QxStack/QxVDI-OA/HC High-Density Optimized SKU_ Reference Architecture

Version 2.0

CONTENT

TABLES

FIGURES

1. Executive Summary

In the fast-changing business world, the role of IT department is getting more and more complex. With diverse business models and services provided by enterprises, IT departments are devoted to providing a reliable and elastic production-level environment with prominent performance and sufficient storage resources to support diverse workloads. Security is also a significant concern. On top of that, all these requirements need to be fulfilled within limited budgets.

In order to assist customers and partners to accelerate data center transformation, Quanta Cloud Technology (QCT), a global data center solution provider, provides a ready-to-use software-defined data center solution-QxStack/QxVDI-OA/HC High-Density Optimized SKU with following benefits:

- Ultra-performance and sufficient storage capacity: gain ultra-performance and sufficient storage resource in one appliance for diverse scenarios.
- Accelerated time to value: minimize the time required to deploy new infrastructures.
- Reliability: provide a confident choice with VMware vSAN ReadyNode™ certification.
- Economical total cost: reduce TCO by high-density design with a shared energy system.

In this document QCT validates different use cases, including Virtual Desktop Infrastructure (VDI) with OA (office application) and HC (high computing) application as well as mixed workload scenarios, proving the feasibility and validated-overall performance of the solution architectures.

With the above-mentioned benefits and the validated use cases, QCT's QxStack/QxVDI-OA/HC High-Density Optimized SKU is a valid choice for partners and customers to construct a software-defined data center and stay in a leading position.

2. Introduction

2.1. Purpose

To introduce QxStack/QxVDI-OA/HC High-Density Optimized SKU and validate the solution's outstanding performance and availability in multiple use cases like VDI and mixed-workload environment.

2.2. Scope

- Introduce overall structure of QxStack/QxVDI-OA/HC High-Density Optimized SKU and the benefits of this solution.
- Illustrates the hardware configuration and software stack discreetly selected by QCT in the solution.
- Simulates different use cases in the data center including VDI and IaaS with mixed workloads, and demonstrates the ultra-performance and scalability of QxStack/QxVDI-OA/HC High-Density Optimized SKU.

2.3. Audience

The intended audiences of this document are IT professionals, technical architects, and sales engineers who would like to replace traditional desktop and adopt software-defined infrastructure to build VDI or IaaS environment.

3. Solution Overview

3.1. Manageability, Scalability, and Efficiency- Hyper-Converged Infrastructure

QxStack/QxVDI-OA/HC High-Density Optimized SKU is a hyper-converged infrastructure solution. Traditionally, IT technicians face the challenges of resource management and scalability since compute and storage resources are separated. Hyper-Converged Infrastructure (HCI) is a novel technology which can integrate compute, storage, and virtualization resources in a single hardware box. Every single node is capable of delivering compute and storage resources at the same time. Several benefits are listed in detail below.

Simplified Management

In legacy-converged architecture, since compute and storage resources are provided by different servers and storage devices, configuration settings and operation management are independent, meaning IT administrators must manage two devices through different management tools. To provide a storage device to a server, administrators need to configure settings from LUN and Volume and then mount the storage device to the server host for VM to access. In hyper-converged architecture, compute and storage can be regarded as a system. Administrators can manage both compute and storage resources with a single management portal. By achieving full “policy-driven management”, IT administrators only need to define their own compute and storage resources. The allocation process can be automatically completed by a single management portal, which significantly reduces the management effort.

Scalability and Efficiency

The hyper-converged infrastructure integrates compute and storage resources into a basic unit, called building block. By implementing the clustered architecture, administrators can add more building blocks to the cluster to expand the overall performance and capacity. This also makes the expansion of the hyper-converged architecture simple and predictable, as shown in Figure 1.

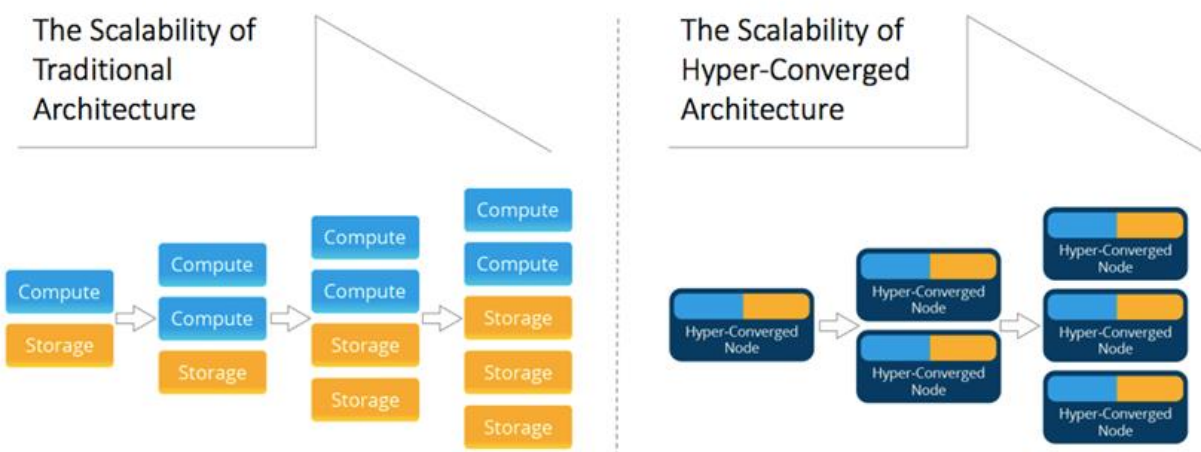

Figure 1. Comparison between Traditional Architecture and HCI.

3.2. Compute Capability and Storage Capacity

The server used for QxStack/QxVDI-HC High-Density Optimized SKU is the QuantaPlex T42S-2U, which is a 2U 4-node server. Compared to the general 1U 1-node server with the same storage capacity, QuantaPlex T42S-2U provides 50% more compute power. Besides, QxStack/QxVDI-OA/HC High-Density Optimized SKU adopts the latest technology - Intel® Xeon® Scalable Processors to unleash its computing performance to a new level. According to QCT's testing results, Intel® Xeon® Scalable Processors perform better than the previous CPU product family in IOPS and throughput under hyper-converged infrastructure. With the high performance and sufficient storage capacity, QxStack/QxVDI-OA/HC High-Density Optimized SKU can support customer's business more efficiently.

3.3. Accelerated Time to Value

Building up a data center can be a tough and time-consuming challenge. To select the best-fit system customers have to identify the demands of their companies first and strive to understand the pros and cons of each system structure. After that, customers still need to search for suitable components and concern about the compatibility, not to mention the complicated deployment process. In the lengthy process, customers need to put lots of efforts and time which may result in the IT department losing the focus on core business. In view of these factors, QCT provides a pre-configured and pre-validated total solution with software-installed for customers to accelerate time to value.

Pre-configured Hardware

To ensure hardware compatibility, QCT discretely selected the components and passed the strict testing process from system integration to electromagnetic interference to thermal testing. VMware also validates the feasibility of these components to make sure the compatibility between the components and VMware's software. The server's features and disk group ratio best practices design suggested by VMware are considered to enhance the performance of the software stack. The configuration design is proven to provide outstanding performance and seamless compatibility.

Pre-loaded Software

In the factory, QCT will install the software and further validate the system. Software pre-loaded allows the customer to save time to deployment, speed up data center establishment, and utilize a higher quality solution.

3.4. Reliability - Pre-Validated Solution

vSAN ReadyNode™ is a program created by VMware® to verify the compatibility between server platforms and VMware®-developed software and to guarantee the performance and stability of a solution. To pass vSAN ReadyNode™ certification, all the details of a solution including hardware components, firmware and driver, and software stack should be strictly examined to meet the rigorous requirements. QCT put a lot of effort towards certification validation of QxStack/QxVDI-OA/HC High-Density Optimized SKU. In the past, it was common for administrators to spend weeks researching and struggling with compatibility issues to deploy a new system. Now, with the solution validated by QCT and VMware®, customers can rest assured of the solution reliability and focus on strategic and productive tasks.

3.5. Economical Total Cost

Economical Server Investment- High Density Design

Compared with general 1U 1-node server, QuantaPlex T42S-2U saves 50% of space consumed in a data center and doubles compute density that can maximize the productivity per square foot in customers' data center. Regular 1U 1-node server needs at least three servers to build up a VMware vSAN™ cluster with high availability (HA). QxStack/QxVDI-OA/-HC High-Density Optimized SKU can build up a vSAN™ cluster with HA in one single server, which simplifies the complexity of cabling, conserves space, and minimizes overall effort in data center management.

Energy Saving- Shared Energy System

Commonly, four 1U 1-node servers use eight power supplies in total and consume 800 watts for each power supply. QuantaPlex T42S-2U is a 2U 4-node server, which shares only two power supplies and consumes 1600 watt for each power supply. Compared with four 1U 1-node servers, QuantaPlex T42S-2U server saves up to 50% energy consumption.

With the ultra-dense and efficient power consumption design, QxStack/QxVDI-OA/HC High-Density Optimized SKU can better boost resource utilization and save the total investment in a data center.

4. Solution Architecture

The four-node server and VMware software are included in the solution and possible host failure is taken into consideration to fulfill HA. The overall solution architecture is shown in Fig. 2 below.

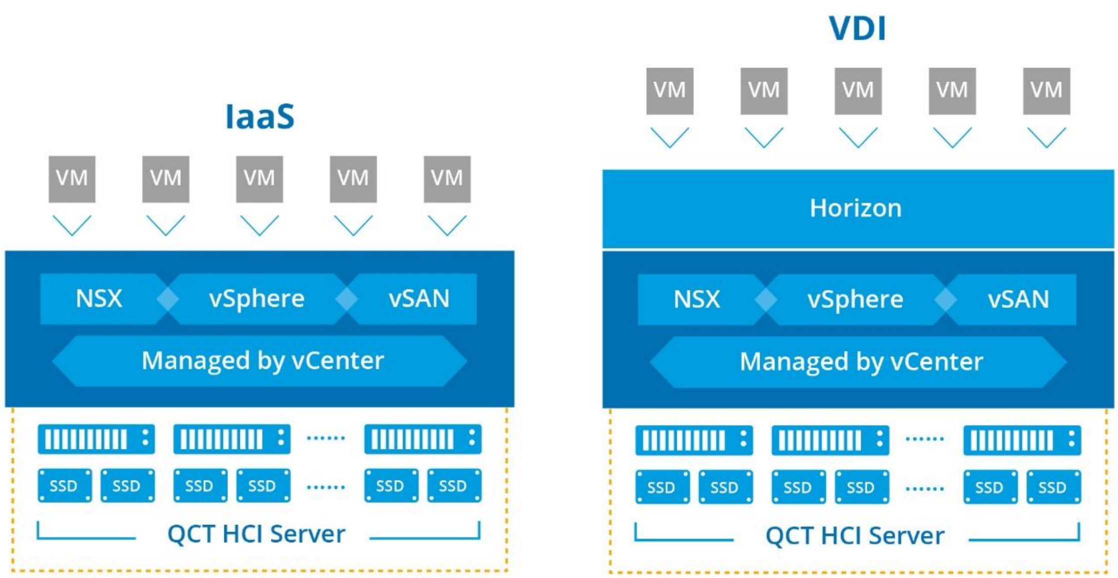

Figure 2. QxStack/QxVDI-HC High-Density Optimized SKU.

4.1. Hardware configuration

In QCT QxStack/QxVDI-OA/HC High-Density Optimized SKU, the server QuantaPlex T42S-2U, a 2U 4-node server, is used to build the infrastructure. The server is capable of carrying 2 CPUs in Intel® Xeon® Processor family per node, 6 SATA/SAS disks per node, and up to 512GB DDR4 RDIMM memory per node. An extra M.2 riser equipped on the MB can support both SATA and PCIe M.2 devices for OS boot. Some important features of QuantaPlex T42S-2U are listed below:

- High floating-point operation per second with Intel Xeon Processor and great memory bandwidth up to 2666MHz provide outstanding performance.
- Ultra-dense design with 4 nodes in just 2-unit height provides prominent space and energy utilization.
- Extra M.2 riser on the motherboard supports both SATA and PCIe M.2 devices, and provides better boot performance for installed OS.

The components in this high-density SKU are carefully evaluated and selected, as show in Table 1:

.

Table 1. Hardware Configuration.

Role	Component	Total Quantity
Server Model Name	QuantaPlex T42S-2U (2U, 4Nodes)	1
CPU	Intel Xeon Gold 5220 2.2GHz	8
Memory	32GB DDR4-2666	32
Storage of SSD	2.5” SATA 960GB	4
Storage of HDD	2.5” SAS 1.8TB	20
Storage Controller	SAS 3008A mezzanine	4
Network Interface Card	Quanta 82599 dual port 10Gb, SFP+	4
Boot Device	M.2 SSD 240GB	4

4.2. Software Configuration

The software stack utilized in this solution includes the VMware vSphere®, VMware vSAN, and VMware Horizon®. The adopted versions are shown in Table 2.

Table 2. Software Configuration.

Software and Service	Version
VMware vSphere®	6.7 enterprise plus Edition™
ESXi™ hypervisor	6.7 U3
VMware vSAN	6.7.0
VMware Horizon®	7.10 enterprise
vCenter Server® Appliance	6.7 U3

VMware vSphere®

VMware vSphere® is the leading virtualization software for cloud foundation. The primary features of the VMware vSphere® are:

- Server virtualization: VMware ESXi™ hypervisor virtualizes server’s hardware into manageable resources.
- Centralized management: vCenter Server®, a virtual appliance, provides central management platform across ESXi™ hosts.
- Friendly UI: vSphere web client, a web management user interface, enables administrators to easily manage vCenter Server® or ESXi™ hosts.
- Easy VM migration: the function vMotion® in VMware vSphere enables VM migration between hosts which is vital for server redundancy.

