QCT™

QxStack VMware® Edition
-vSAN ReadyNode™

Cost/Capacity Optimized SKU

Reference Architecture

Table of Contents

31

Quanta CLOUD TECHNOLOGY

QCT CONFIDENTIAL

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.

1   Introduction

Quanta Cloud Technology (QCT) is a global data center solution provider, combining the hyperscale hardware with infrastructure software to innovate next-generation data center design and conquer operation challenges. QCT serves cloud service providers, telecoms, and enterprises for running public, hybrid, and private clouds.

Software-Defined Everything (SDx) has already impacted the world such as data center, hosting company, small and medium enterprises, university, etc. IT organizations are considering to transfer the architecture from traditional Physical Defined Data Center (PDDC) to Software Defined Data Center (SDDC) virtualization environment. It is time to transform the way of building cloud since the leading technology x86 server with VMware Software-Defined Storage (SDS) can absolutely deliver simplified management, agility, and lower TCO.

QCT QxStack VMware Edition-vSAN ReadyNode™ is a hyper-converged solution, developed to replace the traditional storage and silos. By integrating and validating QCT ultra-design hardware and VMware software stack, QxStack VMware Edition-vSAN ReadyNode™ solution provides three optimized SKUs to fulfill multiple workloads and different scenarios such as Infrastructure as a Service (IaaS) and Remote Office and Branch Office (ROBO). According to VMware in July, 2017, VMware vSAN™ users who adopt VMware's hyper-converged infrastructure software are rapidly growing to 8,000 users. The 64% of users adopts vSAN™ to run their business-critical applications across a variety of industries. SDS is adopted from small-scale industry to large-scale industry. It's no longer a technology preview. QxStack VMware Edition- vSAN ReadyNode™ is designed to step into the "leading technology", and provide simplified centralized management and multi-purpose hyper-converged infrastructure.

1.1 Purpose

The purpose of this reference architecture is to introduce a pre-validated QxStack VMware Edition- vSAN ReadyNode^{TM} solution and provide several use cases adopted in the solution. To design the reference architecture, the components are carefully selected by following VMware's compatible guide. The proposed solution is based on the information and experiences from QCT and VMware which minimizes the efforts and risks of the deployment.

QxStack VMware Edition- vSAN ReadyNode^{TM} is a series of hyper-converged IaaS appliances certified by VMware vSAN ReadyNode™ certification. It solves the complexity implementation and management problems in a simply way which enables users to easily get started. The highlighted key advantages and benefits that QCT QxStack VMware Edition- vSAN ReadyNode^{TM} solution offers are listed in the following sections.

1.2 Scope

This document demonstrates the architecture of QCT **QxStack VMware Edition-vSAN ReadyNode^{TM} Cost/Capacity Optimized SKU** and its validation process. It provides a reference architecture for users to build the best Cost/Capacity infrastructure in VMware environment, as shown in Fig. 2. Three QuantaGrid D51PH-1ULH servers with pre-installed VMware vSphere®, VMware vSAN™, and VMware vCenter® software are adopted. To deliver a best-practice solution, this reference architecture will demonstrate hardware and software descriptions, test environment, validation steps, and scenarios.
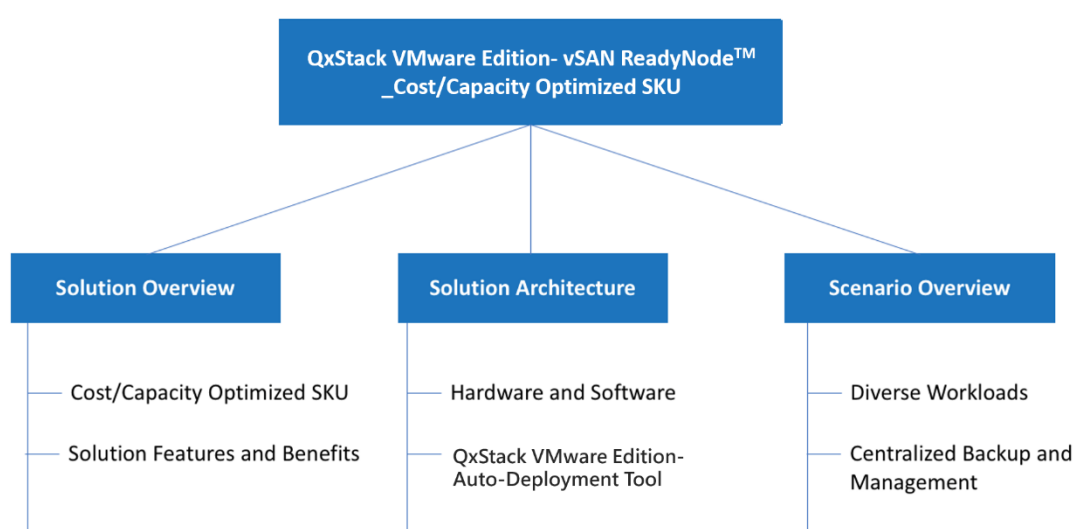

Figure 1. Structure of the Proposed Reference Architecture.

2   Solution Overview

QCT (Quanta Cloud Technology) provides a high-capacity appliance powered by VMware virtualization software. By integrating ultra-designed servers and hyper-converged VMware software, QCT ensures an outstanding performance when supporting multiple workloads.

