifications to ensure compatibility. (System configuration with equivalent CPU cores but featuring a decrease in memory and storage capacity)

Purpose: This test aims to isolate the impact of CPU processing power on database performance. Replacing SATA Disks with NVMe Disks potentially leading to faster execution times.

Test Case 3: Enhanced Computing Power: This test will investigate the impact of increased computing resources on database performance. We will achieve this by running the database on a system with a significantly bolstered computational capacity.

Purpose: This test investigates the influence of NVMe disks on database performance, Boosting compute resources by a factor of two. Ample RAM allows for efficient caching of frequently accessed data, reducing disk I/O operations and potentially improving query response times. Increased CPU cores have the potential to enhance database performance, particularly when combined with NVMe storage that minimizes I/O bottlenecks.

Hardware Specifications

	Test Case 1 16SATA_48C_512g	**Test Case 2 4NVME_48C_384g**	**Test Case 3 8NVME_96C_768g**
Server Model	Dell R750	Dell R7625 (1 Socket)	Dell R7625 (2 Sockets)
Processor	2x Intel Xeon Gold 6342 CPU 2.80GHz (24C)	1x AMD EPYC 9454 48-Core	2x AMD EPYC 9454 48-Core
Memory Capacity	16x DIMM Slots, 32GB DDR4	12x RDIMM Slots, 32GB DDR5	24x RDIMM Slots, 32GB DDR5
Storage Controller	No PERC	No PERC	No PERC
Network Adapter	2x Mellanox ConnectX-5 Ex, Dual Port 100 GbE QSFP28 Adapter	2x Mellanox ConnectX-6 Dx Dual Port 100 GbE QSFP56 Adapter	2x Mellanox ConnectX-6 Dx Dual Port 100 GbE QSFP56 Adapter
Drive Bays	16x U.2 SSD Samsung MZ7L33T8 3.84TB	4x E3.S NVMe Samsung MZ3LO15THBLA 16TB	8x E3.S NVMe Samsung MZ3LO15THBLA 16TB
Power Supply	2400W	2400W	2400W
Rack unit	2U	2U	2U

Table 3: VMware Tanzu Greenplum Segment Node H/W Specifications

Disk IO performance results

- NVMe 4EA improves Disk IO performance over SATA SSD 16EA
- Disk IO performance improves as much as the number of NVMe

Hardware Performance Benchmark

To verify the baseline hardware performance, we will leverage the industry-standard iperf tool for network performance in conjunction with hardware benchmarks using the gpcheckperf provided by VMware Tanzu Greenplum.

Read/write speed comparison

	Test Case 1 16SATA_48C_512g	Test Case 2 4NVME_48C_384g	Test Case 3 8NVME_96C_768g
Disk Read MBytes/ Node	8,467	11,994	21,140
Disk Write MBytes/ Node	6,972	8,491	15,705
Disk Read MBytes/ Cluster	33,868	47,977	84,559
Disk Write MBytes/ Cluster	27,890	33,966	62,818

Table 4: Disk IO performance by H/W configuration

Performance Benchmark Test Results

Overall Test Summary

The TPC-DS test results revealed significant performance improvements for NVMe configurations, particularly during multi session (5 multi-session). We anticipate observing even greater performance improvements when running more demanding workloads concurrently.

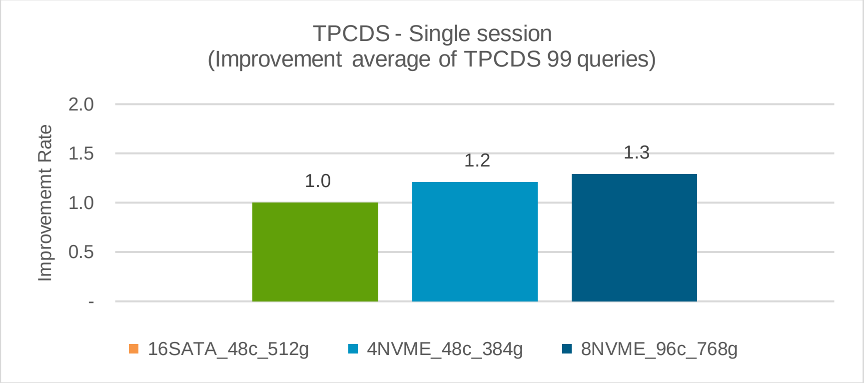


Figure 2: TPCDS single session performance comparison by HW configuration

Mixed workload test results

- [SATA SSD 16EA, 48C, 512GB] vs [**NVMe 4EA**, 48C, 384GB] Performance improvement **1.8 times**
- [SATA SSD 16EA, 48C, 512GB] vs [**NVMe 8EA, 96C, 768GB**] Performance improvement **3.6 times**