- Auto resource balance: Distributed Resource Scheduler (DRS) can dynamically balance the shared computing resources for VMs within a cluster.
- Host redundancy: High availability (HA) function monitors hosts within a cluster to deal with the host failure. It migrates VMs to other available hosts when host failure occurs.
- Virtualized switch for VMs: Virtual Switch (VS) creates virtualized network on each ESXi™ host and provides the network to each virtual machine. Virtual Distributed Switch (VDS), which is similar to virtual switch, further aggregates NICs, VMkernals, and portgroups together for a data center. The virtualized switch thus simplifies the management on server network.

VMware vSAN™

vSAN™ is the software-defined storage which can be utilized to:

- aggregate local hosts' storage devices into the shared data store and provide access across hosts within the same cluster.
- provide the hyper-converged infrastructure with simple management and provision.
- provide vSAN™ storage policies to define availability factors such as failure to tolerate.
- integrate with vSphere hypervisor layer which means 100% compatibility.
- enable high scalability that can be either scale-out or scale-up on demands quickly.

VMware Horizon®

VMware Horizon® is the key software component of VDI with the combination of vSphere® and Horizon®. VDI brings an easy method to manage user desktop pools and apps environment with higher security and efficient resource utilization. VMware Horizon® delivers a centralized virtual desktop management platform, different forms of virtual desktop deployment, a simple-updated desktop image for desktop groups, and simple security control of virtual desktops. Meanwhile, it effectively minimizes the management time and cost, and brings end users virtual desktops across devices and geo-locations with full utilization.

VMware Horizon® 7 enterprise edition is the version adopted in the solution and the related components/services utilized are shown in Table 3 below.

Table 3. VMware Horizon® Software Features.

Name	Description
Horizon® Connection	Connection server acts as a bridge for client connections. Connection Server authenticates users through Microsoft Active Directory and directs the request to the appropriate VM, physical or blade PC, or Windows Terminal Services server. Connection server provides management capabilities such as enabling SSO, authenticating client users, entitling client users to specific desktops or pools, creating links between users and desktops, etc. A connection server is used in the solution as the VDI management platform.
Horizon® Client	This client software which accesses remote desktops and applications can run on a tablet, a phone, a notebook, etc. After logging in, users are authorized to use a list of remote desktops and apps. Permission may require Active Directory credentials, a UPN, a smart card PIN, an RSA SecureID, or an authentication token. In the solution, the View Clients are installed and used to log in to the virtual desktop services.
Active Directory	The Active Directory is developed from Microsoft for the Windows system domain networks. The Horizon® needs Active Directory infrastructure to perform the user validation and management. One Active Directory server is used in the solution to manage the VDI topology.
Horizon® Agent	Acting as a bridge between Horizon® and VMs' guest OS, the Agent installed on the source parent VMs is used for the communication between client and virtual machines.

VMware NSX®

VMware NSX® is a software-defined network that delivers scalable and flexible network architecture in the vSphere®. In contrast to traditional server networking, VMware NSX® reproduces Layer 2 to Layer 7 network model in the software form to simplify network management. Its logical functions can be created on demand any time to provide the flexibility of server infrastructure on either security management or security control. In addition, its programmable API feature is also critical to achieve data center automation.

VMware NSX® consists of several main components, as shown in Table 4. VMware NSX® 6 enterprise edition is the version adopted in this solution.

Table 4. VMware NSX® Software Features.

Name	Description
NSX® Manager™	NSX® Manager™ is the management plane and centralized network management component of NSX®, installed as a virtual appliance in the vCenter Server®. It determines the core configuration of the whole system.
NSX® Controller™	NSX® Controller™ is the control plane that controls virtual networks and overlay tunnels' information such as logical switches and logical routers. NSX® Controller™ is a central control point for all logical switches. It processes the information of all virtual machines, hosts, and logical switches. The Controller™ supports two control plane modes, namely, Unicast and Hybrid. These modes make NSX® decouple from the physical network.
NSX® Edge™ Service Gateway	The Edge™ Service Gateway (ESG) is a logical router which provides north-south traffic for a data center. It consists of built-in services such as routing, load balancer, DHCP, parameter firewall, NAT, SSL VPN, etc.
NSX® Distributed Logical Router	NSX® Distributed Logical Router (DLR) is able to provide east-west distributed routing with tenant IP address space and data path isolation between switches. The key function of DLR allows VMs or workloads that reside on the same host with different subnets to communicate with each other without having to traverse a traditional routing interface; thus, the traffic is optimized. DLR also supports both dynamic routing and static routing, and provides the functions such as L2 bridging, DHCP relay, etc.
NSX® Distributed Firewall	NSX® Distributed Firewall (DFW) is a hypervisor kernel-embedded firewall that provides visibility and control for virtualized workloads and networks. L2 to L4 access control policies can be created based on VMware vCenter® objects such as data centers, clusters, virtual machine names, and tags, and network construction such as IP/VLAN/VXLAN addresses as well as user group identity from the Microsoft Active Directory. The nature of this firewall can automatically extend firewall capacity when additional hosts are added to the data center.

5. Solution Use Case

QxStack/QxVDI-OA/HC High-Density Optimized SKU is the solution which can be provided for several common use cases as follows in an enterprise environment.

5.1. Virtual desktops services provisioning

Nowadays, enterprises invest considerable resource and expense in providing personal computers with diverse OS and applications for employees to fulfill different working demands. Due to the inevitable demand of devices, software management, workstation maintenance, and troubleshooting are thus getting more complicated and time-consuming for administrators. To tackle these difficulties, QxVDI-OA/HC High-Density Optimized SKU can deliver up to 300 Windows desktops with high-efficiency application and desktop image management in a single T42S-2U server.

5.2. Mixed workloads and stable storage performance

Enterprises are concerned about the quality of stable services from a data center to provide end users great experience since different VM workloads such as web services, database, and e-commerce service are commonly executed in a data center. The standard to evaluate the infrastructure's performance is thus regarded as the most important index for enterprises. QCT's QxStack/QxVDI-OA/HC High-Density Optimized SKU can load up and execute mixed workloads, and provide vSAN™ shared datastore in a data center.

5.3. Micro-segmentation on services

Since cyber-attacks issues are complicated and may utilize multiple methods and malware agents to infiltrate a data center, traditional firewalls only provide limited protection for the cloud environment. This limitation is primarily due to a traditional firewall’s effectiveness for on-premises environments only. To overcome this, QxStack/QxVDI-OA/HC High-Density Optimized SKU offers an integrated security model for the overall data center. By using this SKU, flexible security policies can span across all the hosts and services easily that isolate and segment every object wherever it moves and regardless of its network topology.

6. Solution Validation

6.1. Virtual Desktop Infrastructure- View Planner

Tests are conducted to validate the VDI capacity and performance of our solution using "View Planner". View Planner is a VMware designed virtual desktop benchmark tool that analyses VDI environment's performance and desktop experience. The benchmark mechanism is controlled by the "View Planner harness" and it manages the targeted desktops which are installed with "View Planner agent" to launch specific workloads and test. View Planner defines "work profile" to categorize which applications are selected for testing and defines "run profile" to categorize selected benchmark/infrastructure settings to perform test. The test will be executed in either local or remote mode, the former generates workloads directly on target desktops while the latter utilizes additional desktops to launch sessions and generate workloads on target desktops. Eventually, the View Planner will collect the latency data of the desktops and produce report.

6.1.1. Virtual Desktop Infrastructure (VDI) Test: Overview

The View Planner offers local, remote and passive remote mode for benchmarking. The passive remote mode, which maps client desktops and targeted desktops in one-to-multiple mapping, is adopted to perform the benchmarking. Primarily, View Planner calculates the "QoS latency groups" of CPU and storage, and "ratio of actual to expected operations", which is the ratio of a workload being executed successfully. After the test completes, the quality of service, ratio of actual to expected operations and performance will be collected, and list on the test report. A compliant test result must accomplish the following conditions: QoS latency groups under respective threshold and more than 0.9 of the ratio of actual to expected operations. In verdict, reaching these conditions meaning that the system is compliant with VMware's standard which provides users a reliable virtual desktop experience.

The goal of the tests is to showcase a formal performance reference from the approximately saturated system for various scenarios therefore the tests are performed to reach its capability of theoretical upper limit. QCT chose two work profiles for two scenarios respectively, one is customized profile for light-weight Office Application environment while the other is a view planner predefined standard profile for typical High Compute intensive environment. Several view planner workloads are involved in the test, which are MS Office applications, Adobe Reader, Chrome browser and Windows Media Player. The two scenarios' related work profile name and respective workloads from view planner are listed in table 5:

Table 5. View Planner workprofile and workloads

Scenario	View Planner workprofile	View Planner workloads
Office Application	TestProfile_Office	vp_Adobe, vp_msExcel, vp_msOutlook, vp_msPowerPoint, vp_msWord
High Compute	standardTestProfile_chrome	vp_Adobe, vp_Chrome, vp_ChromeWebAlbum, vp_msExcel, vp_msOutlook, vp_msPowerPoint, vp_msWord, vp_winMediaPlayer

In the test environment, a T41S-2U server is utilized as client server to load all the service VMs and client desktop pool. On the T42S-2U server, only the target desktop pool is deployed as test target. When the View Planner starts the benchmark, it controls the agents that install on both client and target desktops to execute tasks. The client desktops will launch the Horizon Client application and link to the target desktops in which applications will be launched by agent to perform automatic stress test.

To ensure that the desktop experience and system performance are acceptable when system is maximum resource use is soon reached, QCT monitored several parameters from the view planner and categorized them with specific thresholds as success criteria in table 6:

Table 6. View planner parameters threshold.

Parameter	threshold
CPU sensitive	<1
Storage sensitive	<6
Ratio of actual to expected operations	>0.9
Discarded desktop count	<1%
Memory Usage of any one of the test hosts	<90%
Memory Balloon of any one of the test hosts	=0

6.1.2. Virtual Desktop Infrastructure (VDI) Test: Technical Configuration

On the client server, vCenter is deployed to provide central control of the data center, the Horizon Connection service is installed on a Windows Server 2016 to provide instant-clone desktops for all the tests. The Active directory and domain name service are installed on a second Windows Server 2016 to provide integrated domain user service for VDI. The DHCP service is installed on a third Windows Server 2016 providing IP addresses for the desktop pool of VMs. A virtual desktop pool is deployed and acts as client to launch the Horizon Client's sessions.

In the test cluster, a virtual desktop pool is deployed to be test target in which the workloads are being executed.

The overall topology of the test environment is shown in figure 3 below and the details of service VMs that are located on server T41S-2U is listed in table below.

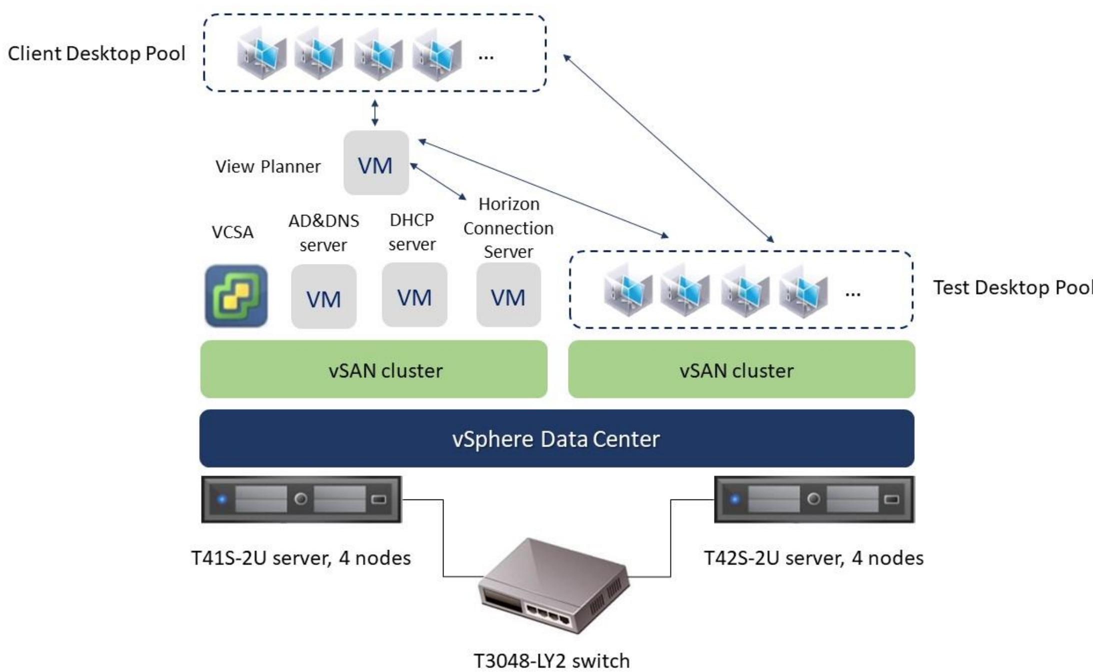

Figure 3. VDI Testing Topology.

Table 7. virtual server spec.