QCT has three pre-validated and optimized solutions to meet users' requirements, including IOPS-Optimized SKU, Compute-Storage Optimized SKU, and Cost/Capacity Optimized SKU, as shown in Fig.3. IOPS-Optimized SKU is a performance-oriented solution to fulfill the requirements of low latency, high IOPS, and high throughput. Compute-Storage Optimized SKU provides high-compute density with 24 disks in a 2U 4-node box. Cost/Capacity Optimized SKU is a capacity-oriented solution. It is a market-leading solution, providing high capacity, agility, and quick implementation in a 1U system. In this document, QCT mainly focus on demonstrating the Cost/Capacity Optimized SKU, QCT QxStack VMware Edition-vSAN ReadyNode[TM]. By adopting this solution, users can maximize usable storage for a better business productivity.

IOPS Optimized SKU	Compute-Storage Optimized SKU	Cost/Capacity Optimized SKU
Support up to 10*NVMe in a system	24x 2.5" hot-plug	Support up to 4*SSD+12*HDD in a system
Provide highest IOPS in the VMware vSAN™ environment.	Deliver compute-storage balance in the VMware vSAN™ environment.	Deliver the best Cost/Capacity in the VMware vSAN™ environment.

Figure 2. QCT QxStack VMware Edition- vSAN ReadyNode[TM] Optimized SKUs.

Solution Validation

QCT make lots of efforts on pre-integrated, pre-configured and pre-validated total solution to maximize value and minimize all risky issues while integration. Hardware components are validated by QCT, Intel, and VMware to ensure cloud performance and reliability, as shown in Fig. 1.
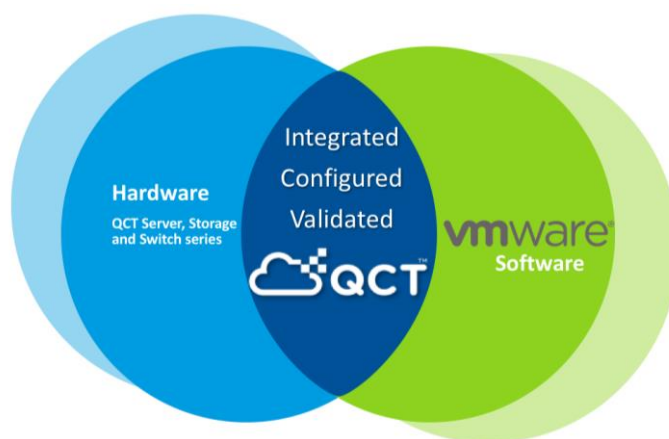

Figure 3. QCT Solution Validation.

Time Efficiency

QCT's solution powered by VMware simplifies the implementation process. With pre-validated solutions, QCT ensures to save users configuration time and implementation time, and reduces operational risks.

2.1 Hyper-Converged Infrastructure

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 highly 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 and several benefits are listed below.

Simplified Management

In legacy-converged architecture, since the compute and storage resources are provided by different servers and storage devices, the configuration settings and operation management are independent which means IT administrators must manage two devices through different management tools. To provide a storage device to a server, users 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. Users 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, users 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 4.
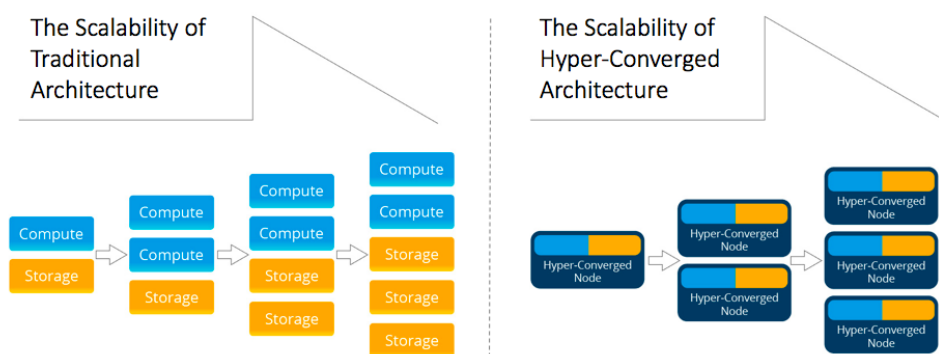

Figure 4. Comparison between Traditional Architecture and HCI.

2.2 vSAN ReadyNode™

VMware vSAN™

VMware vSAN™ is a mature software-defined storage technology for hyper-converged solutions. Uniquely embedded in the hypervisor, vSAN™ delivers high-performance, flash-optimized, easily-scaled hyper-converged storage for any virtualized application. vSAN™ cluster server with attached flash devices and/or hard disks provide a flash-optimized, highly-resilient shared data store for a variety of workloads, including business-critical applications, virtual desktops, remote IT, DR, and DevOps infrastructure.