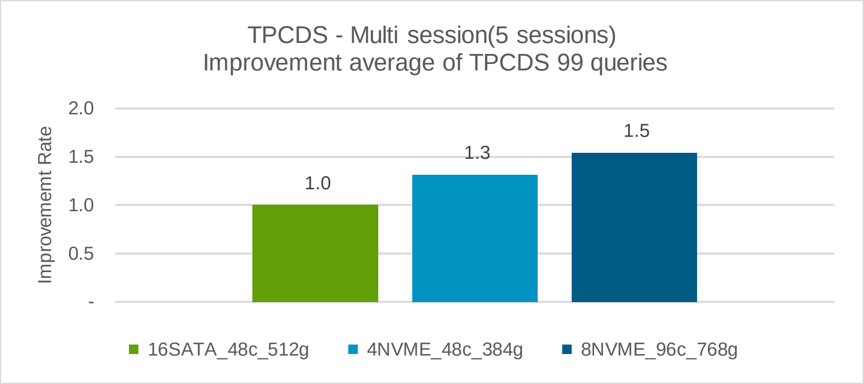


Figure 3: TPCDS multi session performance comparison by HW configuration

To simulate real-world conditions, we combined additional workloads during benchmark execution. Even under this increased load, NVMe configurations demonstrated a significant performance improvement for database operations compared to baseline SATA configuration.

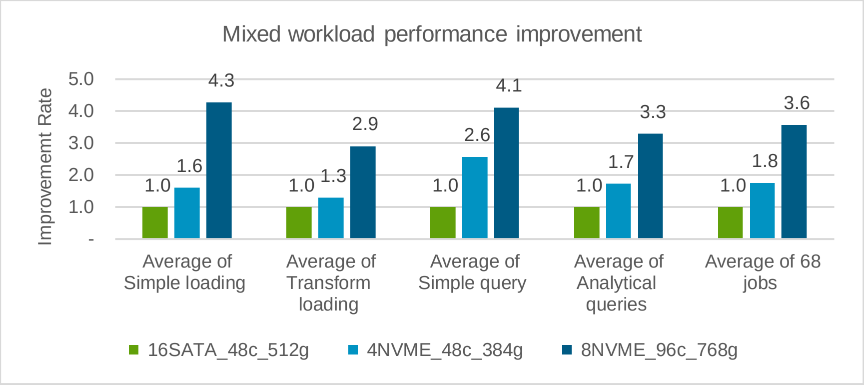


Figure 4: Comparison of mixed workload test performance by HW configuration

Assuming a two-fold performance improvement from the NVMe configuration and considering a doubling of compute resources, we might anticipate an overall database performance gain of approximately four times. The results summarized below demonstrate that transform queries, characterized by their high computational demands, achieved 3.6x improvement in response times. Notably, focusing on frequently encountered short queries, a key user experience factor, reveals the potential for even greater performance gains, up to twelvefold.

Mixed workload test results

- Applying high spec from SATA SSD to NVMe, performance is significantly improved in short queries that users can feel highly.
- Average: 3.6 times, Short Query: 5~10 times

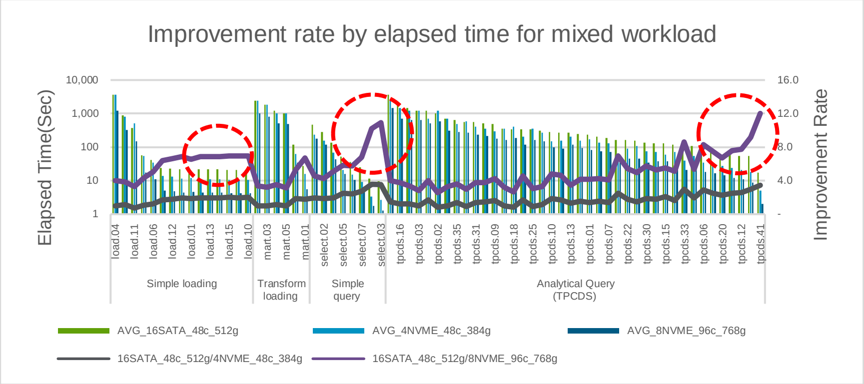


Figure 5: Average elapsed time and improvement rate of each query in mixed workloads

Test 1: SATA vs NVMe

While the SATA storage configuration utilized 16 disks, the NVMe storage configuration achieved a remarkable 20-40% performance increase despite using only 4 disks per node.