Server	role	amount	vCPU	vRAM	VM Disk
vCenter server	Data center management	1	4	16GB	32GB
Windows server 2016	Active Directory, DNS	1	2	2 GB	
Windows server 2016	DHCP	1	2	2 GB	
Windows server 2016	Horizon Connection service	1	4	10 GB	
View planner harness	View Planner central control	1	8	8 GB	

There are two Windows 10 x64 virtual desktop configurations selected for the VDI tests. A 300 VMs desktop pool is deployed and its primary virtual hardware setting is 2 vCPU, 2GB vRAM, one 32GB disk and a VMXNET3 vNIC. Another 200 VMs desktop pool is deployed and its primary virtual hardware setting is 2 vCPU, 4GB vRAM, one 32GB disk and a VMXNET3 vNIC. The two scenarios and the respective test configurations of VM are listed in table 8.

Table 8. VM Profiles for OA and HC.

Scenario	OS	Desktop Amount	vCPU	vRAM	VM disk size	vNIC
Office Automation	Windows 10 x64	300	2	2GB	32GB	1
High Compute		200		4GB		1

In the view planner tests, the "ramp up time" was set to 600 and 460 seconds for the custom and standard test respectively, to which the view planner waits for a random amount of time between 0 and this specific number before each iteration being executed in multiple VMs, and this time setting reduces system' CPU loading. Both the two tests are set with 5 seconds think time, 5 iterations, and BLAST display protocol. As listed in table 9 below:

Table 9. View Planner configurations.

Scenario	Office Application	High Computing
Test workload Profile	TestProfile_Office	standardTestProfile_chrome
Number of VMs	300	200
Ramp Up time	600	460
Think Time	5	
Number of iterations	5	
Display Potocol	BLAST	
Type of Desktop	VDI	
Remote sessions per client	3	

6.1.3. Virtual Desktop Infrastructure (VDI) Test: Result

The report revealed the statistics of test including configurations and performance. The key parameters are the "Group A" and "Group B" latencies chart, which represent the CPU and storage sensitive in the virtual desktops respectively, and the Ratio of actual to expected operations. Both the tests are completed with no desktop sessions failed and with compliant result therefore the system is stable for acceptable user experience.

The test result summary of the two test profiles are listed in table 10 which included the group A and group B value of CPU and storage sensitivity, ratio of actual to expected operations and discarded desktop count.

Table 10. VDI Test Results Statistic

Test Profile	OA	Standard
CPU Sensitive	0.7859 Sec	0.7759 Sec
Storage Sensitive	4.7526 Sec	5.5912 Sec
Ratio of actual to expected operations	0.99	1.0
Discarded desktop count	0	0

According to the test result for scenario "OA", the run completed with group A latency 0.7859 second, group B latency 4.7526 second, the Ratio of actual to expected operations reached 0.99 and the discarded desktop count is 0, meanwhile, the memory usage and memory balloon of the four server nodes are recorded in view planner as well. In verdict, the CPU and storage latencies are within the threshold, the workloads' successful execution rate is 99%, no desktops were failed to test and the memory usage and balloon are controlled within the threshold. All the outcome matched the success criteria and details are shown in figure 4 – 7 below.

2.1 Workload Profile

Work Profile Name	TestProfile_Office

2.2 Run Profile

Run Profile Name	passive_300vm_office
Number Of Vms	300
Ramp Up Time	600
Think Time	5
Number Of Iterations	5
Discarded Desktop Count	0
Display Protocol	blast
Type Of Desktop	vdi

Figure 4. Scenario: OA-profile.

3. View Planner Score

Test Name	nov20_300vm_office
Test Mode	passive
Latency Data Mode	local
Test Start Time	2019-11-20 03:01:59
Test End Time	2019-11-20 04:54:56
Test Status	Completed

3.1 Workgroup - worker: QoS Summary

Latency Group	95th Percentile Latency in Seconds (Lower is Better)	Threshold in Seconds
Group A (CPU Sensitive)	0.7859	<= 1
Group B (Storage Sensitive)	4.7526	<= 6

Figure 1. 95th Percentile latency

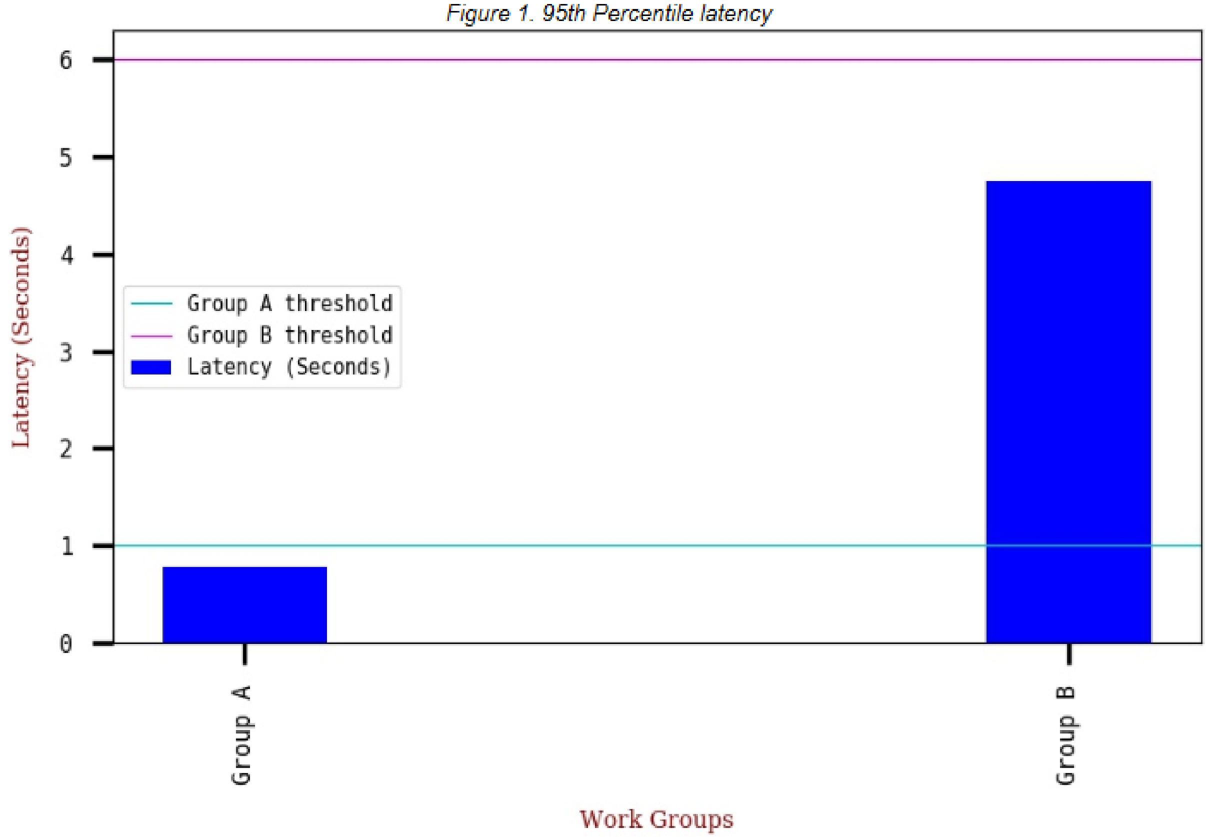

Figure 5. Scenario: OA-performance.

4. Operation Details

4.1 Workgroup - worker: Executed Vs Expected Operation Ratio

Ratio of Actual to Expected Operations (0.0 to 1.0)	0.99

4.2 Workgroup - worker: Application Response Time

Table 1. Workload Event Latency Statistics

Operation	Group	Executed / Expected Count	Mean	Median	Variance
excel_close	Group A	899/900	0.0881	0.0892	0.0003
excel_data_entry	Group A	4495/4500	0.1173	0.1039	0.0039
excel_maximize	Group A	899/900	0.4916	0.4538	0.0085
excel_minimize	Group A	899/900	0.3004	0.2981	0.0006
excel_open	Group B	899/900	1.6276	1.5102	0.2009
excel_saveas	Group B	899/900	0.3902	0.3782	0.0054
excel_sort	Group A	4495/4500	0.185	0.1732	0.0098
outlook_close	Group A	856/900	0.3621	0.3508	0.0021
outlook_open	Group B	860/900	0.7096	0.6873	0.0091
outlook_read	Group B	4280/4500	0.566	0.4969	0.0692
outlook_restore	Group B	856/900	0.13	0.1049	0.0281
pdf_browse	Group A	18000/18000	0.0859	0.081	0.0003
pdf_close	Group A	900/900	0.6637	0.5623	0.0281
pdf_maximize	Group A	900/900	0.9095	0.9031	0.0029
pdf_minimize	Group A	900/900	0.6386	0.5525	0.0167
pdf_open	Group B	900/900	1.1714	1.0837	0.1052
pptx_append_slide	Group A	3600/3600	0.2343	0.0988	0.111
pptx_close	Group A	900/900	0.2181	0.21	0.0016
pptx_maximize	Group A	900/900	0.2576	0.2545	0.0001
pptx_minimize	Group A	900/900	0.2943	0.2893	0.0012
pptx_modify_slide	Group A	3600/3600	0.1509	0.1409	0.0049
pptx_open	Group B	900/900	4.3286	4.2605	0.3619
pptx_run_slide	Group A	9000/9000	0.0322	0.0136	0.0012
pptx_saveas	Group None	900/900	4.6465	4.6629	0.1508
word_close	Group A	897/900	0.8052	0.8153	0.2654
word_maximize	Group A	900/900	0.909	0.9026	0.0105
word_minimize	Group A	900/900	0.5815	0.5557	0.0065
word_modify	Group A	4485/4500	0.0768	0.0698	0.001
word_open	Group B	900/900	4.6665	4.6692	0.5587
word_save	Group B	897/900	3.2489	3.1922	0.086

Figure 6. Scenario: OA-operation details.

Resource Usage of 10.0.0.5

Measurement	Average	Min	Max
CPU Usage (percent)	46.69	5.84	100.0
CPU Core Utilization (percent)	65.08	14.85	100.0
CPU Utilization(percent)	43.97	8.95	100.0
Memory Usage (percent)	71.27	71.22	71.36
Active Memory (GB)	84.61	31.59	116.38
Consumed Memory (GB)	190.32	190.18	190.55
Memory Balloon (GB)	0.0	0.0	0.0
Network Receive Rate (KBPS)	29198.31	2051	136346
Network Transmit Rate (KBPS)	27258.79	1902	117013

Resource Usage of 10.0.0.6

Measurement	Average	Min	Max
CPU Usage (percent)	45.4	5.15	100.0
CPU Core Utilization (percent)	63.76	13.33	99.77
CPU Utilization(percent)	42.11	7.97	92.86
Memory Usage (percent)	69.79	69.75	69.88
Active Memory (GB)	82.56	30.8	114.49
Consumed Memory (GB)	186.36	186.26	186.6
Memory Balloon (GB)	0.0	0.0	0.0
Network Receive Rate (KBPS)	27606.41	1982	124317
Network Transmit Rate (KBPS)	30290.51	1972	114167

Resource Usage of 10.0.0.7

Measurement	Average	Min	Max
CPU Usage (percent)	45.75	5.01	100.0
CPU Core Utilization (percent)	63.87	12.95	100.0
CPU Utilization(percent)	42.33	7.81	99.79
Memory Usage (percent)	70.47	70.44	70.53
Active Memory (GB)	83.83	31.37	114.77
Consumed Memory (GB)	188.19	188.1	188.32
Memory Balloon (GB)	0.0	0.0	0.0
Network Receive Rate (KBPS)	37321.01	2167	169669
Network Transmit Rate (KBPS)	34236.03	2083	156020

Resource Usage of 10.0.0.8

Measurement	Average	Min	Max
CPU Usage (percent)	45.79	5.51	100.0
CPU Core Utilization (percent)	64.0	14.36	100.0
CPU Utilization(percent)	42.77	8.67	100.0
Memory Usage (percent)	70.56	70.5	70.65
Active Memory (GB)	83.82	30.8	115.37
Consumed Memory (GB)	188.41	188.27	188.65
Memory Balloon (GB)	0.0	0.0	0.0
Network Receive Rate (KBPS)	28462.97	2078	151885
Network Transmit Rate (KBPS)	30354.88	1885	143040

Figure 7: Scenario: OA- resource usage.

According to the test result for scenario "HC", the run completed with group A latency 0.7759 second, group B latency 5.5912 second, the Ratio of actual to expected operations reached 1.0 and the discarded desktop count is 0, meanwhile, the memory usage and memory balloon of the four server nodes are recorded in view planner as well. In conclusion, the CPU and storage latencies are within the threshold, the workloads' successful execution rate is 100%, no desktops failed and the memory usage and balloon are controlled within the threshold. All the outcome matched the success criteria and details are shown in figure 8-11 below.

2.1 Workload Profile

Work Profile Name	standardTestProfile_chrome

2.2 Run Profile

Run Profile Name	passive_200vm_chrome
Number Of Vms	200
Ramp Up Time	460
Think Time	5
Number Of Iterations	5
Discarded Desktop Count	0
Display Protocol	blast
Type Of Desktop	vdi

Figure 8. Scenario: HC-profile.

