



VMware Greenplum Reference Architecture Using Samsung PCIe Gen-5 NVMe Value Kit Drives

Table of contents

Executive summary	3
Background	3
Reference platform architecture overview	4
Reference Samsung PM1743 drives.	5
Samsung SSD Value Kit (SVK)	6
Segment vs. coordinator machine architecture.	6
Greenplum platform recommendations	7
Key hardware components	7
BIOS setup and settings	8
Linux tuning	9
Network setup	9
Number of Greenplum segments per machine	10
Performance benchmark test suites	11
System diagram of server used for testing	11
Description of previous generation hardware	12
Performance benchmark test results.	12
Architecture cost analysis	13
Key benefits of the solution	14
Future work: performance optimization for RAID	16
Conclusion	17
Appendix	17

Executive summary

In a market that is continually evolving, an organization's choice of hardware can make a large impact on the performance and cost of a VMware Greenplum deployment. In working with Samsung, VMware has defined a performance, power and cost optimized reference server platform that benefits from the latest Samsung solid state drive (SSD) storage solutions technology.

This white paper will provide an overview of the VMware Greenplum reference architecture that features Samsung PCIe Gen 5 drives and the Samsung SSD Value Kit, which can enhance performance with energy savings. Benefits to Greenplum customers that use this reference architecture can include:

- Significant improved scan query performance
- Extended performance over concurrent queries
- Reduced power consumption across segment nodes
- Improved total cost of ownership benefit and performance/Watt benefit
- VMware validated platform performance

Such benefits can further enhance the Greenplum customer experience.

Background

As businesses grapple with the escalating pace of big data expansion, there is a compelling need to rethink how they manage, process and interpret this data for more informed decision-making. New technologies are coming to the fore, notably high core count CPUs and innovative storage solutions such as Samsung PCIe Gen5 NVMe Value Kit drives, that are reshaping server architectures fundamental to data processing and analytics.

Use of these advanced storage technologies with powerful processing capabilities can markedly enhance the speed and efficiency of handling large data sets. This technological synergy is geared towards bolstering business intelligence initiatives, enabling the exploration of new use cases, managing extensive data sets, and driving down the total cost of ownership for analytics systems.

Modernization of data systems for analytics and data warehouses is becoming increasingly relevant in this context. Recent trends highlight the integration of cutting-edge storage technologies and robust CPUs with massively parallel processing (MPP) SQL data systems, such as VMware Greenplum. This integrated approach offers businesses a practical pathway to augment the efficiency of their infrastructure for data processing, enabling it to be adaptable, future-ready, and offer a higher return on investment.

VMware Greenplum exemplifies this shift towards modernized, scalable solutions. As an MPP database technology, Greenplum is geared to manage

large-scale data volumes, encompassing everything from terabytes to petabytes. Its shared-nothing architecture, designed to optimize data distribution, could potentially lead to accelerated analytics and enhanced query performance across large data sets. Additionally, with in-database machine learning capabilities, Greenplum aids businesses in performing predictive analytics on a large scale, promoting a culture of data-driven decision-making.

This white paper presents an overview of a reference server platform that strategically combines hardware and software for cost-effective scalability and maximum price performance of data storage and processing. This forward-thinking approach can equip organizations with a robust, scalable and modern data system for analytics and data warehouses, positioning them well to navigate the big data landscape of the future.

Reference platform architecture overview

The reference platform architecture we are discussing in this white paper is centered on a base server model. Each server in this model houses eight Samsung PM1743 drives, a line of cutting-edge storage solutions known for their fast data read and write capabilities, which facilitate efficient handling of large data volumes. [**Note:** Coordinator machines have only two Samsung PM1743 disks, see [Segment vs. coordinator machine architecture](#)].

This base server model is designed to be scaled horizontally, creating a multimachine architecture. The horizontal scalability allows the system to grow in line with increasing data demands, adding additional servers as needed to maintain performance and efficiency levels. This can be a significant advantage, providing both flexibility and cost-effectiveness by allowing for incremental infrastructure expansion rather than requiring large upfront investments.