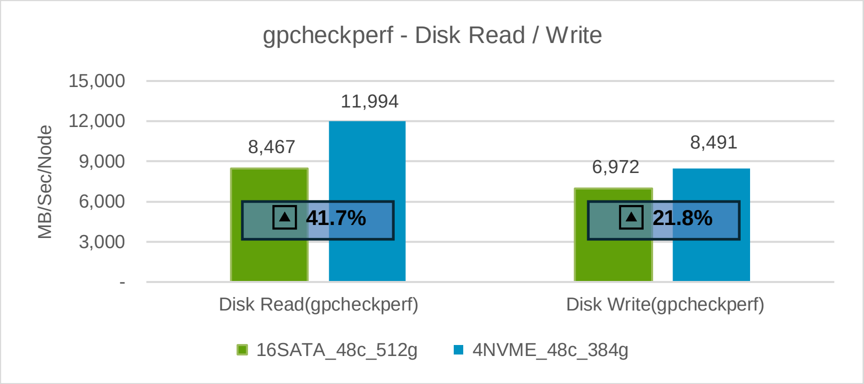


Figure 6: Disk IO performance between 16 SATA SSDs and 4 NVMEs

When the TPC-DS benchmark was executed in isolation, the TPC-DS results revealed that the NVMe configuration delivered approximately 20-40% faster performance compared to the baseline configuration. Our results demonstrate a strong correlation between improved disk performance and TPC-DS scores.

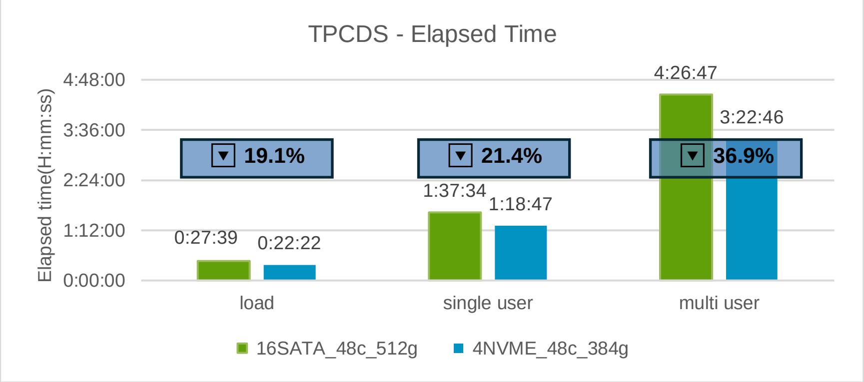


Figure 7: TPCDS elapsed time between 16 SATA SSDs and 4 NVMEs

However, since the benchmark tests above did not fully utilize the system resources, we combined additional workloads to simulate a more demanding environment. With these mixed queries running concurrently, the NVMe configuration delivered substantial performance gains, reaching 80%.

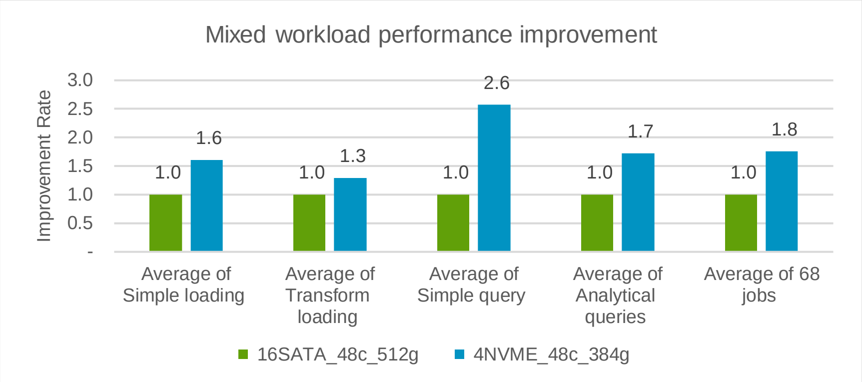


Figure 8: Mixed workload performance improvement ratio between 16 SATA SSDs and 4 NVMEs

When examining the running counts of each tested query, it reveals a dramatic performance improvement for NVMe configuration compared to SATA configuration, particularly for short queries completing in under 100 seconds.