QCT QxStack VMware Edition- vSAN ReadyNode™ is a solution certified by VMware vSAN ReadyNode™ program and its designed architecture is shown in Fig. 5. The solution adopts 3 units of QCT servers certified by VMware and VMware software (i.e., vSphere®, vSAN™, vCenter®, and vRealize® Operations Manager™) to design QCT vSAN ReadyNode™ optimized SKU.
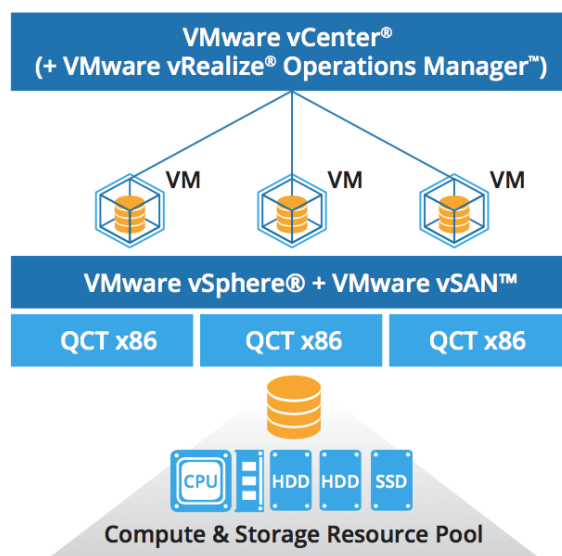

Figure 5. Architecture of QxStack VMware Edition- vSAN ReadyNode™.

vSAN ReadyNode™ is a program created by VMware to guarantee the performance and the stability of HCI. Users can choose the vSAN™ version for deployment, select a ReadyNode profile, and choose one of the available models. In the vSAN ReadyNode™, two main categories are Hybrid SKU and All-Flash SKU. The Hybrid SKU consists of four profiles, HY2, HY4, HY6, and HY8, as shown in Fig. 6. The All-Flash SKU consists of three profiles, AF4, AF6, and AF8, as shown in Fig. 7.

The differentiation among these profiles are hardware configurations such as CPU core, DIMM, and caching performance, and capacity performance. QxStack VMware Edition- vSAN ReadyNode™ Cost/Capacity Optimized SKU is certified by VMware vSAN ReadyNode™ HY4. The hardware for this solution will be elaborated in the following section.

VMware vSAN ReadyNode-Hybrid SKU				
	HY2	HY4	HY6	HY8
Number of VM per Node	Up to 20	Up to 30	Up to 50	Up to 100
IOPS per Node	Up to 4K	Up to 10K	Up to 20K	Up to 40K
Raw Storage Capacity per Node	2TB	4TB	8TB	12TB
CPU	1x6 core	2x8 core	2x10 core	2x12 core
Memory	32GB	128GB	256GB	384GB
Capacity Tier Flash	2x 1TB SAS 7.2K RPM	4x 1TB SAS 7.2K RPM	8x 1TB SAS 7.2K RPM	12x 1TB SAS 10K RPM
Caching Tier Flash	1x200 GB SSD or Endurance Class >=B Performance Class >=B	1x200 GB SSD Endurance Class >=C Performance Class >=D	2x200 GB SSD Endurance Class >=C Performance Class >=D	2x400 GB SSD Endurance Class >=D Performance Class >=E
IO Controller	Queue Depth >=256	Queue Depth >=256	Queue Depth >=256	Queue Depth >=512
NIC	10 GbE	10 GbE	10 GbE	10 GbE

* VM density can vary based on use case
** Assumes latest generation CPU architecture
*** Pass through mode recommended; additional controller may be required depending on maximum number of drives supported by the controller

Figure 6. VMware vSAN ReadyNode™ — Hybrid Profile.

VMware vSAN ReadyNode-All Flash SKU			
	AF4	AF6	AF8
Number of VM per Node	Up to 30	Up to 60	Up to 120
IOPS per Node	Up to 25K	Up to 50K	Up to 80K
Raw Storage Capacity per Node	4TB	8TB	12TB
CPU	2x10 core	2x10 core	2x12 core
Memory	128GB	256GB	384GB
Capacity Tier Flash	4x1 TB SSD Endurance Class A or above Performance Class C or above	8x1 TB SSD Endurance Class A or above Performance Class C or above	12x1 TB SSD Endurance Class A or above Performance Class C or above
Caching Tier Flash	1x200 GB SSD Endurance Class C or above Performance Class C or above	2x200 GB SSD Endurance Class C or above Performance Class D or above	2x400 GB SSD Endurance Class D or above Performance Class F or above
IO Controller	Queue Depth >=256	Queue Depth >=512	Queue Depth >=512
NIC	10 GbE	10 GbE	10 GbE

* VM density can vary based on use case
** Assumes latest generation CPU architecture
*** Pass through mode recommended; additional controller may be required depending on maximum number of drives supported by the controller

Figure 7. VMware vSAN ReadyNode™ — all flash profile.

3 Solution Architecture

In this session, we will focus on illustrating the architecture of QxStack VMware Edition- vSAN ReadyNode^{TM} Cost/Capacity Optimized SKU, including hardware configuration, software stack, and QxStack VMware Edition－Auto-Deployment Tool.

3.1 Hardware Configuration

Tailored for hyper-scale data centers and software-defined storage solution, QuantaGrid D51PH-1ULH features hybrid-tiered storage architecture in an ultra-dense hot-swappable 1U platform. The external cable management arm is not required which significantly reduces the system deployment and rack assembly time, and simplifies the rack cable routing in the rear.

QuantaGrid D51PH-1ULH is a rackmount server with Intel® Xeon® processor E5-2600 v3, v4 product family and can be expanded to 1TB memory capacity. It is equipped with 12 hot-swappable 3.5" disk drives and 4 hot-swappable 2.5" SATA SSD (7mm) which can help technicians to highly minimize efforts and reduce downtime. This server provides both extreme storage density and compute power to hyper-scale cloud data centers. The specification of QuantaGrid D51PH-1ULH is illustrated in Table 1.