Each machine in this architecture operates VMware Greenplum software. The MPP capabilities of Greenplum, coupled with its shared-nothing architecture, can provide an ideal platform for dealing with large, complex data sets and workloads. Its ability to leverage the resources of a multinode setup can enhance the overall performance and efficiency of the system. Finally, each machine runs an enterprise-grade Linux distribution, providing a robust, secure and stable operating environment.

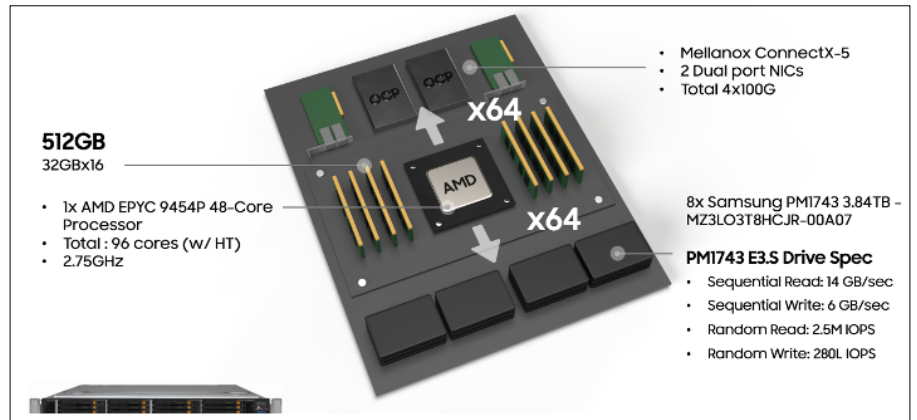


Figure 1: The Samsung Gen-5 NVMe drives, specifically the PM1743 models, have been chosen for this architecture.

Reference Samsung PM1743 drives

The Samsung PCIe Gen5 NVMe drives, specifically the PM1743 models, have been chosen for this architecture owing to their impressive performance and capacity specifications. A notable feature of these drives is their ability to balance exceptional speed with substantial storage capability, making them a prime choice for data-intensive tasks.

In the discussed architecture, each server utilizes eight of these Samsung PM1743 drives, each boasting a storage capacity of up to 3.84TB. This configuration amasses a raw capacity of over 30TB per server, an impressive reservoir of storage that can comfortably handle expansive data sets.

Read/write speed comparison

	PM1743 Gen5 SSD	PCIe Gen4 SSD*
Sequential read	14,000MB/s	6,700MB/s
Sequential write	6,000MB/s	4,100MB/s
Random read	2,500 K IOPS	1,000 K IOPS
Random write	280 K IOPS	180 K IOPS

Table 1: Improvement in read/writes as observed in PCIe Gen5 NVMe, specifically the PM1743 models against the *PM9A3 as PCIe Gen4 model.

In total, the aggregated read rate sums up to a striking 120GB/sec. This groundbreaking read speed can have profound implications for a hardware platform running VMware Greenplum. The drive's ability to quickly pull large

data sets from disk into memory liberates the CPU from the intensive task of sourcing data from storage. The result is a system that can operate more efficiently, with the CPU freed up to focus on computational tasks rather than data retrieval. This capability fundamentally enhances the system's effectiveness in dealing with massive data loads and complex analytical tasks.

Samsung SSD Value Kit (SVK)

Samsung PCIe Gen5 drives are complemented by the optional Samsung SSD Value Kit, which provides a set of tools to optimize SSD performance at the system level. For the proposed architecture here, SVK is used to optimize system tunings in an operating system around RAID and network path tuning, described later, to realize higher efficiency out of Samsung SSDs. Additional SVK features will include power optimizer and RAID performance optimizer tools.

Segment vs. coordinator machine architecture

The bare metal deployments of VMware Greenplum are characterized by two categories of hardware: coordinator machines and segment nodes. The deployment is structured such that two coordinator machines process the inbound workload from users, optimizing queries before distributing them to the segment nodes for execution. The number of segment nodes, also known as worker nodes, is contingent on both the volume of data that requires storage and the extent of processing necessitated by this data. Conversely, the architecture consistently encompasses two coordinator machines, regardless of the deployment's size or complexity.