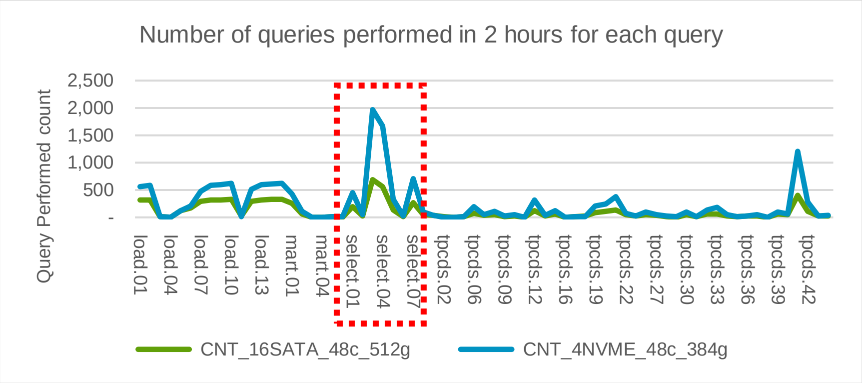


Figure 9: Number of queries performed in 2 hours for each query in mixed workload between 16 SATA SSDs and 4 NVMEs

The Graph below illustrates the performance improvements in simple and transform queries, shorter elapsed times indicate more performance improvement rate. While overall performance improved, transform queries did not show a significant increase under the current workload.

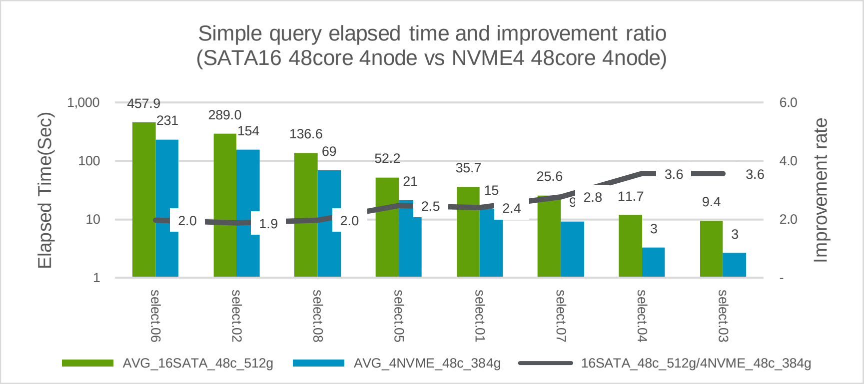


Figure 10: Simple query elapsed time and improvement ratio in mixed workload between 16 SATA SSDs and 4 NVMEs

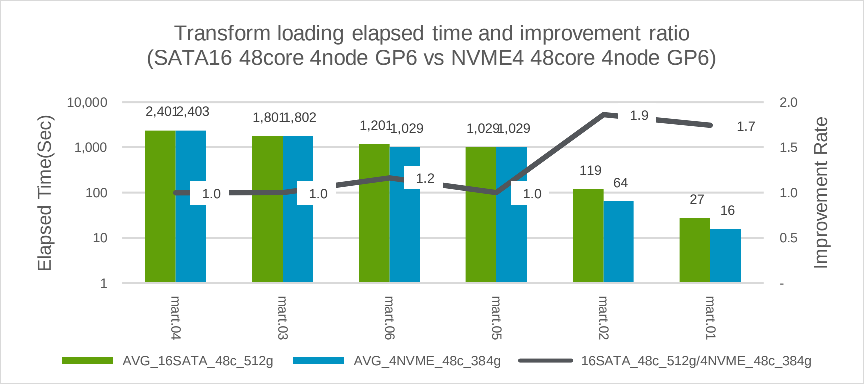


Figure 11: Transform loading query elapsed time and improvement ratio in mixed workload between 16 SATA SSDs and 4 NVMEs

The NVMe configuration demonstrates significant gains, with some queries experiencing 3.6x improvement, particularly for shorter queries. However, the performance gains for complex queries are less pronounced. This is because complex queries require more compute resources for calculation, which becomes the bottleneck rather than disk performance.