3.1 Workgroup - benchmark: QoS Summary

Latency Group	95th Percentile Latency in Seconds (Lower is Better)	Threshold in Seconds
Group A (CPU Sensitive)	0.7759	<= 1
Group B (Storage Sensitive)	5.5912	<= 6

Figure 1. 95th Percentile latency

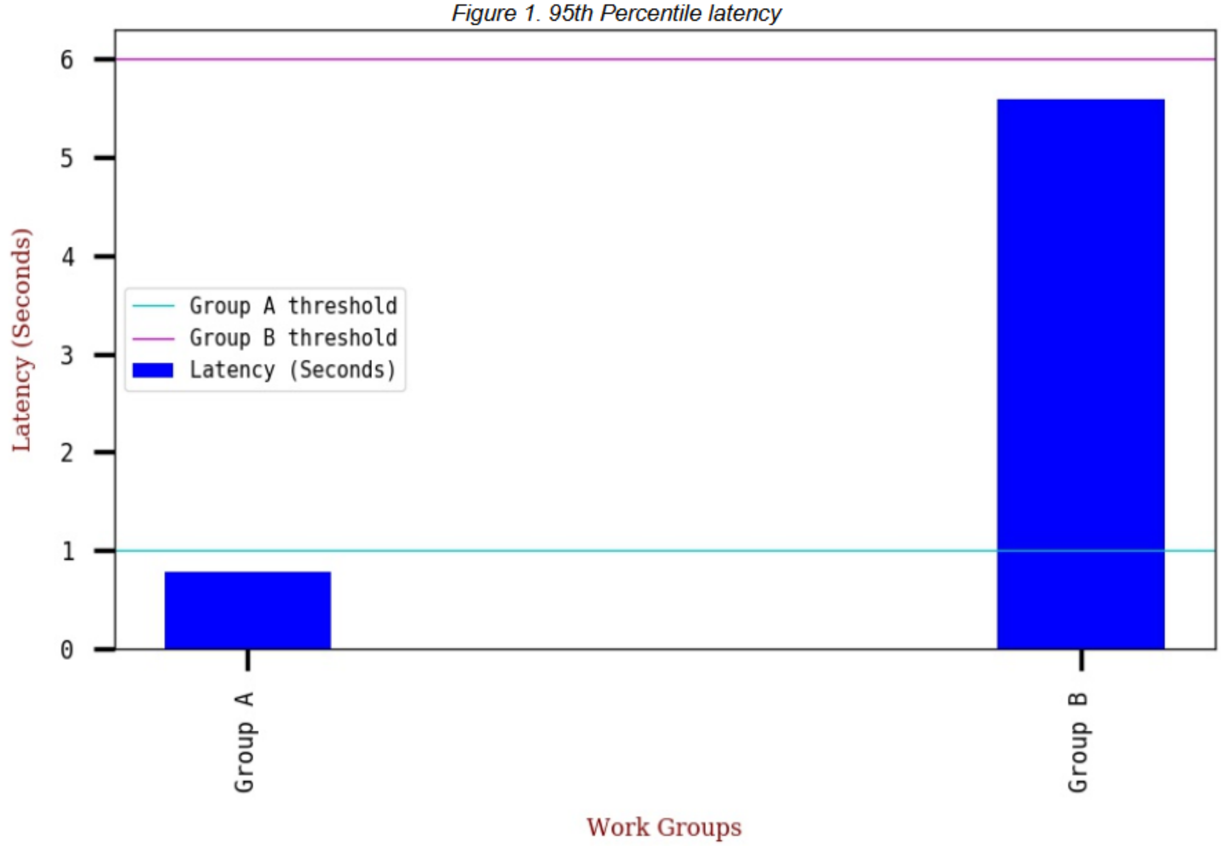

Figure 9. Scenario: HC-performance.

4. Operation Details

4.1 Workgroup - benchmark: Executed Vs Expected Operation Ratio

Ratio of Actual to Expected Operations (0.0 to 1.0)	1.0

4.2 Workgroup - benchmark: Application Response Time

Table 1. Workload Event Latency Statistics

Operation	Group	Executed / Expected Count	Mean	Median	Variance
chrome_close	Group B	600/600	2.4928	2.2365	0.2417
chrome_navigateApacheDoc	Group A	16200/16200	0.5752	0.5257	0.0323
chrome_open	Group B	600/600	5.5555	5.5661	0.1012
chrome_webalbum_close	Group B	600/600	2.4731	2.2283	0.2262
chrome_webalbum_navigate	Group A	8400/8400	0.2968	0.2738	0.0042
chrome_webalbum_open	Group B	600/600	5.5638	5.5743	0.0903
excel_close	Group A	599/600	0.0797	0.0795	0.0002
excel_data_entry	Group A	2995/3000	0.1074	0.0968	0.0023
excel_maximize	Group A	599/600	0.4703	0.4385	0.0075
excel_minimize	Group A	599/600	0.293	0.2909	0.0004
excel_open	Group B	599/600	1.4078	1.3901	0.0407
excel_saveas	Group B	599/600	0.3439	0.337	0.0051
excel_sort	Group A	2995/3000	0.1598	0.1551	0.0047
outlook_close	Group A	597/600	0.3503	0.3428	0.0012
outlook_open	Group B	597/600	0.6252	0.6176	0.0021
outlook_read	Group B	2985/3000	0.54	0.4619	0.0567
outlook_restore	Group B	597/600	0.1154	0.0957	0.013
pdf_browse	Group A	12000/12000	0.0837	0.0791	0.0003
pdf_close	Group A	600/600	0.668	0.5508	0.0328
pdf_maximize	Group A	600/600	0.8789	0.876	0.0018
pdf_minimize	Group A	600/600	0.5876	0.5285	0.0122
pdf_open	Group B	600/600	0.9676	0.9497	0.0312
pptx_append_slide	Group A	2388/2400	0.2285	0.097	0.101
pptx_close	Group A	597/600	0.2038	0.1981	0.001
pptx_maximize	Group A	597/600	0.2571	0.2543	0.0001
pptx_minimize	Group A	597/600	0.2898	0.2839	0.0013
pptx_modify_slide	Group A	2388/2400	0.1337	0.1253	0.0023
pptx_open	Group B	597/600	3.9184	3.9251	0.1165
pptx_run_slide	Group A	5970/6000	0.0303	0.013	0.001
pptx_saveas	Group None	597/600	4.3785	4.4465	0.0967
video_close	Group A	600/600	0.2589	0.256	0.0001
video_open_and_play	Group C	600/600	0.5229	0.5193	0.0061
word_close	Group A	600/600	0.6672	0.6237	0.1988
word_maximize	Group A	600/600	0.8894	0.8882	0.0084
word_minimize	Group A	600/600	0.5744	0.5486	0.0078
word_modify	Group A	3000/3000	0.0694	0.0617	0.0012
word_open	Group B	600/600	4.1087	4.067	0.2818
word_save	Group B	600/600	2.9187	2.9548	0.0745

Figure 10. Scenario: HC-operation details.

Resource Usage of 10.0.0.5

Measurement	Average	Min	Max
CPU Usage (percent)	40.93	4.58	100.0
CPU Core Utilization (percent)	57.95	11.32	98.7
CPU Utilization(percent)	34.92	6.46	72.39
Memory Usage (percent)	88.94	54.43	89.83
Active Memory (GB)	90.15	49.22	168.35
Consumed Memory (GB)	237.5	145.35	239.88
Memory Balloon (GB)	0.0	0.0	0.0
Network Receive Rate (KBPS)	13975.98	2608	83203
Network Transmit Rate (KBPS)	13270.01	2386	70623

Resource Usage of 10.0.0.6

Measurement	Average	Min	Max
CPU Usage (percent)	41.19	4.86	100.0
CPU Core Utilization (percent)	58.31	11.92	100.0
CPU Utilization(percent)	35.05	6.8	75.54
Memory Usage (percent)	89.39	56.0	89.87
Active Memory (GB)	89.77	46.87	169.15
Consumed Memory (GB)	238.7	149.55	239.97
Memory Balloon (GB)	0.0	0.0	0.0
Network Receive Rate (KBPS)	14383.91	2441	79526
Network Transmit Rate (KBPS)	15296.3	2433	66028

Resource Usage of 10.0.0.7

Measurement	Average	Min	Max
CPU Usage (percent)	40.13	4.56	100.0
CPU Core Utilization (percent)	57.36	11.53	99.55
CPU Utilization(percent)	34.81	6.5	84.66
Memory Usage (percent)	89.54	56.34	89.87
Active Memory (GB)	90.09	45.96	173.72
Consumed Memory (GB)	239.11	150.45	239.98
Memory Balloon (GB)	0.0	0.0	0.0
Network Receive Rate (KBPS)	17087.14	2301	84621
Network Transmit Rate (KBPS)	16470.44	2305	76982

Resource Usage of 10.0.0.8

Measurement	Average	Min	Max
CPU Usage (percent)	40.48	5.01	100.0
CPU Core Utilization (percent)	57.72	12.38	100.0
CPU Utilization(percent)	34.91	7.13	75.86
Memory Usage (percent)	89.34	63.67	89.94
Active Memory (GB)	89.47	48.45	170.97
Consumed Memory (GB)	238.55	170.02	240.17
Memory Balloon (GB)	0.0	0.0	0.0
Network Receive Rate (KBPS)	12970.19	2186	66764
Network Transmit Rate (KBPS)	15072.62	2190	80349

Figure 11. Scenario: HC-resource usage.

According to the test results, the following conclusions are determined:

- The test results matched QCT's success criteria and are compliant to view planner's requirement therefore the overall system provides a certified quality of service for virtual desktop environment.
- The actual upper limit of the two test cases were not reached therefore the system can provide more desktops in the two scenarios, however, the tradeoff would be the performance of system.
- The test results are comparable between servers and vendors because of the standardized view planner scoring methodology.

6.2. Infrastructure as a service- VMmark®

To validate the overall performance on the solution, the benchmark tool, VMware VMmark®, is adopted.

6.2.1. Test Overview

VMmark® is a tile-based benchmark tool developed by VMware for vendors to measure performance, power consumption, and scalability of virtualization platforms. It creates reliable performance score of the virtualized data center and provides the comparison between tested systems with industrial standard.

The target server, also called system under test (SUT), is tested and loaded up with tile workloads. VMmark® executes a set of virtual machines with diverse workloads including web simulation, e-commerce simulation, and standby system, and integrates application-level benchmarking software, the Weathervane, to place workload of realistic applications. These apps implement real-time auction website and simulate users' practical behaviors. DVDstore benchmark, an online e-commerce test application with database, simulates clients' web server login and catalog's browse using basic queries. Totally, the VMmark® simulation test includes 20 VMs in a so called "tile", and each tile requires a client VM to generate workloads upon the tile. The test procedures include the following items, as listed below. Generally, the VMmark® Prime Client appliance controls the tile VMs to perform app test and migration in the infrastructure, as shown in Fig. 12.

- Scalable web workload simulation.
- E-commerce transaction simulation.
- VM cloning and deployment.
- Dynamic VM relocation between servers.
- Dynamic VM relocation across storages.
- Simultaneous server and storage virtual machine relocation.
- Automated load balancing.

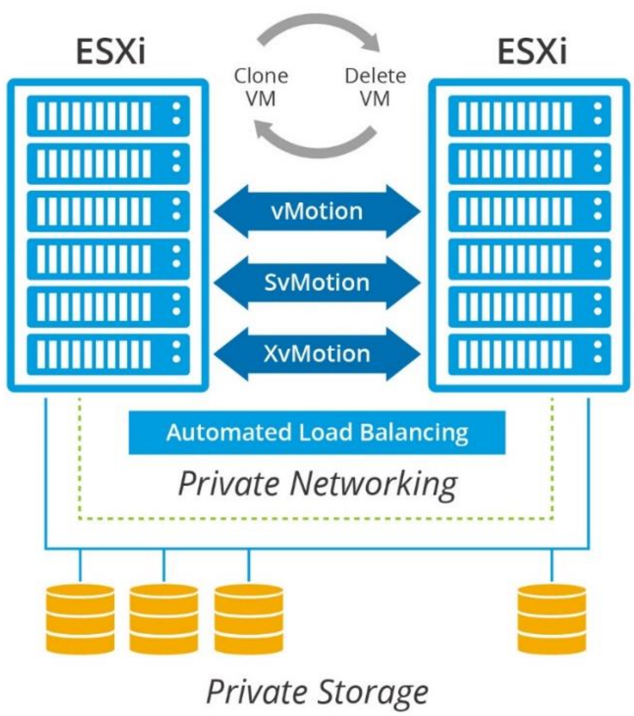

Figure 12. VMmark® Test Procedure.

6.2.2. Test Configuration

The SUT server's hardware is listed in table 1 (page 7). Each VMmark® tile requires 47 vCPUs, 166GB memory, and 891GB storage space. The test topology of VMmark® benckmark for our solution including one SUT - T42S-2U server, one T41S-2U server to load Client VMs, and one Windows iSCSI server for providing additional shared storage. All these servers are connected to one physical switch.

The vCenter Server® VM, VMmark® Prime Client, and Client VMs are located on T41S-2U server rather than SUT to avoid any resource consumption that affect the tile VMs. The tile VMs are deployed and operated only on SUT, as shown in Fig. 13.