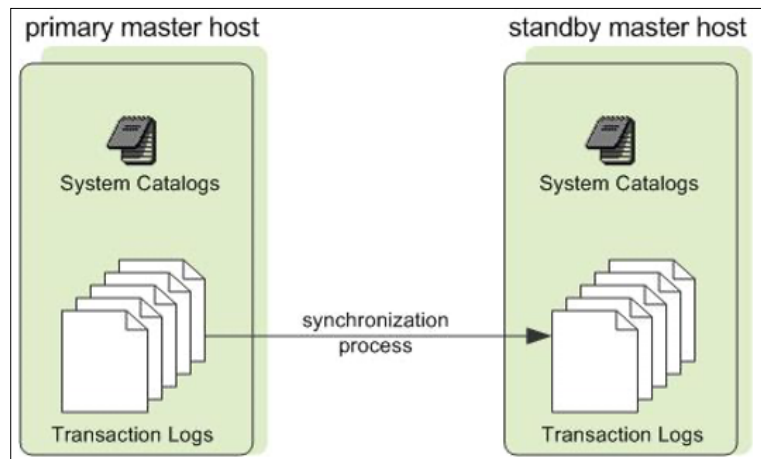


Figure 2: Dual coordinator architecture.

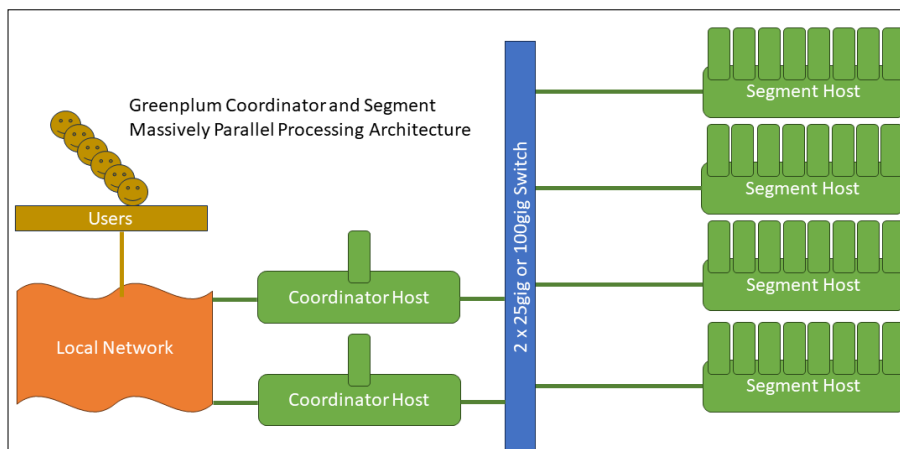


Figure 3: The Greenplum coordinator and segment architecture.

The hardware depicted in this document applies to both the segment and coordinator machines. However, a crucial distinction regarding the number and configuration of drives in each machine category must be emphasized.

Typical configurations include eight Samsung NVMe drives per machine—a necessity for segment nodes. However, the two coordinator machines demand a different setup. Each coordinator node requires merely two Samsung NVMe drives, configured with RAID 1. RAID 1, or disk mirroring, provides a number of advantages including data redundancy, which enhances data reliability and read performance. This redundancy ensures the availability of a backup, if one drive fails, ensuring the continued operation of the system.

This distinction in hardware configuration and data management strategy between coordinator machines and segment nodes is a vital aspect of the Greenplum architecture when procuring hardware.

Greenplum platform recommendations

The subsequent recommendations are proposed as beneficial adjustments specifically tailored to enhance the performance of the VMware Greenplum use case. Any more drastic alterations would lead to a deviation from the following reference architecture:

Key hardware components

CPU

The reference architecture has used AMD EPYC 9454P 48-Core Processor. As prices continue to decrease following the publication of this paper, the selected model in this reference architecture should be regarded as the base specification. Procuring processors with higher core counts is likely to improve performance on the VMware Greenplum concurrent workloads. A transition to a dual-socket system would constitute a deviation from this reference

architecture, as it would necessitate considerations related to non-uniform memory access (NUMA) performance.

Memory

This reference architecture includes 512GB of RAM per server, which represents a minimum recommended specification. Many VMware Greenplum users opt for servers with 768GB or 1024GB of RAM to afford more flexibility in allocating memory per query, particularly for high-concurrency analytics workloads.