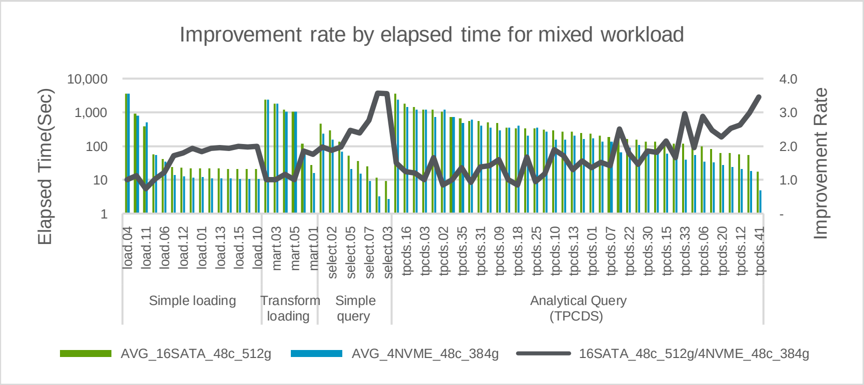


Figure 12: Improvement rate by elapsed time of each query in mixed workloads between 16 SATA SSDs and 4 NVMEs

Test 2: NVMe vs NVMe with High-end Machine

The test results above conclusively demonstrate significant performance gains achieved by replacing SATA storage with NVMe storage, even when utilizing fewer NVMe devices. This suggests that the NVMe configuration likely possesses minimal I/O bottlenecks, enabling it to handle increased workloads efficiently. The following test investigates the performance scalability of Tanzu Greenplum on an NVMe configuration by analyzing performance improvements with increased resource allocation.

We doubled our compute resources, including CPU, memory, and disks. This increase in hardware resources was well reflected in our gpcheckperf benchmark results below.

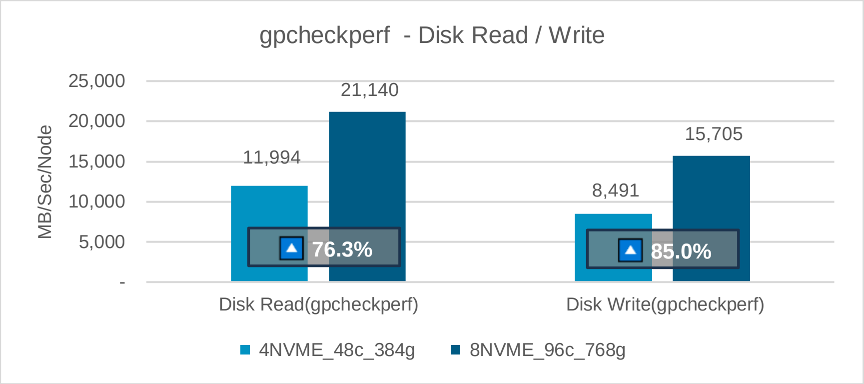


Figure 13: Disk IO performance between 4 NVMEs and 8 NVMes with high-end machine

Our TPC-DS benchmark results showed that, when executed in isolation, the high-end NVMe configuration did not yield significant performance gains compared to the SATA vs NVMe comparison. The high-end configuration only delivered modest performance improvements of around 3-10%. This result suggests that the TPC-DS benchmark may not have fully saturated the additional compute resources.

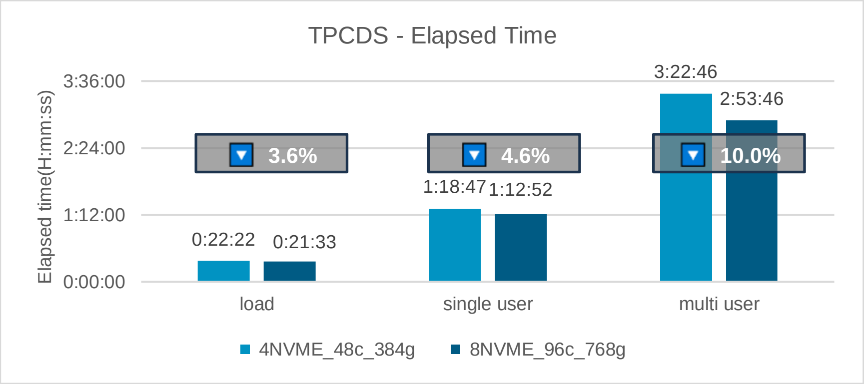


Figure 14: TPCDS elapsed time between 4 NVMEs and 8 NVMes with high-end machine

In contrast, when running mixed queries concurrently, the high-end NVMe configuration delivered significant performance improvements, exceeding those of the NVMe baseline configuration by more than a factor of two. Most notably, transform queries, which are computationally intensive, exhibited a dramatic increase in performance.