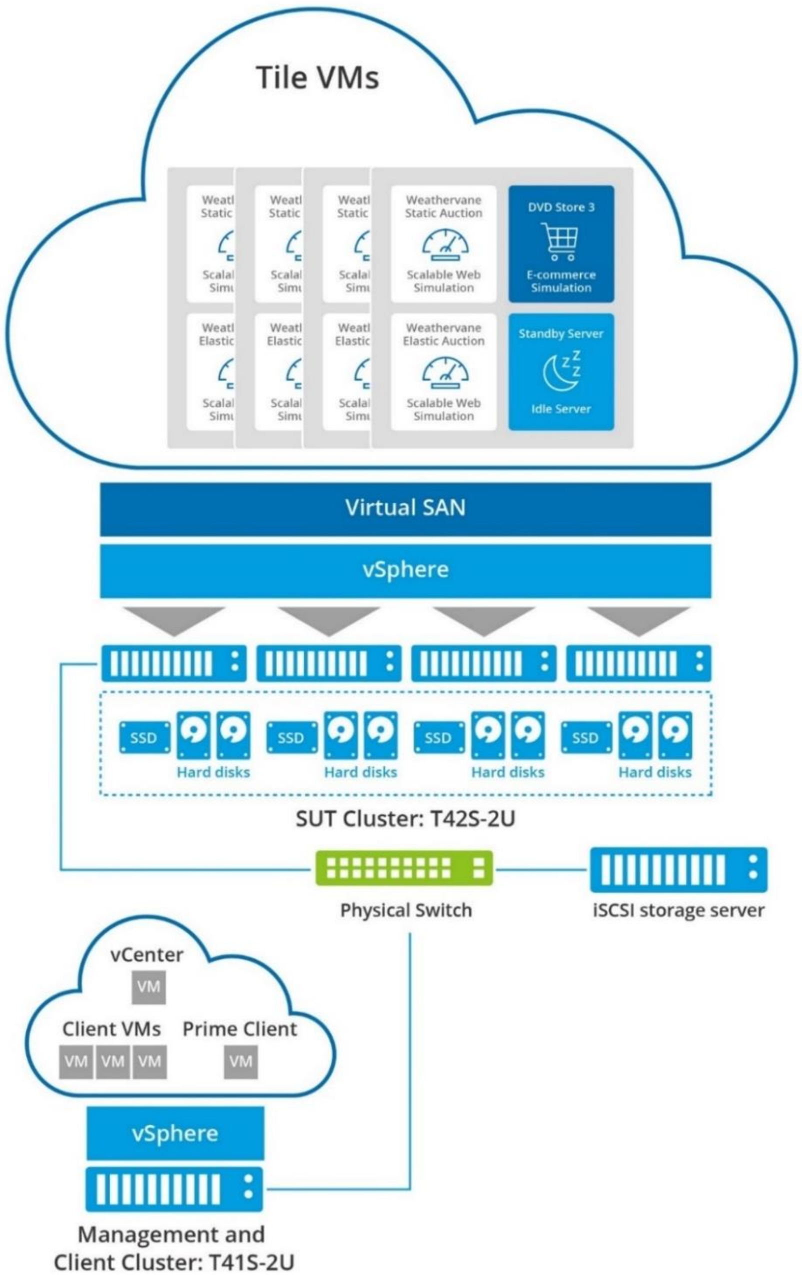

Figure 13. VMmark® Test Topology.

6.2.3. Test Result

After careful planning and repeated benchmarking, a 6-tile test was completed eventually which the maximum capability was reached in the system. According to the results, the overall system maintains good quality of service when executing an intensive stress test of 6-tiles. The test result has been approved by VMware and published, and the screenshots of the test result is shown in figure 14 below. For the full test result, please refer to [QCT VMmark® test report]() published on VMware VMmark® website.

VMmark® 3.1 Results

Server Vendor & Model: QuantaPlex T42S-2U Storage Vendor & Model: QuantaPlex T42S-2U Hypervisor: VMware ESXi 6.7 U3 (Build 14320388) Datacenter Management Software: VMware vCenter Server 6.7 U3 (Build 14367737)	VMmark 3.1 Score = 5.23 @ 6 Tiles

Number of Hosts: 4	Uniform Hosts [yes/no]: yes	Total sockets/cores/threads in test: 8/144/288
Tested By: Quanta Cloud Technology		Test Date: 12-09-2019
Performance Section Performance	Configuration Section Configuration	Notes Section Notes for Workload

Performance

	weathervane			weathervaneE			dvdstoreA			dvdstoreB			dvdstoreC			
TILE_0	Actual	Ratio	QoS(nRT|MaxPctF)	Actual	Ratio	QoS(nRT|MaxPctF)	Actual	Ratio	QoS(ms)	Actual	Ratio	QoS(ms)	Actual	Ratio	QoS(ms)	GM
p0	3538.43	0.98	0.95 | 0.13	562.62	0.98	0.96 | 0.61	756.12	1.03	1292.05	526.35	1.05	1467.92	372.98	1.08	1572.96	1.02
p1	3546.51	0.99	0.82 | 0.01	560.89	0.98	0.75 | 0.41	755.77	1.03	1267.18	550.95	1.10	1428.49	411.23	1.19	1527.76	1.05
p2	3531.69	0.98	0.88 | 0.11	556.77	0.97	0.90 | 0.55	730.55	0.99	1379.02	514.10	1.03	1528.07	344.48	0.99	1653.63	0.99
TILE_1	Actual	Ratio	QoS(nRT|MaxPctF)	Actual	Ratio	QoS(nRT|MaxPctF)	Actual	Ratio	QoS(ms)	Actual	Ratio	QoS(ms)	Actual	Ratio	QoS(ms)	GM
p0	3552.53	0.99	0.78 | 0.03	563.78	0.99	0.77 | 0.37	693.27	0.94	1553.64	491.80	0.98	1764.77	341.52	0.98	1943.67	0.98
p1	3559.90	0.99	0.78 | 0.01	559.45	0.98	1.03 | 0.62	737.15	1.00	1411.48	500.05	1.00	1649.11	361.18	1.04	1805.83	1.00
p2	3543.07	0.98	0.91 | 0.18	556.71	0.97	0.67 | 0.39	710.58	0.97	1496.46	490.95	0.98	1683.62	337.45	0.97	1870.77	0.98
TILE_2	Actual	Ratio	QoS(nRT|MaxPctF)	Actual	Ratio	QoS(nRT|MaxPctF)	Actual	Ratio	QoS(ms)	Actual	Ratio	QoS(ms)	Actual	Ratio	QoS(ms)	GM
p0	3545.73	0.99	0.84 | 0.03	566.73	0.99	1.04 | 0.74	763.08	1.04	1287.27	549.73	1.10	1450.47	404.20	1.17	1585.80	1.05
p1	3557.20	0.99	0.96 | 0.33	560.95	0.98	0.70 | 0.35	759.08	1.03	1281.28	527.40	1.05	1470.21	349.45	1.01	1602.54	1.01
p2	3540.91	0.98	0.93 | 0.05	558.30	0.98	0.82 | 0.40	756.92	1.03	1295.41	543.67	1.09	1455.44	407.18	1.17	1549.29	1.05
TILE_3	Actual	Ratio	QoS(nRT|MaxPctF)	Actual	Ratio	QoS(nRT|MaxPctF)	Actual	Ratio	QoS(ms)	Actual	Ratio	QoS(ms)	Actual	Ratio	QoS(ms)	GM
p0	3540.44	0.98	0.80 | 0.11	566.94	0.99	0.99 | 0.66	726.35	0.99	1365.25	512.38	1.02	1532.99	376.88	1.09	1633.85	1.01
p1	3554.70	0.99	0.73 | 0.00	565.03	0.99	0.51 | 0.42	768.55	1.05	1231.73	539.05	1.08	1400.55	373.82	1.08	1532.27	1.03
p2	3533.51	0.98	0.74 | 0.01	564.74	0.99	1.02 | 0.63	739.17	1.01	1324.19	544.42	1.09	1441.25	404.95	1.17	1540.76	1.04
TILE_4	Actual	Ratio	QoS(nRT|MaxPctF)	Actual	Ratio	QoS(nRT|MaxPctF)	Actual	Ratio	QoS(ms)	Actual	Ratio	QoS(ms)	Actual	Ratio	QoS(ms)	GM
p0	3536.41	0.98	1.14 | 0.24	563.70	0.99	0.91 | 0.58	768.15	1.05	1283.99	529.67	1.06	1480.48	360.50	1.04	1676.57	1.02
p1	3532.94	0.98	0.81 | 0.01	563.44	0.98	0.71 | 0.35	755.02	1.03	1290.90	540.77	1.08	1472.11	385.00	1.11	1581.02	1.04
p2	3460.45	0.96	0.89 | 0.16	558.94	0.98	0.85 | 0.56	762.67	1.04	1282.88	526.23	1.05	1471.87	379.45	1.09	1646.61	1.02
TILE_5	Actual	Ratio	QoS(nRT|MaxPctF)	Actual	Ratio	QoS(nRT|MaxPctF)	Actual	Ratio	QoS(ms)	Actual	Ratio	QoS(ms)	Actual	Ratio	QoS(ms)	GM
p0	3558.03	0.99	1.03 | 0.25	564.46	0.99	0.95 | 0.58	772.08	1.05	1210.89	550.25	1.10	1329.74	381.70	1.10	1451.01	1.04
p1	3541.78	0.98	0.77 | 0.01	559.54	0.98	0.98 | 0.57	781.15	1.06	1176.91	572.42	1.14	1315.63	436.05	1.26	1358.12	1.08
p2	3515.94	0.98	1.07 | 0.21	557.35	0.97	0.61 | 0.42	794.77	1.08	1149.47	561.23	1.12	1291.01	377.98	1.09	1396.87	1.05
p0_score:	6.13															
p1_score:	6.22															
p2_score:	6.13															

Infrastructure_Operations_Scores:	vMotion	SVMotion	XVMotion	Deploy
Completed_Ops_PerHour	52.00	32.00	24.00	12.00
Avg_Seconds_To_Complete	15.94	204.72	260.51	583.91
Failures	0.00	0.00	0.00	0.00
Ratio	2.00	1.78	1.33	1.50
Number_Of_Threads	2	2	2	2

Summary	Run_Is_Compliant	Turbo_Setting:0
	Number_Of_Compliance_Issues(0)*	Median_Phase(p0)
Unreviewed_VMmark3_Applications_Score	6.13	
Unreviewed_VMmark3_Infrastructure_Score	1.63	
Unreviewed_VMmark3_Score	5.23	

Figure 14. VMmark® test result (VMmark® is a product of VMware, Inc.)

To understand the cluster resources' utilization, the SUT cluster's statistics are recorded via vCenter server. In the overall test phase, the average utilization of CPU is 33.161%, memory is 71.25%. Spiking did not appear in vSAN performance chart, which indicates stability. As shown in figure 15-17 below.

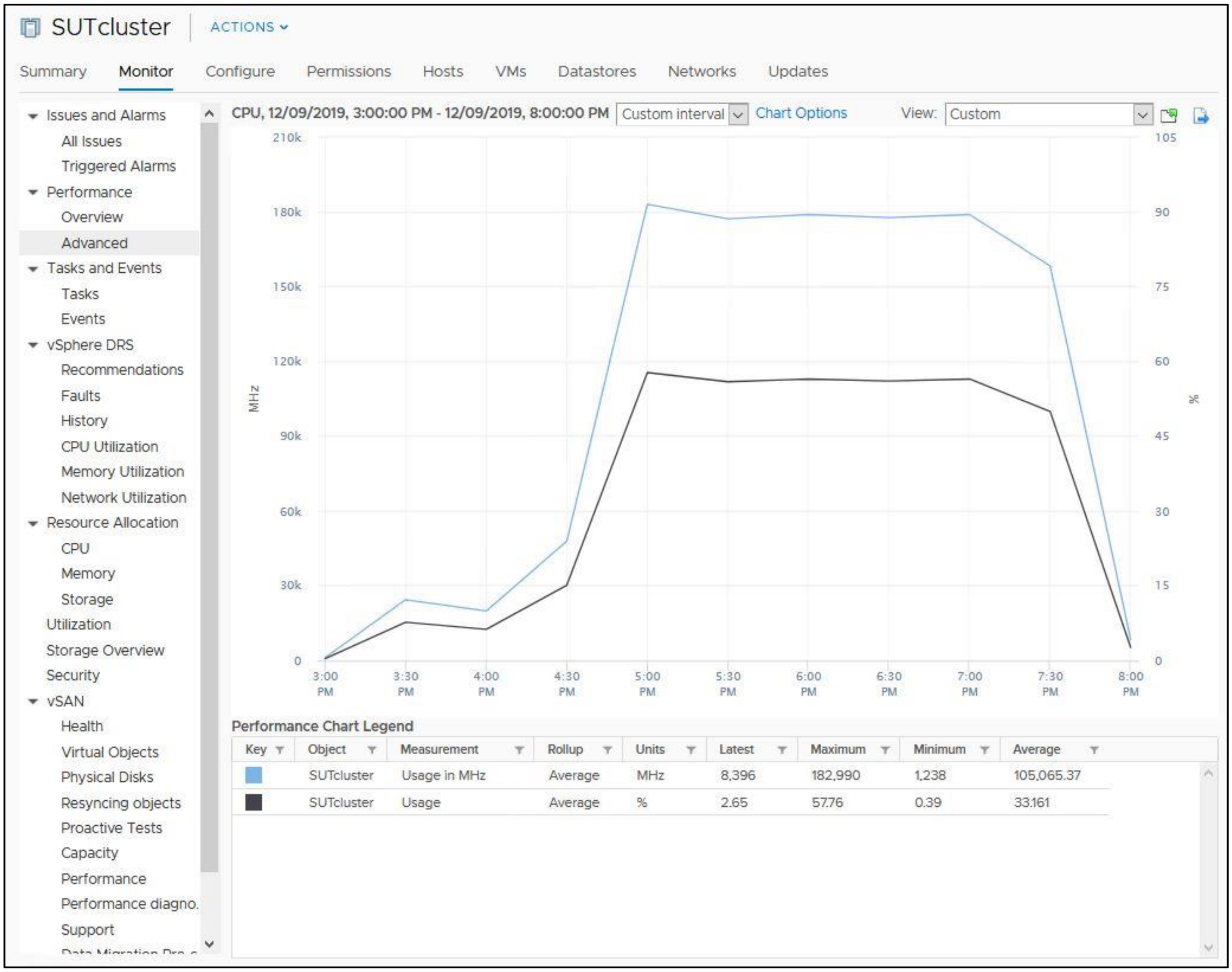

Figure 15. SUT clusters statistics (1)