Caching is prerequisite not only in today's software-defined storage solution but also virtualization applications to boost performance and IOPS. The hybrid architecture is ideal for tier storage planning in which the solid-state drives are required to accelerate IOPS and throughput without sacrificing large data storage capacity.

Table 1. Specification of QuantaGrid D51PH-1ULH.

Processor	Dual Intel® Xeon® processor E5-2600 V3/V4 family, TDP up to 135W
Chipset	Intel® C610
Memory	(16) 2133/1866/1600/1333MHz DDR4, support 8GB, 16GB and 32GB RDIMM/LRDIMM
Form Factor	1U Rack Mount
Dimensions (WxHxL)	17.6" x 1.7" x 35" (448.2mm x 43.2mm x 881mm)
Expansion Slot	(1) SAS Mezzanine x8 (1) OCP LAN Mezzanine slot x8
Storage	(12) 3.5"/2.5" hot-plug 12Gb/s SAS or 6Gb/s SATA HDD/SSD (4) 2.5" hot-plug 7mm 6Gb/s SATA SSD (1) Internal SATA DOM
Front I/O	(1) USB 2.0 port
Power Supply	700W Platinum PSU, 1+1 redundant power supplies, 100- 240V AC
Fan	(6) 40x56 dual rotor fan
System Management	IPMI v2.0 Compliant, on board "KVM over IP" support
Operating Environment	**Operating temperature:** 5°C to 35°C (41°F to 95°F) **Non-operating temperature:** -40°C to 65°C (-40°F to 149°F) **Operating relative humidity:** 50% to 85%RH Non-operating relative humidity: 20% to 90%RH
Weight (Max. Configuration)	~62.8lb (~28.5 Kg)
Storage Controller	Wellsburg PCH RSTe
Network controller	Intel® I350 dual-port 1 GbE Dedicated 10/100/1000 management port **Option:** QCT Intel® I350 dual-port OCP mezzanine QCT Intel® X540 dual-port 10GbE BASE-T OCP mezzanine QCT Intel® 82599ES dual-port 10G SFP+ OCP mezzanine QCT Mellanox CX3PRO 10G SFP+ OCP mezzanine with RDMA support
Onboard Storage	(1) SATADOM
Video	Integrated AST2400 with 8MB DDR3 video memory
Rear I/O	(2) USB 3.0 ports (1) VGA port (1) RS232 serial Port (2) 1GbE RJ45 ports (1) GbE RJ45 management port
TPM	Yes (TPM 2.0/1.2, LPC)

Hardware Optimization for vSAN™

QuantaGrid D51PH-1ULH server is the suitable hardware with both extreme storage density and compute power for the hyper-converged solution. The hardware configuration design in this reference architecture is listed on VMware VCG portal, as shown in Fig. 8. In this design, 3 QuantaGrid D51PH-1ULH servers contribute its local

storage to build a vSANDatastore. Each server uses two Intel® Xeon® E5-2620 v4 CPUs with 8 cores to provide outstanding performance for compute and storage services. 256GB memory capacity for each node is used and can be extended to 1TB memory capacity.

vSAN ReadyNode™ Details	Model : HY4-QCT-QuantaGrid D51PH-1ULH Profile: HY-4 Series	Partner Name: Quanta Computer Inc Type: Hybrid	Generation: Gen2 - 12G

Components	Details		Quantity
SKU	QuantaGrid D51PH-1ULH_HY4		
ESXi Pre-Installed?	Yes		
System	Model : QuantaGrid D51PH-1ULH	System Type : Rackmount	1
CPU	Intel E5-2620 v4		2
Memory	32GB		8
Caching Tier	Model : Intel® SSD DC S4600 Series SSDSC2KG019T7(1.92TB,2.5") SATA Device Type: SATA Performance Class: Class E: 30,000-100,000 writes per second	Partner Name: Intel Capacity: 1920 GB TBW Endurance Class: Endurance Class D >=7300 TBW	4
Capacity Tier	Model : SAS 8TB ST8000NM0075 Device Type: SAS	Partner Name: Seagate Capacity: 8000 GB	12
Controller	Model : Quanta SAS 3008 Mezz card Driver: lsi_msgpt3 version 15.00.00.00-1OEM	Firmware: 09.00.00.00	1
NIC	Model : Quanta Computer Inc ON 10GbE 82599ES Network Max_Interface_Speed: N/A	Driver: N/A Firmware: N/A	1
Boot Device	Model : SATADOM 32GB		1

Figure 8. QxStack VMware Edition- vSAN ReadyNode™ Cost/Capacity Optimized SKU_HY4 list in VMware Website.