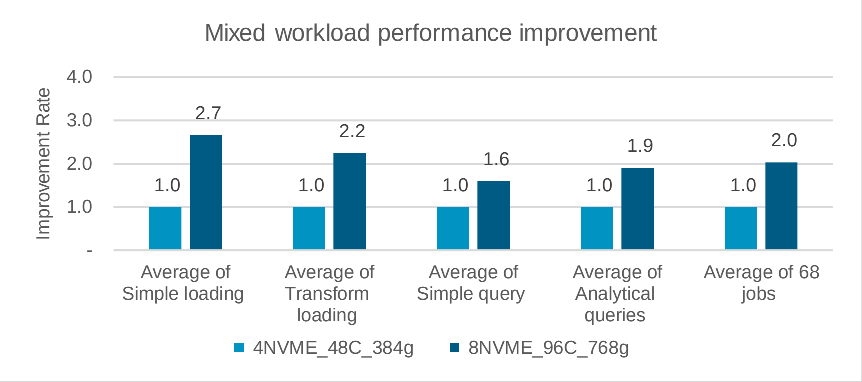


Figure 15: Mixed workload performance improvement ratio between 4 NVMEs and 8 NVMes with high-end machine

Analysis of individual query execution counts demonstrates a substantial performance uplift for the NVMe configuration within overall areas.

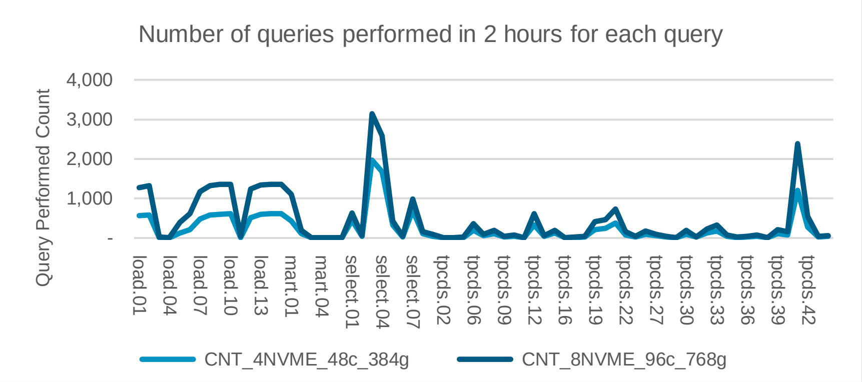


Figure 16: Number of queries performed in 2 hours for each query in mixed workload between 4 NVMEs and 8 NVMes with high-end machine

While the high-end configuration demonstrated overall performance improvements in high-end configuration, simple queries did not exhibit gains as significant as our initial expectation of a twofold increase. However, transform queries showed significant performance improvements, unlike what we observed in the SATA versus NVMe comparison in the first test.

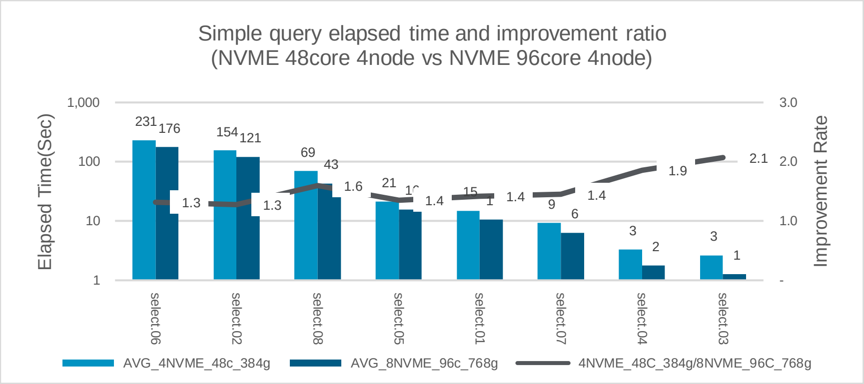


Figure 17: Simple query elapsed time and improvement ratio in mixed workload between 4 NVMEs and 8 NVMes with high-end machine