Disk

In this architecture, eight Samsung PM1743 drives are utilized. The size of the drives, however, is not a critical factor influencing performance. The recommended design offers 30TB of raw disk capacity per server. Increasing the storage capacity can be easily achieved by replacing the current 3.84TB drives with larger ones. Drives with a capacity of 7.x or 15.xTB, for instance, would be suitable substitutions when dealing with storage-dependent workloads or the need to store large volumes of data.

Network

This architecture uses two Mellanox ConnectX-5 network cards, each with two ports, providing 400Gb aggregate bandwidth per server. The specific model of the network interface card is not fundamental to this reference architecture. Different brands can be used as substitutes, provided they offer the same bandwidth. VMware Greenplum standard practice advises for a minimum of two network interface ports. Some users choose to extend this to six ports for external connectivity of the VMware Greenplum nodes to other systems and networks. The use of 100Gb ports is recommended in this architecture, but not all enterprises support such networks. Thus, if necessary, 25Gb ports can be substituted with minimal risk. However, using ports slower than 25Gb would constitute a significant deviation from this reference design.

BIOS setup and settings

The BIOS setup and configuration play a crucial role in tuning the hardware platform for optimal performance. Throughout the testing phase of this architecture, a series of BIOS settings were altered to ensure that performance was maximized. The following changes were made:

Nodes per socket (nps) set to 2

This setting is found under the **advanced/acpi** section of the BIOS. Setting the number of nps to 2 optimizes the NUMA characteristics of the system, balancing core usage and memory locality to enhance the overall system performance.

[**Note:** This might not be needed and will be verified on future tests.]

Global C-state disabled

The global C-state, found under **advanced/cpu**, was disabled. The disabling of C-state has been a standard practice with VMware Greenplum since 2010. This change reduces the CPU's power-saving modes, keeping it ready for

immediate, high-performance computation. The resultant performance boost is particularly beneficial for heavy data processing tasks, typical in a VMware Greenplum workload.

Input-output memory management unit (IOMMU) disabled

The IOMMU, located under **advanced/nb**, was disabled. The IOMMU maps device-visible virtual addresses to physical addresses, which can add latency to I/O operations. By disabling this, the system can bypass this additional step, potentially improving I/O performance.

Max power level for performance

The power management setting was adjusted to the max power level for the performance setting, indicating that the system should prioritize performance over energy conservation. This is a crucial adjustment when maximum computational performance is required, as in the case of large-scale data processing and analytics.

[**Note:** While these BIOS settings were chosen to maximize performance during testing, they might not necessarily be universally recommended. Alternative configurations might be more appropriate depending on the specific system, workload and operational context. It is crucial that these settings be evaluated in the context of your specific organization.]

Linux tuning

Optimizing the Linux settings in accordance with the product's specifications can significantly influence the performance of VMware Greenplum. The configurations were adjusted as per the official VMware Greenplum recommendations, available in the [detailed guide](#).

Network setup

A comprehensive network setup is crucial to fully harness the capabilities of the system. To this end, we recommend assigning multiple IP addresses to each physical machine. The aim of this configuration is to distribute the VMware Greenplum segments evenly among the network interfaces on each physical machine.

This strategic distribution allows for optimal utilization of all available network interface cards and ports, which in turn are used by the multiple VMware Greenplum segments on each host. We can effectively manage network traffic by ensuring such an allocation, enhancing data transfer rates, and reducing potential bottlenecks.

This is not unique to this architecture, but rather, a common practice in VMware Greenplum deployments. For a deeper understanding of this practice and to guide you through the implementation process, refer to the following sections of the VMware Greenplum documentation: [gp_segment_configuration](#), [initialization of Greenplum](#), and [gpinitssystem](#) tool.

Number of Greenplum segments per machine

The parallelism in VMware Greenplum can primarily be modulated by three key aspects: the number of physical or virtual machines, the count of Greenplum primary segment instances per machine, and the permissible concurrency of queries as defined within the resource management groups of VMware Greenplum. The count of physical hosts within the system is largely dictated by the data storage requirements and the volume of CPU cores necessary for data processing. When designing the architecture, the count of physical hosts can typically be considered as a predetermined factor. Consequently, the count of primary segments and the permitted concurrency emerge as the primary variables in managing parallelism. Each long-running query within VMware Greenplum necessitates a minimum of one Linux process for every primary segment. In the case of complex queries, multiple processes may be generated for each primary segment.