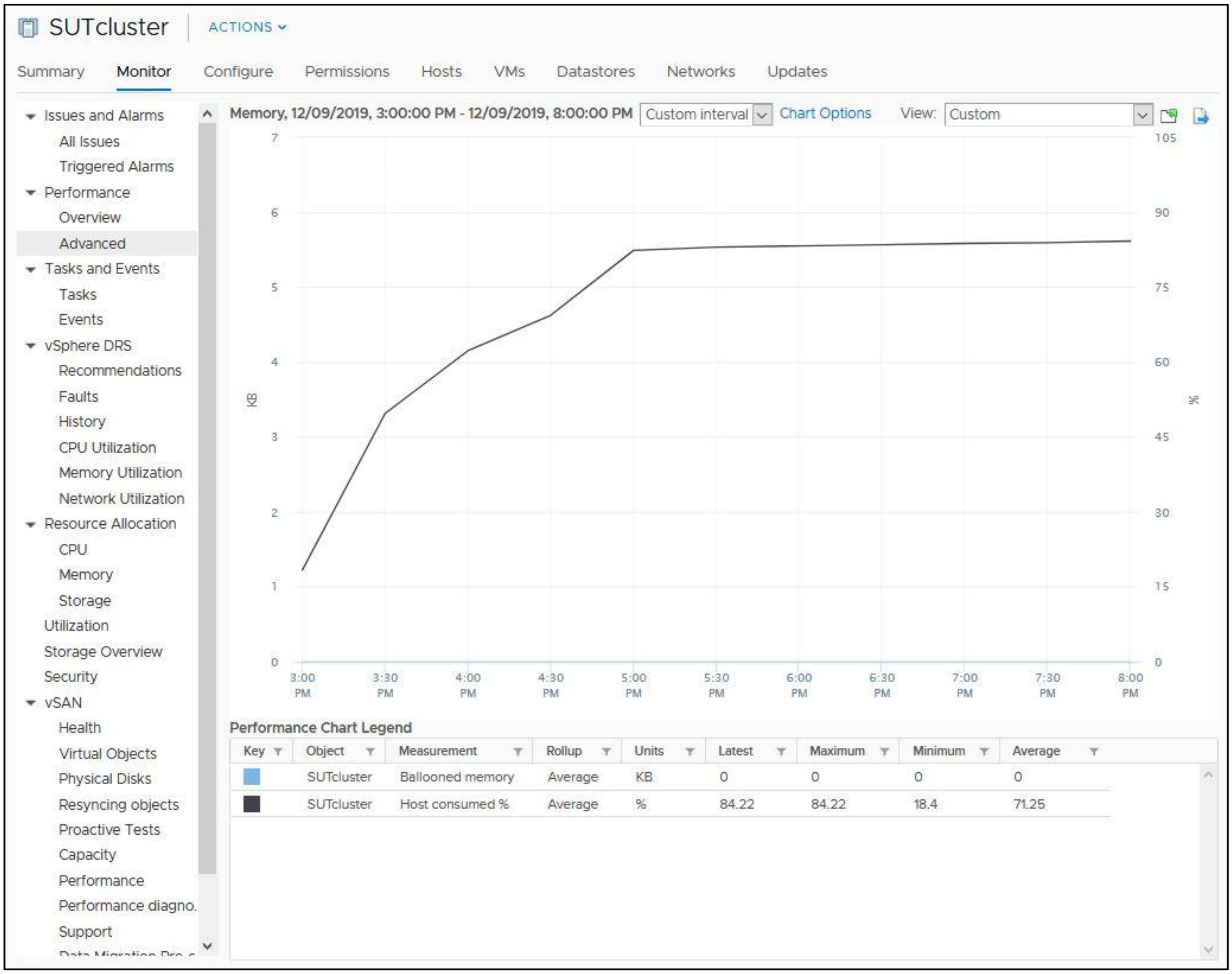

Figure 16. SUT clusters statistics (2)

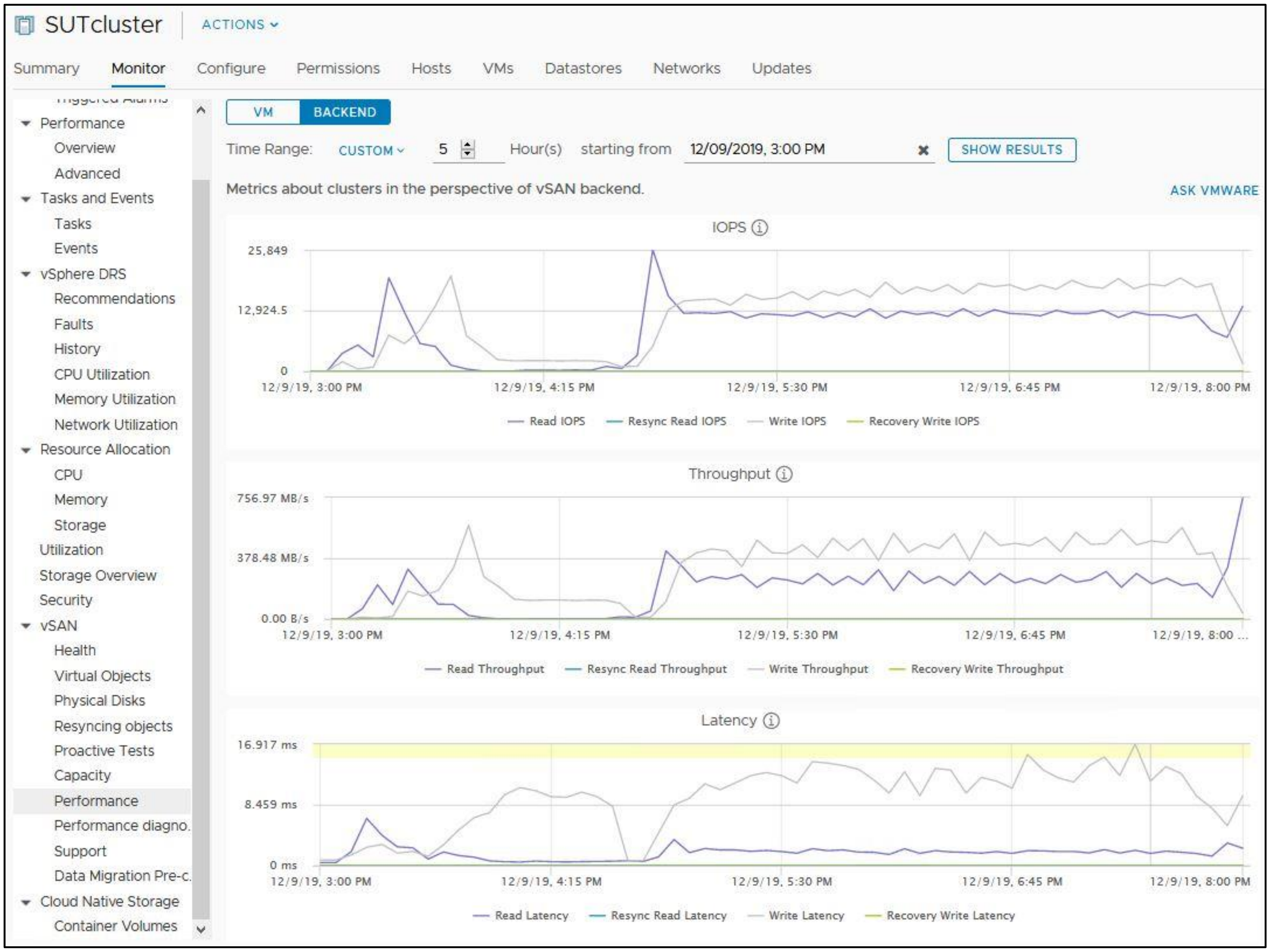

Figure 17. SUT clusters statistics (3)

In conclusion, VMware has already approved the test result of the T42S-2U platform meaning that the overall system is able to provide up to 120 VMs (6 tiles) of various workloads with acceptable performance for typical enterprise use. Additionally, system resources did not reach an upper, allowing flexibility to load more services without adversely affecting performance.

6.3. HCI storage performance- HCIBench Test

To validate the storage performance, a benchmark tool - HCIBench is adopted.

6.3.1. Test Overview

Hyper-Converged Infrastructure Benchmark (HCIBench) is an automation testing tool developed by VMware to evaluate the storage performance of an HCI cluster such as virtual SAN. It is an Open Virtual Appliance (OVA) file which can be deployed on the vSphere® platform and is composed of test controller VM, Ruby vSphere® Console, automation bundle, etc. HCIBench integrates the open source benchmark tool, Vdbench, to execute HCI tests.

The HCIBench appliance and a DHCP VM are installed in the supporting cluster to provide test-related services such as testing guest VMs and IPs. The parameters such as block size, read percentage, and test time are configured in the HCIBench web UI. Once the test task is initiated, the guest VMs are created and deployed in the target vSAN™ cluster.

The test environment is divided into two parts, T41S-2U support server and T42S-2U target cluster. One vCenter Server® Appliance, one DHCP VM, and one HCIBench appliance are deployed on the T41S-2U support server while eight guest VMs are deployed on the T42S-2U target cluster to perform the test process, as shown in Fig. 18.

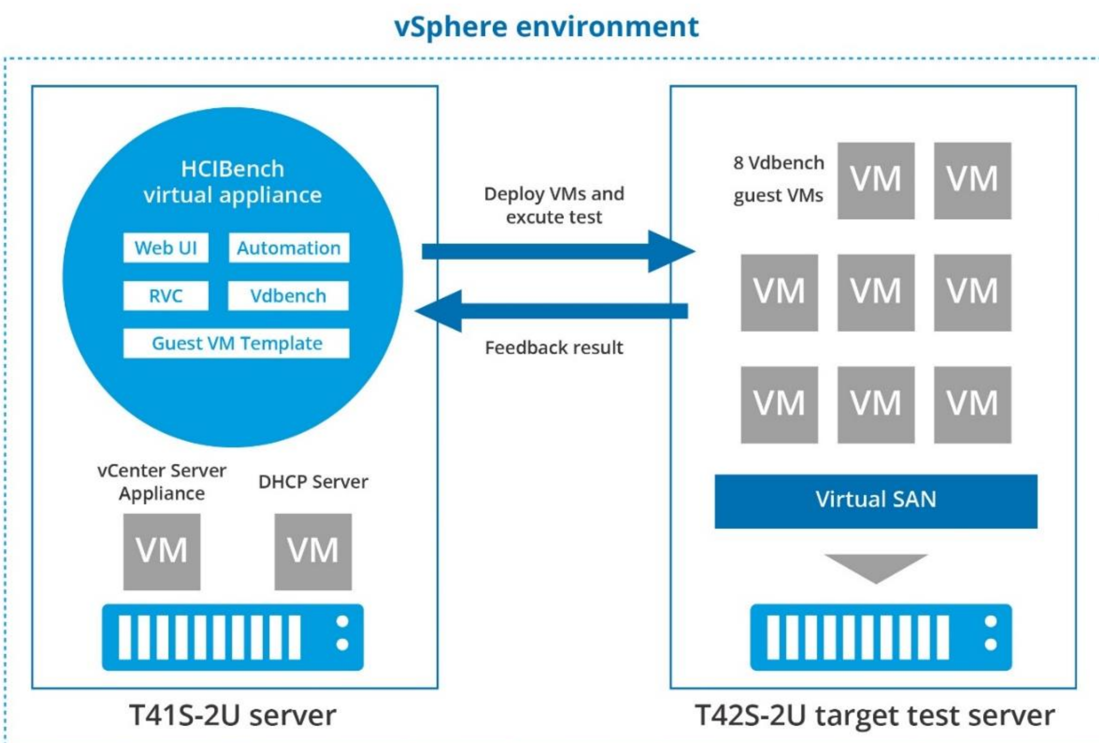

Figure 18. HCIBench Test Topology.

6.3.2. Test Configuration

The six different test cases were executed on the T42S-2U vSAN™ cluster. Several workload parameters defined for all the 6 test cases are listed below:

1. Thread and Queue Depth

In the storage performance tests, each thread is a single-threaded I/O operation and queue depth (QD) is the number of I/Os per thread. Different workloads are stored in different block sizes and the storage performance will be impacted by the choice of size. In the tests, 128 threads per host is set as a baseline.

2. Block size

The block size, also referred to as I/O request size, is the maximum length of a sequence of bytes or bits that applications use to perform I/O operation on storage devices. The block size impacts both the IOPS and throughput since throughput = IOPS x block size. In the tests, the block sizes 4KB, 8KB, 64KB, and 512KB are chosen to be tested.

3. Read and write percentage

The setting of read and write percentages in storage processing depends on the workloads. For example, the web file service normally performs 100% read and 0% write process. The summation of the read/write operation is 100%, that is, if the write percentage is set to 30%, the read percentage will relatively be 70%. In the tests, the predefined percentages are 100% read/ 0% write, 100% write/ 0% read, and 70% read/ 30% write.