Each Quanta SAS 3008 IO controller hosts 4 Intel® Solid-State Drives (SSDs) DC S3710 (1.2TB, 2.5 inches) for cache tier and 12 Hard Disk Drives (HDDs) for the capacity tier under vSAN™'s architecture. Each QuantaGrid D51PH-1ULH host contributes its local disks to a vSANDatastore and the disks are organized into disk groups. Totally, each host consists of four disk groups. A disk group can be seen as a "fault domain." If the cache device fails, all HDDs and SSDs in the same disk group will be impacted. However, the design of multiple disk groups can reduce the impact of the overall disk group when cache device fails. More than three HDDs in the same disk group will prolong the overall time to rebuild the data. Therefore, each disk group with 1 SSD and 3 HDDs are designed to reduce the risk of cache failure and time for rebuilding the data in vSAN™, as shown in Fig. 9.
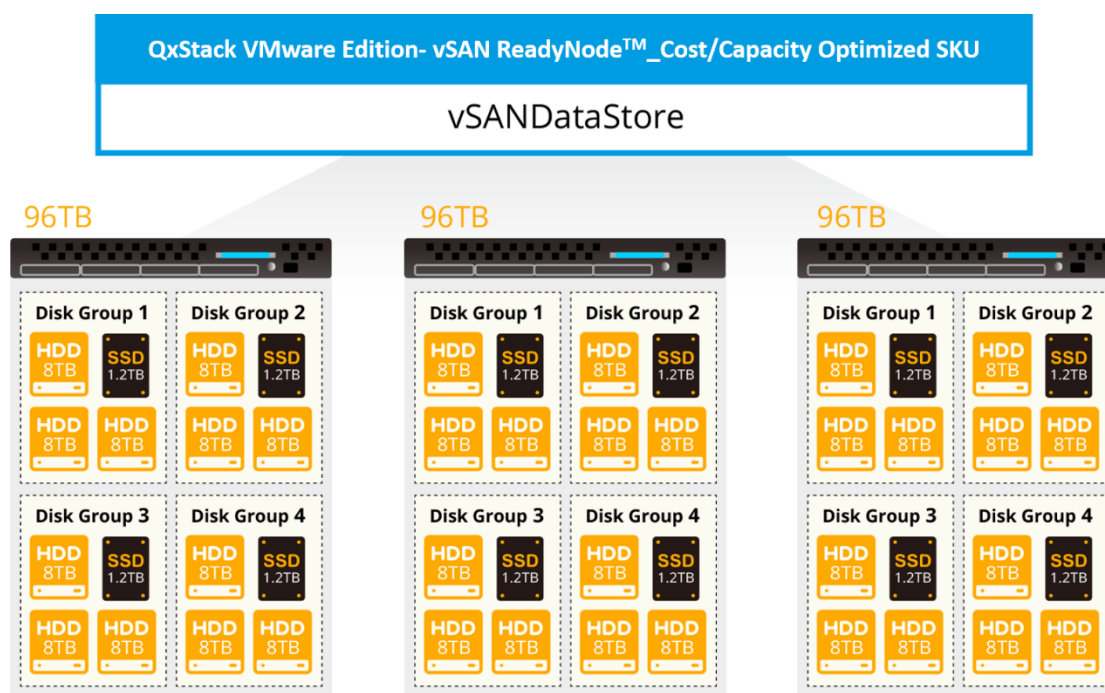
Figure 9. vSANDataStore Configuration of QxStack VMware Edition- vSAN ReadyNode™ Cost/Capacity Optimized SKU.

Intel® SSD is selected to provide an extraordinary performance and endurance for vSAN™ hybrid cache tier. Cache tier is designed to accelerate the I/O processing and maximize both read and write performances. The optimal value of the flash capacity size is based on the actual workload. Generally, cache device consumes at least 10 percent of storage capacity in hybrid vSAN™ configurations since most of the virtualized applications requires a working set around 10% suggested by VMware. The purpose is to keep the active working sets in the cache tier as much as possible and achieve a better performance. However, 5% is set as a baseline value in this SKU. The percentage difference can be neglected since the design is certified by vSAN ReadyNode™ which guarantees an achievable performance. In each disk group, three 8TB Seagate® SAS hard drives are selected to highlight the extreme storage for the capacity tier. The 1.2TB cache size is calculated based on the 5% of the 24TB hard drive capacity consumed by virtual machine storage.

Network Planning

A well-designed network topology ensures that vSAN™ traffic can run efficiently, correctly, and availably between hosts. VMware suggests 10 Gigabit Ethernet (GbE) network for vSAN™ to avoid the bottleneck of network bandwidth. In our network design, a VMkernal port is created on the vSwitch, dedicated for vSAN™ traffic. If administrator needs to run multiple traffic on the same NIC, Quality of Service (QoS) can be implemented using Network I/O Control (NIOC). NIOC can allocate the dedicated network bandwidth to the vSAN™ traffic. To achieve the network redundancy on the uplink network adapter, NIC teaming is a way to "team" two NICs together. The two uplink adapters "vmnic1" and "vmnic2" are teamed on a vSwitch to either "two active uplinks" or "an active/standby uplink" for failover and redundancy purposes. In Fig.10, two physical switches with link aggregation are prepared at the upper layer. The real network topology is different depending on users' requirements.
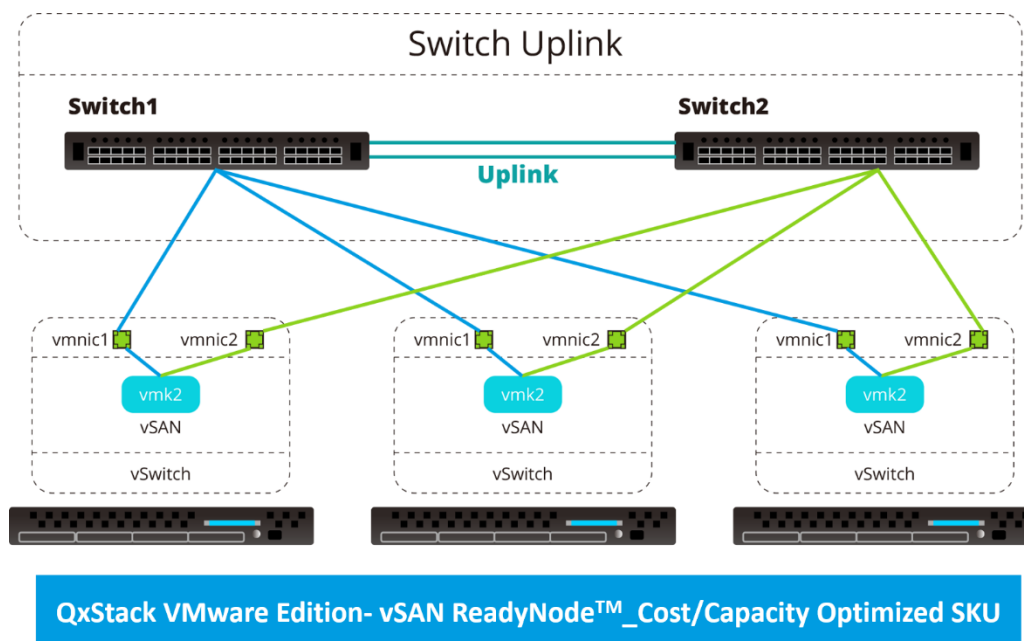
Figure 10. Network planning of QxStack VMware Edition- vSAN ReadyNode™ Cost/Capacity Optimized SKU.