In low-concurrency environments, a greater number of primary segments is suggested to achieve inherent concurrency. Conversely, in high-concurrency environments, a reduced number of primary segments is recommended, given the naturally high parallelism due to concurrent queries. Importantly, the count of primary segments should almost never exceed the number of CPU cores on each machine, to avoid creating unnecessary operating system scheduling overhead and inefficiencies. As such, architects are strongly encouraged to prioritize throughput as a primary measure of success rather than the level of concurrency. For instance, a target of 1 million queries per hour might be more useful than 10,000 concurrent queries.

Based on these considerations, we propose a recommended configuration for this reference architecture. There are three variations of the recommendation, each tailored to a different use case for the VMware Greenplum system:

- **Type 1** – High concurrency mixed workload multiple-purpose data system, with concurrency frequently exceeding 30 concurrent queries
- **Type 2** – Low concurrency large data processing, typical concurrency between 10 to 30 queries, focused on big data reporting and ETL
- **Type 3** – Ultra-low concurrency bulk load and reporting system, typically less than 1 to 5 concurrent queries conducting long data processing

Corresponding primary segment configuration recommendations are as follows:

- **Type 1** – 8 primary segments per physical host
- **Type 2** – 16 primary segments per physical host
- **Type 3** – 32 primary segments per physical host

We advise consultation with an experienced Greenplum architect for a review of your use case when making this decision.

Performance benchmark test suites

A TPC-DS derived test suite was selected as the core data warehousing benchmark used to evaluate the performance of this reference architecture. See the [benchmark script](#).

The TPC-DS test suite includes a diverse set of queries, and the data model comprises multiple tables with a wide range of sizes and level of normalization/denormalization. The benchmark is characterized by high CPU and I/O load, aiming to stress the system under test to its maximum capabilities.

In this paper, we consider three primary business outcomes, which are integral to evaluating the performance of the architecture:

- **Data loading time** – Given that VMware Greenplum is utilized for substantial data processing, the ability to load data rapidly with minimal resource utilization becomes a crucial performance metric for any reference architecture.
- **Non-concurrent data warehouse (DW) query batch completion time** – This measures the duration required to execute one complete round of all the queries in the benchmark suite. Primarily, this test assesses the hardware's data processing throughput.
- **Concurrent DW query batch completion time** – This evaluates the time needed to execute all queries in the benchmark multiple times, with each batch running concurrently, thereby inducing system concurrency. This test illustrates the system's scalability under increasing concurrent workload.

For the purposes of this paper, we employ a benchmark scale factor of 3,000 and 15 parallel streams to evaluate different hardware configurations comparatively.

System diagram of server used for testing

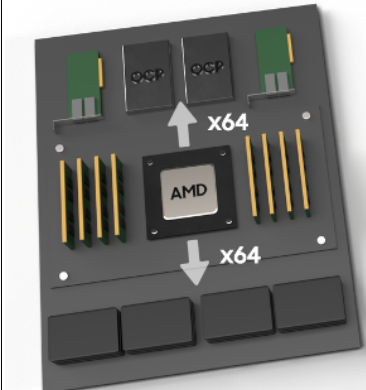
	Processor	Single AMD EPYC™ processor GENOA (AMD Socket SP5)
	Memory Capacity	24x DIMM slots, 2DPC ECC DDR5 designed for up to 4800MT/s
	Expansion	<ul style="list-style-type: none"> • 2x PCIe 5.0 x16 slots • 2x PCIe 5.0 x16 AIOM (OCP 3.0)
	Networking and I/O	<ul style="list-style-type: none"> • 1x RJ45 Dedicated IPMI LAN port • 1x VGA ports and 1x Serial port
	Drive Bays	<ul style="list-style-type: none"> • 16x E3.S SSD(x4) • 2x PCIe 3.0 NVMe 22110 M.2
	Power Supply and System Cooling	<ul style="list-style-type: none"> • 1600W Titanium redundant supply • 8x 40mm PWM fans
	System Management	Built-in Server management tool (IPMI 2.0, KVM/media over LAN) with dedicated LAN port
	Dimension	19" 1RU

Figure 4: Specifications for the SMC Type 3 server.