4. Random and sequential percentage

The storage I/O process can be classified as random access and sequential access. The random access refers to the disk head that picks up data randomly with no specific order while the sequential access refers to the data accessed from the first to the end orderly. The speed of sequential access is theoretically much higher than random access. In testing, either 100% random or 100% sequential configuration is selected for different cases.

5. VMDK size

Virtual machine disk (VMDK) is a virtual disk used by VM. In vSAN™ testing, different VMDK size could produce different write buffer usage. If the buffer usage is higher than 30% of the cache device, the testing process will result in high write latency.

Among all the test cases, some basic parameters are decided in the following ways:

- Total VMDK size per host=960G*50%= 480G.
- Total VMs per host is 2.
- Total VMDK size per VM=480/{(FTT+1) *2} =120G.
- VMDK number per VM is 8.
- Threads number per VM=128 queue depth/ (8*2) =8.
- Per VMDK size=120/8=15G.

Each test case is executed in 3600 seconds test time. The six test cases and the possible applications are shown below:

- Test case 1: All read and all random 4K workload test for web file server.
- Test case 2: All write and all random 4K workload test for web server logging.
- Test case 3: Mixed read and write with all random 4K workload test for Exchange Email server.
- Test case 4: Mixed read and write with all random 8K workload test for OS drive and database online transaction processing.
- Test case 5: All write and all sequential 64K workload test for database logging and file restoring.
- Test case 6: All read and all sequential 512K workload test for video streaming.

Table 11. HCIBench Test Cases and Defined Parameters.

Test Case	1	2	3	4	5	6
Possible scenarios	web file server	web server logging	Exchange Email server	OS drive and database online transaction processing	database logging and file restoring	video streaming
Queue Depth	128					
Block size (KB)	4	4	4	8	64	512
Read (%)	100	0	70	70	0	100
Write (%)	0	100	30	30	100	0
Random (%)	100	100	100	100	0	0
Sequential (%)	0	0	0	0	100	100
VMDK size	15GB					

6.3.3. Test Results

The test results of the 6 cases provided reference values for their correspondent scenarios, as shown in Table 11 and Table 12. With the vSAN™ hardware configuration and specified test parameters, each result indicated the reference IOPS, throughput, and latency in the vCenter® performance chart, as detailed in Fig. 19 to Fig. 24.

Table 12. HCIBench Test Results.

Test Case		1	2	3	4	5	6
Block size (KB)		4	4	4	8	64	512
Read (%)		100	0	70	70	0	100
Write (%)		0	100	30	30	100	0
Random (%)		100	100	100	100	0	0
Sequential (%)		0	0	0	0	100	100
Test result	IOPS	284K	54K	62K	56K	10K	12K
	Throughput (MB/s)	1090	214	246	442	611	5550
	Latency (ms)	10	10	read: 8 write: 7	read: 4 write: 5	13	11

Test Case 1: Testing for the web file services workload profile resulted in IOPS 284K, throughput 1090 MB/s, and latency 10 ms, as shown in Fig. 19.

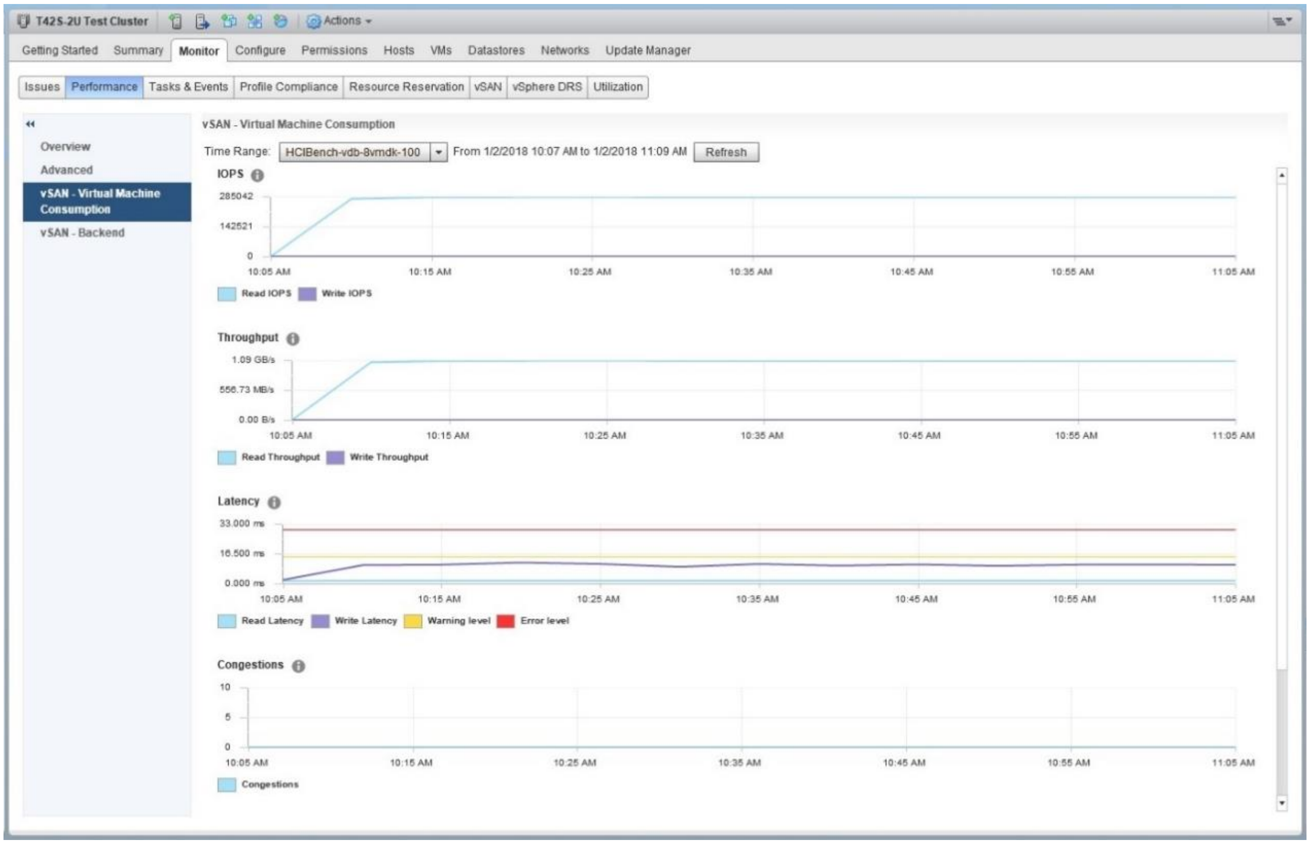

Figure 19. Graphical Results of IOPS, Throughput, and Latency for Test Case 1.

Test case 2: Testing for the web server logging workload profile resulted in IOPS 54K, throughput 214 MB/s, and latency 10 ms, as shown in Fig. 20.

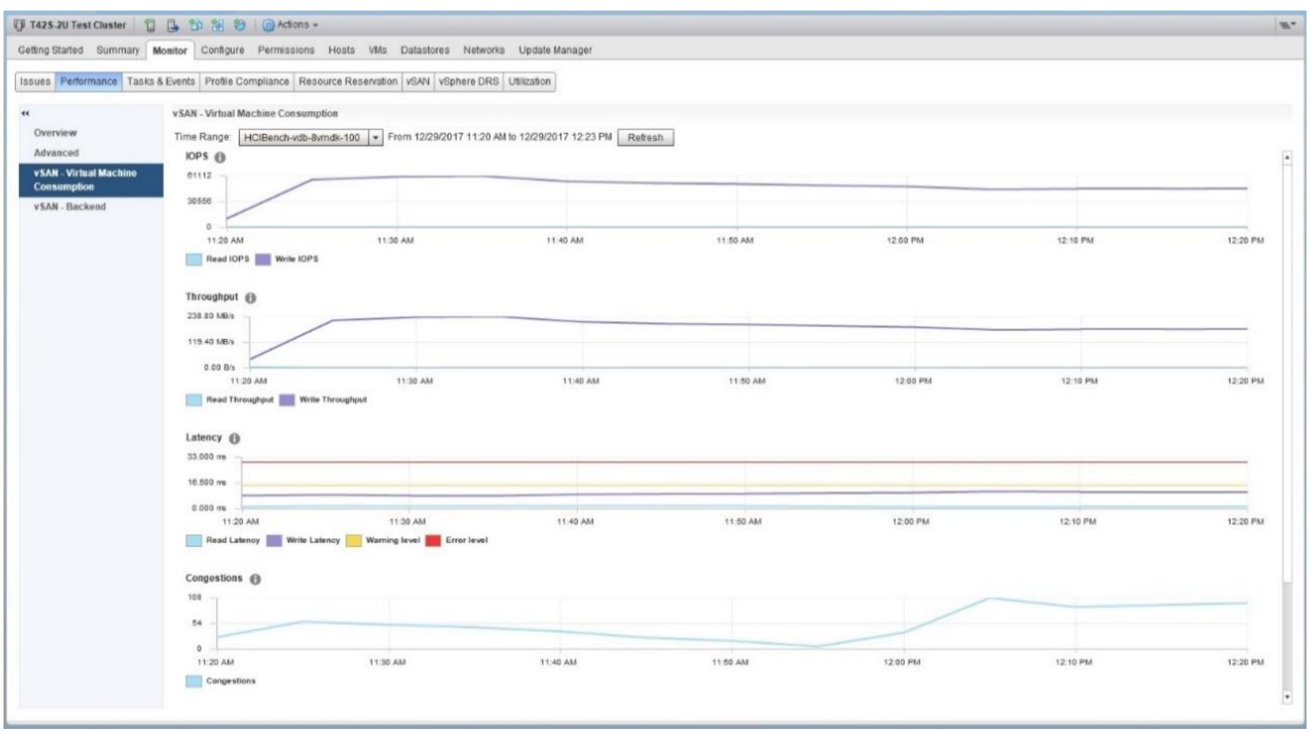

Figure 20. Graphical Results of IOPS, Throughput, and Latency for Test Case 2.

Test case 3: Testing for the Email Exchange server workload profile resulted IOPS 62K, throughput 246 MB/s, and latency 15 ms, as shown in Fig. 21.

Figure 21. Graphical Results of IOPS, Throughput, and Latency for Test Case 3.

Test case 4: Testing for the OS drive or database online transaction processing workload profile resulted in IOPS 56K, throughput 442 MB/s, and latency 9 ms, as shown in Fig. 22.

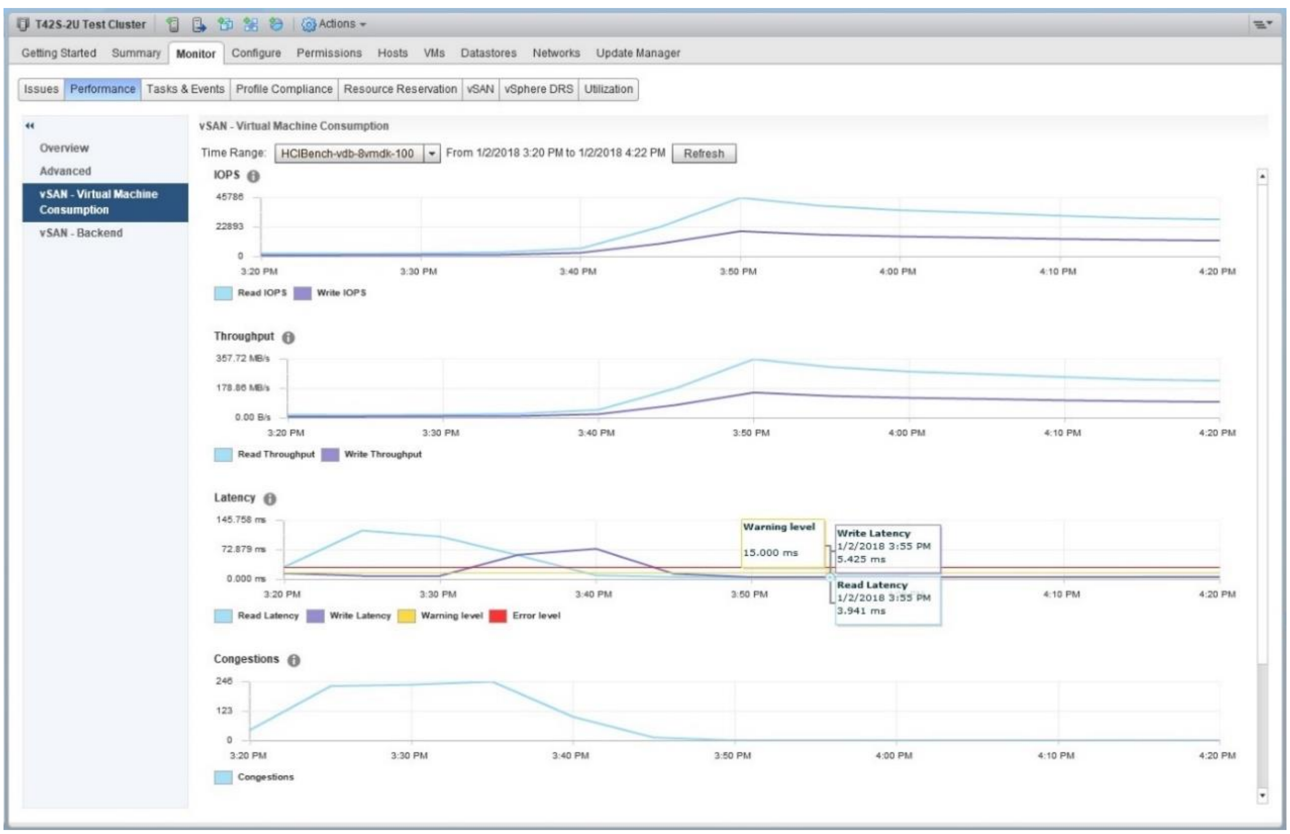

Figure 22. Graphical Results of IOPS, Throughput, and Latency for Test Case 4.