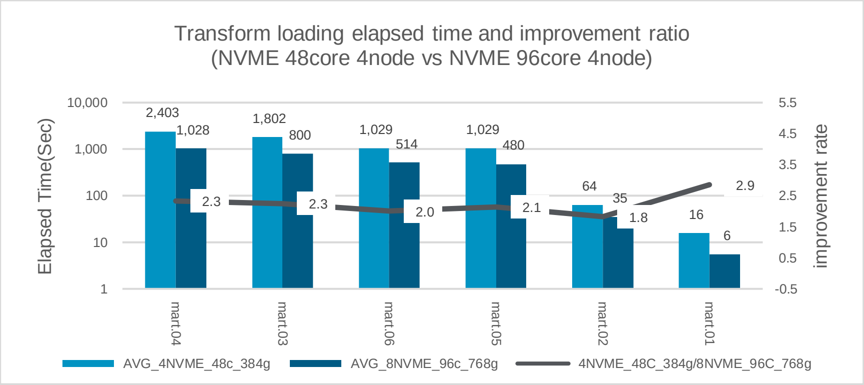


Figure 18: Transform loading query elapsed time and improvement ratio in mixed workload between 4 NVMEs and 8 NVMes with high-end machine

From the test results we performed above, it demonstrates a correlation between increased compute resources and improved database performance as proportional to the amount for resources added. The observed performance improvements were undoubtedly facilitated by the adoption of NVMe storage, which significantly reduced I/O bottlenecks.

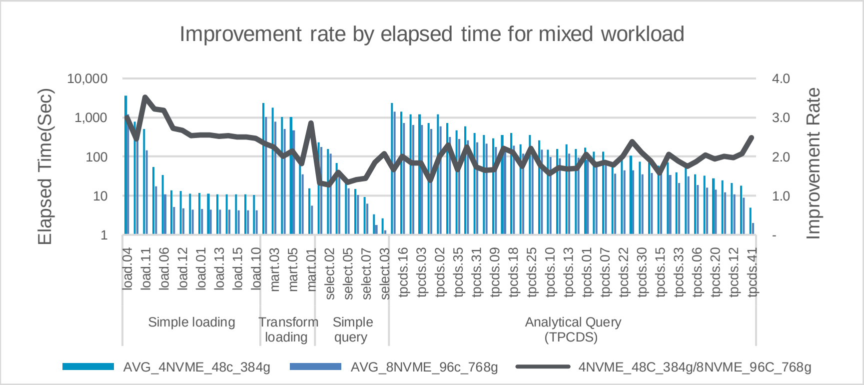


Figure 19: Improvement rate by elapsed time of each query in mixed workloads between 4 NVMEs and 8 NVMes with high-end machine

Based on test results, we anticipate significant cost savings across hardware acquisition, deployment, and electricity consumption. This potential for cost reduction can be achieved while maintaining or even exceeding current performance levels. We will delve deeper into the cost analysis in the next section.

Total Cost of Ownership (TCO) Analysis

From the test results, the NVMe configuration delivers a significant performance improvement of 1.8x compared to the baseline configuration, and a remarkable 3.6x performance increase for the high-end configuration. We can now explore the Total Cost of Ownership (TCO) for each configuration. To determine the number of hosts required, we will assume that the desired performance outcome remains consistent across all configurations.

Hardware Cost Comparison

To estimate hardware costs, we obtained a quotation from hardware manufacturer Dell Technologies. Since the hardware pricing can vary based on Dell's policies and conditions, we have only used estimated prices at this point.

Hardware Cost Comparison

	16SATA_ 48C_512g	**4NVME_ 48C_384g**	**8NVME_ 96C_768g**
Greenplum Performance Improvement (%)	100%	180%	360%
Required Server under identical workload conditions (%)	100%	56%	28%
Required number of Server under identical workload	100ea	56ea	28ea
Per Host Estimated Cost	$150k	$150k	$270k
Per Cluster Estimated Cost	$15M	$8.4M	$7.56M
Total TCO Savings compared to baseline (%)	0%	44%	49.6%

Table 5: Hardware Cost Comparison under identical workload

Key Benefits of using Samsung NVMe with VMware Tanzu Greenplum

- Improve query performance
- Reduced latency variability across all queries
- Total cost saving: H/W cost, Electricity cost, Data center space

Electricity Cost Comparison

To estimate electricity costs, we assumed a daily usage pattern with each server operating under peak load for 12 hours and idle periods for the remaining 12 hours. The calculated percentage improvement in electricity costs is relative to the baseline configuration used in Test Case 1.