Description of previous generation hardware

The EMC Data Computing Appliance (DCA), an earlier-generation bare metal architecture for VMware Greenplum, is widely recognized in this domain. Initially released in 2010 by EMC, the DCA's support phase is nearing its conclusion around the time of this document's publication. Noteworthy contrasts between the DCA structure and the architecture delineated in this paper are as follows:

- Transition from 24 spinning disks to 8 Samsung NVMe drives, which substantially amplifies I/O throughput.
- Shift from a dual-socket CPU configuration to a single-socket CPU setup. This modification mitigates NUMA overhead and scheduling dilemmas arising from multi-CPU socket configurations.
- Adoption of a higher core count CPU design utilizing AMD's EPYC model, which facilitates parallel data processing on a solitary machine, yet at a reduced expense by leveraging a single socket.
- Escalation in network speeds from a standard of 10Gb to 100Gb, thereby enhancing data shuffling capabilities.
- Augmented RAM per host to support additional concurrent workloads on each physical machine.

In summary, the novel architecture is characterized by increased density, with a greater number of CPU cores and I/O per node compared to previous-generation frameworks; yet it manages to maintain a lower cost point by harnessing straightforward commodity server designs.

Performance benchmark test results

A comparison of synthetic I/O benchmarks reveals an outstanding performance improvement of 48x (or 4,800 percent) compared to the previous generation hardware.

Regarding data loading times, this architecture exhibits a remarkable 10.6x (or 1,060 percent) performance boost versus its predecessor. VMware Greenplum has consistently been [recognized for its robust data ingestion rates](#). With this architecture displaying up to a tenfold increase over previous hardware iterations, data loading speed truly emerges as a defining advantage.

The assessment of single-threaded query performance, which is primarily bottlenecked on the CPU, showcased up to a 5.7x (or 570 percent) performance enhancement over the previous generation architecture. It is anticipated that with larger datasets, where scan time plays a significant role in performance, this 570 percent improvement could significantly increase, reaching up to a maximum performance gain of 4,800 percent under ideal conditions.

Performance improvements

Test case	Previous generation Greenplum appliance	Previous PCIe Gen5 NVMe reference architecture	Times faster with new solution
Synthetic I/O	100%	4,800%	48.0x
Data loading	100%	1,060%	10.6x
Single-threaded DW benchmark	100%	570%	5.7x
Concurrent DW benchmark	100%	420%	4.2x

Table 2: A breakdown of performance improvements with new hardware.

In terms of the concurrency test, a performance improvement of up to 4.2x (or 420 percent) was observed with the new hardware. This figure could potentially be higher if an AMD EPYC processor with a greater core count was employed, therefore we recommend considering higher core counts when making purchase decisions. Additionally, we hypothesize that our test scenarios inadvertently favor the older hardware due to their predictable nature. Given the new hardware's vastly superior I/O processor capacity, it is likely to significantly outperform the old hardware in unpredictable workloads with random access. Hence, the 4.2x improvement per node represents a conservative estimate, with the actual expectation for real customer workloads potentially being substantially higher.

Architecture cost analysis

The proposed architecture is forecasted to serve as a cost-efficient solution for large-scale data processing applications, primarily due to the software's efficiency and the hardware's high-density design. It employs a single-socket CPU layout, which is a cost-effective, commodity architecture. Key drivers of cost for this architecture are primarily components like the number of disks, the amount of RAM, and the number of CPU sockets and cores. Other aspects, including network interfaces and miscellaneous components, do not significantly impact the cost.

Comparing the cost of this architecture with hardware from several years ago is not pertinent, as market conditions have evolved, and older hardware is no longer available for purchase. Therefore, to perform a meaningful cost analysis, we must compare this system's price with an alternative architecture for VMware Greenplum that encompasses.

Hardware specifications

	Recommended platform (PCIe Gen5)	Baseline platform (PCIe Gen4)
CPU	Single socket	Dual-socket (with same total number of cores)
Memory	512GB	512GB
Storage	8 NVMe PM1743 drives with value kit	24 SAS drives

Table 3: Cost-effective hardware specification of the recommended PCIe Gen5 platform, against the previous generation baseline of PCIe Gen4.