Test case 5: Testing for the database logging and file restoration workload profile resulted in IOPS 10K, throughput 611 MB/s, and latency 13 ms, as shown in Fig. 23.

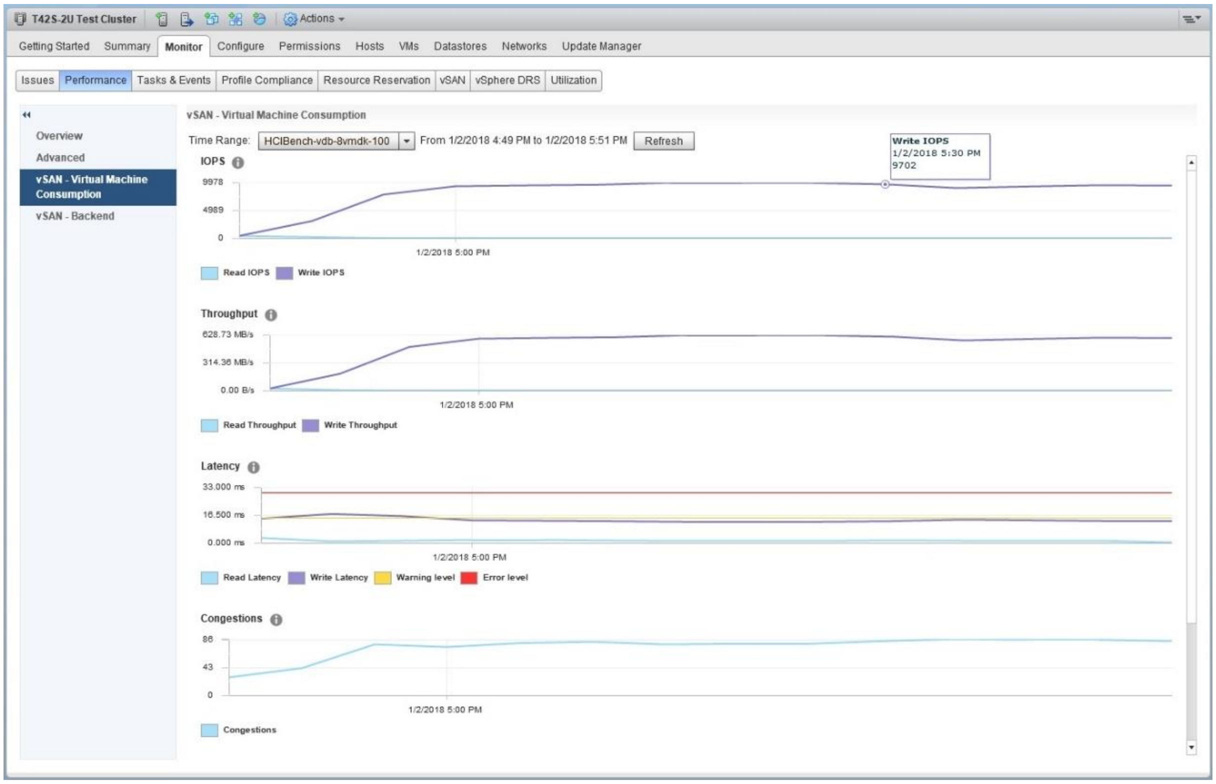

Figure 23. Graphical Results of IOPS, Throughput, and Latency for Test Case 5.

Test case 6: Testing for the video streaming workload profile resulted in IOPS 12k, throughput 5550 MB/s, and latency 11 ms, as shown in Fig. 24.

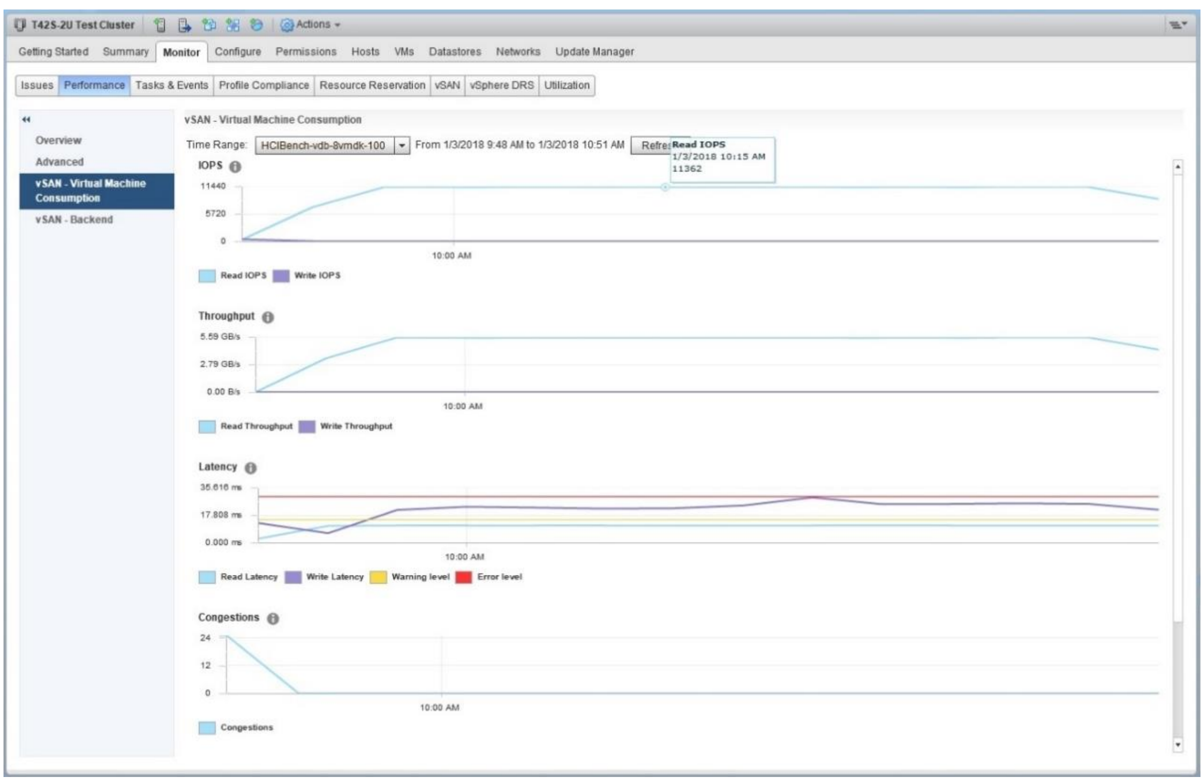

Figure 24. Graphical Results of IOPS, Throughput, and Latency for Test Case 6.

7. Conclusion

Nowadays, data center transformation is considered to be a main trend to address the dynamic business environment. QCT, a global data center solution provider, provides innovative and flexible solutions to keep your organization in a leading position.

QxStack/QxVDI-OA/HC High-Density Optimized SKU is an ultra-dense data center solution rapidly deployed, easily managed, highly qualified, and fully integrated into the industry-leading software-defined storage, vSAN™. QCT executed a strict certification process, including a functional test, stress test, and failure tolerance test at the component level up to the total solution in order to provide customers the most reliability and stability.

This reference architecture has proven that this solution can satisfy diverse use cases, VDI with different scenarios-OA(office application)and HC (high computing) as well as mixed workload scenarios. In the VDI use case, this SKU with OA scenario can support up to 300 virtual desktops; while with HC scenario can support up to 200 virtual desktops. In the mixed-workload use case, this SKU can support up to 6 tiles of mixed workloads in real-time auction or e-commerce scenarios.

According to the test results, the feasibility and stability of the solution are proven in various use cases. By adopting this solution, customers minimize their time and expense in evaluation, selection, deployment, and tuning, ultimately reducing TCO. With QCT's knowledge, customers can leverage the collective results and follow a simplified path to a future-defined data center.

QCT always stays innovative. QCT appreciates any feedback from you. For further inquiry, please visit http://go.qct.io/solutions/

8. Reference

[1] Quanta Could Technology QuantaPlex T42S-2U
https://www.qct.io/product/index/Server/rackmount-server/Multi-node-Server/QuantaPlex-T42S-2U-4Node
[2] VMware vSphere®
https://www.vmware.com/products/vsphere.html
[3] VMware Horizon®
https://www.vmware.com/products/horizon.html
[4] VMware vSAN™
https://www.vmware.com/products/vsan.html
[5] VMware NSX®
https://www.vmware.com/products/nsx.html
[6] VMware VMmark®
https://www.vmware.com/products/vmmark.html
[7] VMware HCIBench
https://labs.vmware.com/flings/hcibench
[8] VMware View Planner
https://www.vmware.com/products/view-planner.html

LEGAL DISCLAIMER

INFORMATION IN THIS DOCUMENT IS PROVIDED IN CONNECTION WITH QUANTA CLOUD TECHNOLOGY (QCT) PRODUCTS. NO LICENSE, EXPRESS OR IMPLIED, BY ESTOPPEL OR OTHERWISE, TO ANY INTELLECTUAL PROPERTY RIGHTS IS GRANTED BY THIS DOCUMENT. EXCEPT AS PROVIDED IN QCT'S TERMS AND CONDITIONS OF SALE FOR SUCH PRODUCTS, QCT ASSUMES NO LIABILITY WHATSOEVER AND QCT DISCLAIMS ANY EXPRESS OR IMPLIED WARRANTY, RELATING TO SALE AND/OR USE OF QCT PRODUCTS INCLUDING LIABILITY OR WARRANTIES RELATING TO FITNESS FOR A PARTICULAR PURPOSE, MERCHANTABILITY, OR INFRINGEMENT OF ANY PATENT, COPYRIGHT OR OTHER INTELLECTUAL PROPERTY RIGHT.

UNLESS OTHERWISE AGREED IN WRITING BY QCT, THE QCT PRODUCTS ARE NOT DESIGNED NOR INTENDED FOR ANY APPLICATION IN WHICH THE FAILURE OF THE QCT PRODUCT COULD CREATE A SITUATION WHERE PERSONAL INJURY OR DEATH MAY OCCUR.

Quanta Cloud Technology (QCT) may make changes to specifications and product descriptions at any time, without notice. Designers must not rely on the absence or characteristics of any features or instructions marked "reserved" or "undefined." QCT reserves these for future definition and shall have no responsibility whatsoever for conflicts or incompatibilities arising from future changes to them. The information here is subject to change without notice. Do not finalize a design with this information.

The products described in this document may contain design defects or errors known as errata which may cause the product to deviate from published specifications. Current characterized errata are available on request.

All products, computer systems, dates, and figures specified are preliminary based on current expectations, and are subject to change without notice. Contact your local QCT sales office or your distributor to obtain the latest specifications and before placing your product order.

ABOUT VMware

VMware software powers the world's most complex digital infrastructure. The company's compute, cloud, mobility, networking and security offerings provide a dynamic and efficient digital foundation to over 500,000 customers globally, aided by an ecosystem of 75,000 partners. Headquartered in Palo Alto, California, this year VMware celebrates twenty years of breakthrough innovation benefiting business and society.

ABOUT QCT

QCT (Quanta Cloud Technology) is a global data center solution provider extending the power of hyperscale data center design in standard and open SKUs to all data center customers.

Product lines include servers, storage, network switches, integrated rack systems and cloud solutions, all delivering hyperscale efficiency, scalability, reliability, manageability, serviceability and optimized performance for each workload.

QCT offers a full spectrum of data center products and services from engineering, integration and optimization to global supply chain support, all under one roof.

The parent of QCT is Quanta Computer Inc., a Fortune Global 500 technology engineering and manufacturing company.

http://www.QCT.io

UNITED STATES

QCT LLC., Silicon Valley office
1010 Rincon Circle, San Jose, CA 95131
TOLL-FREE: 1-855-QCT-MUST
TEL: +1-510-270-6111
FAX: +1-510-270-6161
Support: +1-510-270-6216

QCT LLC., Seattle office
13810 SE Eastgate Way, Suite 190, Building 1,
Bellevue, WA 98005
TEL: +1-425-633-1620
FAX: +1-425-633-1621

CHINA

云达科技, 北京办公室（Quanta Cloud Technology）
北京市朝阳区东大桥路 12 号润诚中心 2 号楼
TEL +86-10-5920-7600
FAX +86-10-5981-7958

云达科技, 杭州办公室（Quanta Cloud Technology）
浙江省杭州市西湖区古墩路浙商财富中心 4 号楼 303 室
TEL +86-571-2819-8650

JAPAN

Quanta Cloud Technology Japan 株式会社
東京都港区芝大門 2-5-8 芝大門牧田ビル 3F, 105-0012
TEL +81-3-5777-0818
FAX +81-3-5777-0819

GERMANY

Quanta Cloud Technology Germany GmbH
Hamborner Str. 55, 40472 Düsseldorf
TEL +492405-4083-1

TAIWAN

雲達科技（Quanta Cloud Technology）
桃園市龜山區文化二路 211 號 1 樓
1F, No. 211 Wenhua 2nd Rd., Guishan Dist., Taoyuan City 33377,
Taiwan