3.2 Software Configuration

VMware vSphere®

VMware vSphere® is an industry-leading virtualization platform which virtualizes and aggregates the physical hardware to provide a virtual resource pool to the data center. By leveraging the virtualization technology, vSphere® can provide a highly available, efficient, and centralized infrastructure for IT administrators to deliver flexible and reliable services. VMware vSphere® provides some benefits:

1. Using proactive high-availability technology to prevent machine downtime.
2. Using predictive load balancing technology to fully exert the datacenter resources.
3. Simplified user experience to deliver a large-scale automation and management.
4. Leveraging virtual machine level encryption technology to reduce risk.
5. Using REST API to promote IT automation and business flexibility.

VMware vSphere® is composed by ESXi™ and vCenter® which will be stated below.

VMware vCenter Server®

vCenter server® is a centralized platform which provides management and operation for VMware virtualization environment. By aggregating all the virtual resources, vCenter server® can provide resource provisioning and monitoring. It provides the capability to provision compute, storage, and other resources to the virtual machine and enables High Availability (HA), Distributed Resource Scheduler (DRS), vMotion, etc. Currently, in vCenter® 6.5, vCenter server® can support up to 2000 hosts.

ESXi™

VMware ESXi™ is an industry-leading hypervisor, installed on a bare-metal physical server. ESXi™ includes its own kernel, called VMkernal based on Linux kernel. ESXi™ enables the virtualization technology which breaks the traditional hardware architecture, including compute, storage, and networking. By sharing the resources of a single hardware across multiple environments, a physical system (x86 server) is capable of executing multiple virtual machines with different operation systems (OS). VMware proposed that "virtualization is the process of creating a software-based (or virtual) representation of something rather than a physical one." Virtualization can be applied to applications, servers, storage, and networks. It is a single effective way to reduce IT expenses and boost efficiency and agility for all-sized businesses.

vSAN™ is a software-defined storage which is built in the vSphere kernel. It is tightly integrated with the hypervisor to minimize the CPU and memory overhead and optimizes the data I/O path to deliver an outstanding performance. vSAN™ is a hyper-converged solution, particularly designed for virtual machines. It minimizes the effort to configure the storage and simplifies the virtual machine deployment.

vSAN™ is built on the industry-standard x86 server and leverages the local storage on the server. By using the virtualization technology, the underlayer physical disk is a concept of abstract resource pool, aggregated into a virtual resource for providing a shared storage to the server in the cluster. It also makes non-disruptive expansion capacity possible by adding hosts to a cluster or adding disks to a host. SSDs are used as a cache tier to accelerate the I/O performance and HDDs as capacity tier to store the data. By leveraging the flash device on server side, vSAN™ can accelerate the read/write I/O processing speed and minimize the storage latency. Administrators can use vSAN™ to define the requirements of virtual machine storage, such as performance and availability. The policy requirements are delivered to vSAN™ layer through the Storage Policy-Based Management (SPBM). SPBM, hence, plays an important role to bridge the upper layer application and the underlayer storage devices. Through pre-defined storage policy, SPBM will drive the vSAN™ layer to adopt these policies when users provision the virtual machine.

3.3 QxStack VMware Edition－Auto-Deployment Tool

Solution Deployment

QxStack VMware Edition — Auto-Deployment Tool simplifies the solution deployment process and provides an easy way to deploy VMware solution, including Virtual Desktop Infrastructure (VDI) and Infrastructure as a Service (IaaS).

Deployment Process

In order to facilitate the deployment process of QxStack VMware Edition- vSAN ReadyNode^{TM}, QCT developed a QxStack VMware Edition－Auto-Deployment Tool to dramatically reduce time and minimize guest effort for building hyper-converge solutions. Auto-Deployment Tool plays a coordinate role such as initializing ESXi™ and deploying vCenter server®.