Utilizing current pricing data from a prominent server vendor, we discerned that our proposed new architecture is up to approximately **30 percent less expensive** than the alternative architecture mentioned above. This finding underscores that the value proposition of this novel architecture extends beyond performance to cost-effectiveness, particularly when compared to alternative bare metal architectures for VMware Greenplum software. Therefore, this new architecture not only can serve as a robust solution in terms of performance but can also provide a financially viable alternative to other architectural options.

Key benefits of the solution

This reference architecture can present numerous benefits for enterprises seeking robust, scalable and efficient solutions for their data warehousing, data mart, or analytics environments. Here's a comprehensive breakdown of these potential advantages:

- **Simplicity and cost-efficiency of single socket servers** – The use of single-socket servers can simplify support and reduce costs. These servers offer an easy path to scale horizontally to accommodate large and extremely large capacities. The single socket design can also reduce the complexity of task scheduling by the Linux operating system, as processes don't need to move between different processors within a single system. This characteristic leads to performance enhancements and increased system stability.
- **High concurrency with high core count CPUs** – The inclusion of high core count CPU architectures like AMD EPYC can facilitate high-concurrency query volumes in VMware Greenplum while maintaining a single CPU socket per server. This design feature supports improved performance and cost efficiency.
- **Efficient utilization of PCIe Gen5 Samsung NVMe drives** – These drives offer remarkable performance that can facilitate full CPU saturation of the workload, thus preventing disk I/O bottlenecks, with only eight drives per

server. Moreover, these drives come in various sizes at different price points, giving customers the flexibility to select the drive model based on their storage capacity needs. This versatility does not alter the architectural aspects of the reference architecture or the performance output of the system. This enables the future of Petascale SSD (with multiple SSDs) to provide even better efficiency for large-scale VMware Greenplum deployments.

- **Scalability of VMware Greenplum architecture** – The horizontal scalability of the VMware Greenplum architecture can support growing business needs. As an enterprise's data storage and computational requirements increase, they can easily scale out and add more machines to the infrastructure. This flexibility can not only satisfy growing demands but can also future proof the enterprise against emerging needs and potential expansions.

Investing in this reference architecture can bring several key benefits to an enterprise, particularly in terms of business intelligence, operational efficiency and financial optimization.

- **Improved decision-making** – The architecture is designed to handle high-concurrency query volumes efficiently. This means it can process large amounts of data quickly, enabling faster and more accurate data analysis. As a result, business decision-making can be significantly improved, making it easier to identify trends, anticipate changes and devise effective strategies.
- **Cost efficiency** – This architecture uses commodity hardware and single-socket servers, both of which are generally less expensive than their proprietary and multi-socket counterparts. Moreover, the architecture's scalability means that as your business grows, your infrastructure can grow with it cost-effectively, rather than requiring significant initial over-investment or disruptive upgrades down the line.
- **Operational efficiency** – With high-performing PCIe Gen5 Samsung NVMe drives with value kit, this architecture can handle heavy workloads without creating bottlenecks, maximizing the system's uptime and productivity.
- **Flexibility and scalability** – The architecture's reliance on commodity hardware avoids the need for proprietary infrastructure, leading to potentially significant cost reductions. This adaptability is essential for maintaining competitiveness in a rapidly evolving business landscape.
- **Future proofing** – Given the speed of technological advancement, any investment in IT infrastructure must consider future needs. The scalability and flexibility of this architecture, coupled with its focus on high-performance components like PCIe Gen5 Samsung NVMe drives with SVK, means it is designed to handle the data, system-level SSD optimizations and analytics challenges of the future, protecting your investment in the long term.

Overall, by adopting this reference architecture, enterprises can significantly enhance their business analytics capabilities, improve operational and cost efficiencies, and ensure they have a robust, scalable infrastructure ready to meet the business needs of today and tomorrow.

Future work: performance optimization for RAID

We analyzed RAID features as part of SVK for VMware Greenplum workload with the proposed architecture. We used RAID-5 in our benchmarks since this provides the lowest cost RAID solution where only 1/8th of the disk capacity will be sacrificed for RAID redundancy. The RAID-5 volume was then partitioned to eight separate smaller volumes, which enabled the benchmark to view these as just another set of disks. After running the TPC-DS benchmark and collecting disk statistics with no RAID setup, it was apparent that the benchmark was generating a large I/O where the average size was over 260KB. Since we have eight SSDs in our proposed architecture configuration, it was decided to use a 32KB strip or stripe unit size for RAID. With eight disk RAID-5, one full stripe of data would be $32K \times 7 = 224KB$, so the 260KB will be able to fill at least one full stripe. This should help to reduce RAID read-modify-write cycles.