Electricity Cost Comparison

	16SATA_ 48C_512g	4NVME_ 48C_384g	8NVME_ 96C_768g	Notes
Peak(W)	900	600	1,000	per Segment Host
Idle(W)	500	300	500	per Segment Host
Daily usage (kWh)	17	11	18	The calculation assumes a peak usage period of 12 hours and an idle period of 12 hours per day.
Daily usage assumption under identical workload conditions (kWh)	17	6	5	Daily usage per Server * Required Server Ratio
Total usage assumption over a defined five-year period (kWh)	3,102,500	1,124,200	919,800	Daily usage per Server * 365 Days * 5 Years * Required number of Servers
Electricity Cost Savings Percentage (%)	0%	64%	70%	Cost savings percentage relative to the baseline configuration

Table 6: Electricity Cost Comparison under identical workload

Key Benefits of using Samsung NVMe with VMware Tanzu Greenplum

The test results and TCO analysis presented above highlight the key benefits of using Samsung NVMe with VMware Tanzu Greenplum. These benefits are summarized as follows:

- 3.6x improvement in database performance for simple loading and query operations.
- 80% reduction in elapsed time for basic queries typically completed within 100 seconds.
- Reduced latency variability across all queries, resulting in more consistent response times.
- Reduced server footprint and operational costs through server consolidation
- 70% reduction in hardware and electricity costs

Learn more

To know more about VMware Tanzu Greenplum visit the product page.

For more information or to purchase VMware products

Call 877-4-VMWARE (outside North America, +1-650-427-5000), visit vmware.com/products, or search online for an authorized reseller.

Conclusion

Our test scenarios have showcased that increasing compute power, particularly CPU cores and memories, can further enhance database performance, especially when leveraged alongside NVMe storage. By offering exceptionally low latency and high bandwidth, NVMe effectively eliminates I/O bottlenecks, paving the way for faster data access and retrieval. This combined approach allows databases to process simple and complex queries more efficiently and deliver faster results.

Basic loading operations, typically completed in 20 seconds, now finish in 4 seconds, representing an 80% reduction in elapsed time. Similarly, basic queries previously requiring 9 seconds can now be completed in under a second, achieving an 90% reduction in elapsed time. Additionally, the increased compute power enables complex transform loading operations to potentially be performed in half the time.

Beyond performance improvements, further benefits can be realized in Total Cost of Ownership (TCO). While the hardware cost per host may nearly double, this can be offset by a total cost savings of as much as 50%. This is achieved through a substantial reduction in the required number of hosts. Additionally, substantial electricity cost savings of 70% are anticipated.

The advent of NVMe storage positions Big Data markets to address longstanding challenges and propel themselves towards the next phase of innovation. The exceptional performance and increased storage capacity offered by NVMe solutions are expected to significantly alleviate concerns around cost and space limitations that have long plagued the industry.

References

- VMware Tanzu Greenplum. (2024, May). About the Greenplum Architecture. https://docs.vmware.com/en/VMware-Greenplum/6/greenplum-database/admin_guide-intro-arch_overview.html
- VMware Tanzu Greenplum. (2024, April). A unified platform for BI to AI. https://tanzu.vmware.com/greenplum
- TPC-DS Benchmark Tool. (2024, May). Description of TPC-DS. https://www.tpc.org/tpcds/default5.asp
- Enterprise-SSD PM1743 (2024, May). Unlock the full potential of enterprise. https://semiconductor.samsung.com/us/ssd/enterprise-ssd/pm1743/

Authors

Sanghee Lee, Korea Data Tech Leader | VMware Tanzu Data Solutions

Kyungjin Park, Client Services Consultant | VMware Professional Services, CX

Michael Kim, Head of Samsung Memory Research Center | Samsung Electronics

Cholmin Kim Ph.D., Principal Engineer | Samsung Electronics

Hyungseuk Kim, Principal Engineer | Samsung Electronics

Samuel Shim, Principal Engineer | Samsung Electronics

Item No: VMW-TNZ-Security-WP100 7/24