Administrators can initiate the Auto-Deployment Tool on a single node and type the default IP. Once the tool is powered on, the deployment tool is connected and a view of GUI can show administrators the deployment process. Subsequently, DHCP service needs to be enabled and applied to each server that is going to be deployed, as shown in Fig. 11. After clicking "Save and Next", three different options including "System Initial Only or Scale Out", "Install vCenter® and Prebuild Cluster" and "Install vCenter® and VDI environment" can be selected. To initiate and deploy the environment, "Install vCenter® and Prebuild Cluster" can be selected to execute the appliance deployment process.

Figure 11. DHCP Service Enablement of QxStack VMware Edition—Auto-Deployment Tool.

Administrators need to enter a general system information, including inband management IP information, ESXi™ hosts information, and vCenter® information, as shown in Fig. 12. Once the administrators enter all the information and click the "Install vCenter® and Prebuild Cluster", the system will initiate each ESXi™ hosts and automatically deploy the vCenter® and vSAN™ service. The entire deployment process will take around 20 to 30 minutes, as shown in Fig. 13.

Figure 12. General System Information of QxStack VMware Edition—Auto-Deployment Tool.

System will be deployed automatically, please wait for 20~30 mins.
Do not close the window during deployment.

vCenter environment already initialized.
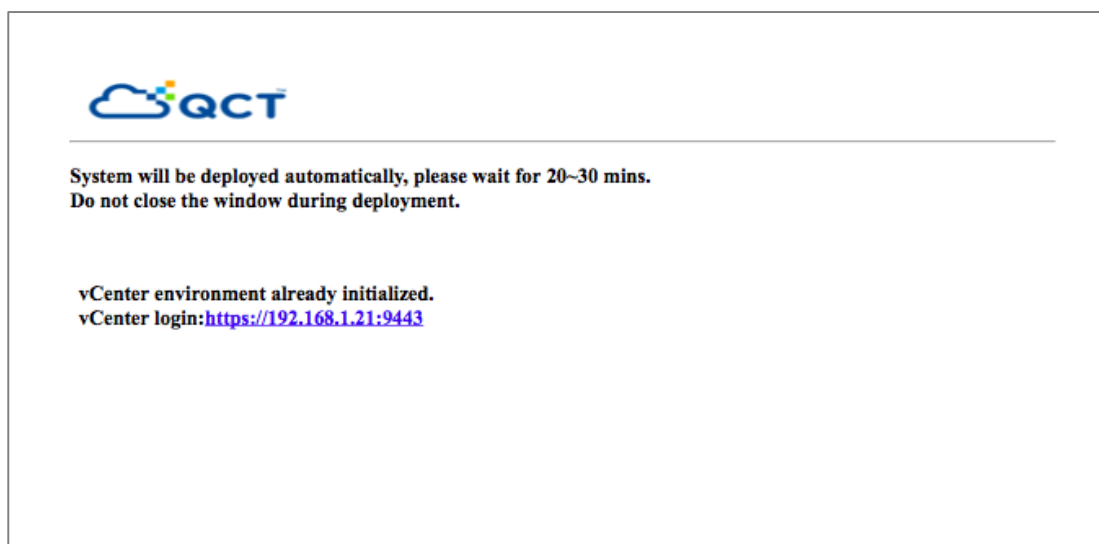vCenter login:https://192.168.1.21:9443

Figure 13. Deployment Process of QxStack VMware Edition－Auto-Deployment Tool.

4   Scenario Overview

The design for QxStack VMware Edition- vSAN ReadyNode^{TM} Cost/Capacity Optimized SKU mainly focuses on the two scenarios "diverse workloads" and "centralized backup and management." In this section, we will present different scenarios and their use cases.

4.1   Diverse Workloads

vSAN™ is a hyper-converged solution, providing not only storage resource but also compute resource. vSAN™ is suitable for a variety of workloads, including business-critical applications, virtual desktops, remote IT, DR, and DevOps infrastructure. The architectures below are proposed to fulfill the diversity of the use cases.