Using the above setup, we ran the TPC-DS benchmark but found that with RAID-5 setup was causing significant data inflation, where about 3x more data was being READ and WRITTEN to disks. Further analysis of this issue showed that this data inflation was caused by RAID-5 partitions where the partition itself was not stripe aligned when it was created. And this caused RAID software to perform large amounts of read-modify-write cycles instead of stripe writes. RAID-5 will cause many more writes and reads to the disks but with this alignment we observed less than 1.5x in our test, which is within expectations.

Other RAID optimizations included merge writes where multiple WRITE operations that are sequential can be merged to reduce RAID overhead as well as merging reads. Write merge seems to help somewhat but was not significant in our benchmark tests. Furthermore, we tested the RAID in degraded mode where one of the disks was forced into failure mode. Even with degraded mode, the performance was not significantly impacted. The major difference in degraded mode is that there will be many more Reads/Writes due to reconstructing of data that must take place whenever failed disk data is accessed. Additional I/O overhead didn't seem to impact performance significantly, but further analysis will be done to better understand this scenario. One small benefit in degraded mode is that whenever there is a stripe write, one less disk is being written.

Other optimizations that were tried included separating XFS journal device as a separate disk device. Basically, for each of the eight XFS file systems, we allocated a separate disk for journaling. This also seems to improve benchmark performance but again was not significant. Since XFS file system meta data sizes are small, there might be some benefit in separating this out into RAID-1 volumes created by allocating a small partition from each of the eight SSDs. Again, this will require more detailed analysis and better measure of its benefits.

In conclusion, this paper at this point is not recommending RAID on this configuration, but likely future modifications will recommend how to optimally configure RAID for use with VMware Greenplum on this hardware configuration.

Learn more

To know more about VMware 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

This white paper focuses on an innovative architecture that seamlessly unites high-performance hardware components, efficient software platforms, and design principles founded on extensive field experience. By leveraging the unprecedented speed and capacity of Samsung's PCIe Gen5 NVMe drives with optional SVK to further increase SSD efficiency at system level, along with the scalable and flexible nature of VMware Greenplum, we assert that we have engineered a formidable solution to cater to businesses facing substantial data processing and analytics challenges.

It is essential to recognize that this architecture was conceived with an emphasis on offering maximum cost performance and the best total cost of ownership for data processing and storage. The blend of computational power, high-speed storage, and efficient software not only enables peak performance but also presents a cost-effective and operationally efficient option for businesses.

Our conviction in this architecture's effectiveness for businesses aiming to modernize their data systems for analytics and data warehousing is steadfast. While each organization has unique needs and circumstances, we assert that this architecture's inherent flexibility allows it to be tailored to meet specific requirements without detracting from its core efficiency and performance features.

In summary, this white paper introduces a robust, scalable architecture that offers a potent framework for businesses looking to modernize their data systems. With a fine balance of adaptability, computational power and strategic design, this reference architecture potentially offers an optimal cost-performance ratio, effectively redefining how enterprises can approach and handle their data processing needs.

Appendix

Selection of VMware Greenplum software version – The version certified on this architecture is Greenplum 6.24. Nevertheless, the authors of this paper strongly recommend using the latest patch release of Greenplum version 6 because of the continuous evolution and improvement of Greenplum. VMware routinely releases updates to its software that often include critical bug fixes, performance enhancements, security patches and new features. By opting for the latest patch release, you ensure that your system benefits from these advancements and maintains its security and reliability. This approach could also preempt potential issues that have been identified and rectified in more recent patches. It is important to note that the shift from the certified Greenplum 6.24 to the most recent patch release is not expected to result in any significant performance degradation. VMware conducts extensive testing and validation to ensure that any updates do not adversely affect the performance of its product. However, it's always recommended to conduct a thorough review and testing process in your environment when applying new patches, to ensure compatibility and performance with your specific use cases and workload.