Use Case 1

Storage Policy Based Management (SPBM) is a policy driven control plane which can dynamically allocate storage resources to virtual machines. SPBM will drive the data plane like vSAN™ to comply with the storage policy. The storage policy contains different storage rules, including "Primary level of failures to tolerate", "Number of disk stripes per object", "Object space reservation", "Flash read cache reservation", and "IOPS limit for object." Administrators can use different combinations of storage rules to achieve the storage Service Level Agreement (SLA). By predefining the storage rules of I/O performance, fault tolerant, availability and space reservation, SPBM can easily assign these rules to virtual machines or their individual disks during the provision, as shown in Fig. 14.
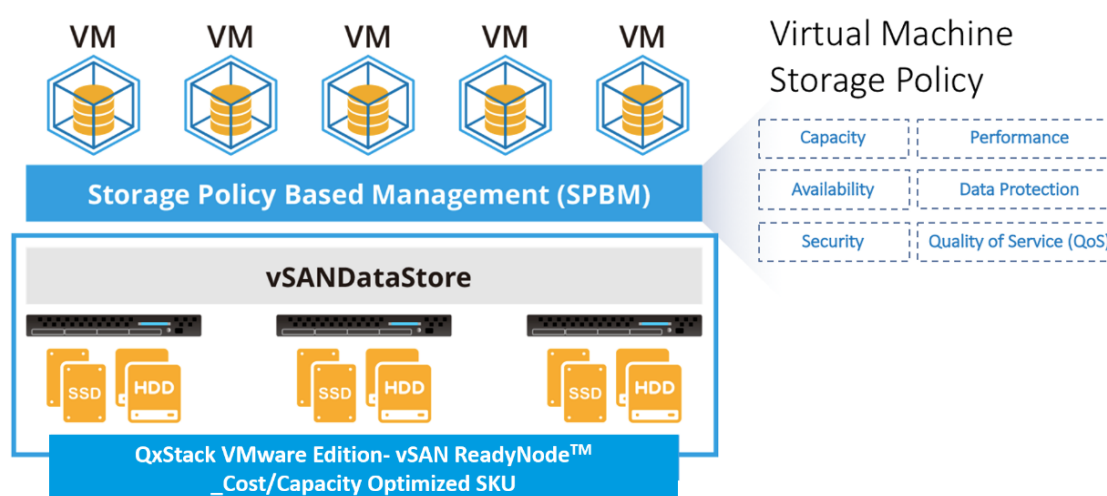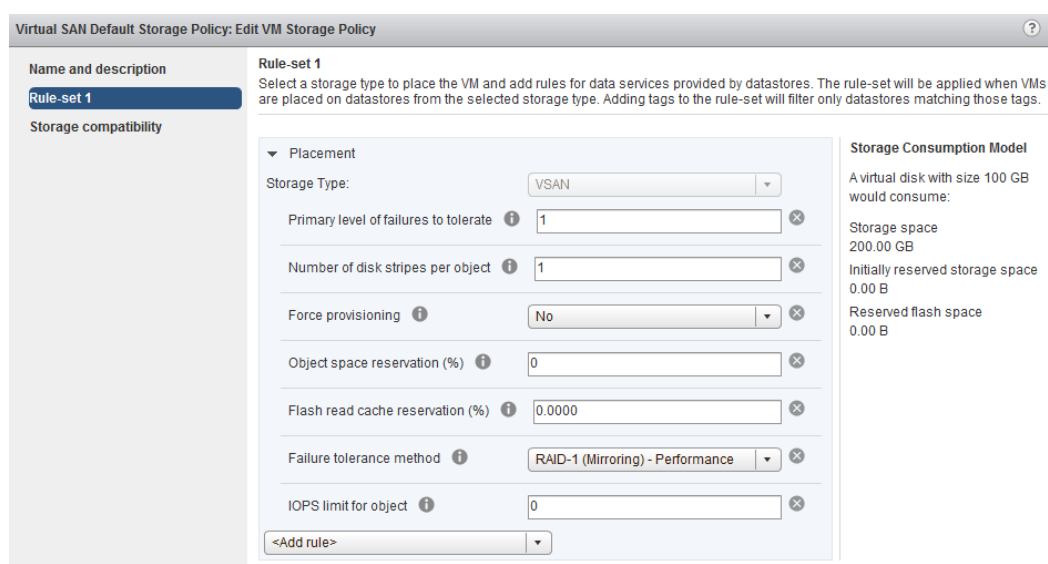
Figure 14. Storage Policy Based Management with Virtual Machine Storage Policy on vSAN™.

The rule "Primary level of failures to tolerate" defines the number of disk, host or fault domain failures a storage object can tolerate. This rule provides a VM with more redundancy at the availablitiy levels.

The rule "Number of disk stripes per object" can be adjusted to improve the performance for some applications that need high performance from the storage. This rule indicates the number of capacity device across each replica of a storage object. To comply this rule, the data will spread across drives and improve the performance from the overall storage which can be used on large streaming write workloads and increasingly de-stage the speed on the storage back end. Another way to improve the performance is to provide the "Flash read cache reservation" for the storage object. To guarantee the performance for business-critcal applications, administrators can set the rule to reserve flash read cache to a specific VMDK.

The rule "Object space reservation" enables administrators to reserve capacity for objects. When the storagae objects are provisioned, the percentage of the storage object will be thick provisioned.

The rule "IOPS limit for object" can define IOPS limitation for an object. If a specific VM consumes more resources than the others, the rule can gaurantee the specific VM will not exceed the IOPS. By setting this rule, it can ensure one VM won't impact the other VMs or the overall vSANDatastore performance. These vSAN™ storage policy rules are shown in Fig. 15.

Figure 15. Rule Set of vSAN™ Storage Policy.

To meet the storage policy driven by SPBM, different rules mentioned above are selectively combined to set up a customized policy for different requirements. The

customized policy are defined as three different levels of applications, namely, "Golen Policy", "Silver Policy", and "Bronze Policy", as shown in Fig. 16.

"Golden Policy" is defined to fit the business-critical applications with high level of the availability and performance. To ensure the availability, the rule "Primary level of failures to tolerate" is set to 1. To increase the performance, the rule "Number of disk stripes per object" is set to 1 or above and the rule "Flash read cache reservation" can be set according to users' requirements. The rule "IOPS limit for object" may need when the application impacts the overall vSANDatastore performance.

"Silver Policy" is defined to fit the general applications which are not performance intensive. vSAN™ default policy is suitable for this general purpose. To ensure the redundancy, the default value of "Primary level of failures to tolerate" is 1. To guarantee the basic performance, the default value of "Number of disk stripes per object" is 1. Both "Flash read cache reservation" or "IOPS limit for object" do not need to be set.

"Bronze Policy" is set to fit the low priority applications such as staging and testing. These testing applications do not need to guarantee the redundancy or performance. The rule "Primary level of failures to tolerate" can set to 0 which means there is no need to have another replica data. In this way, it can reduce the usage of the capacity. The disk space can use thin-provisioned form, that is, "Object space reservation" does not need to be set.
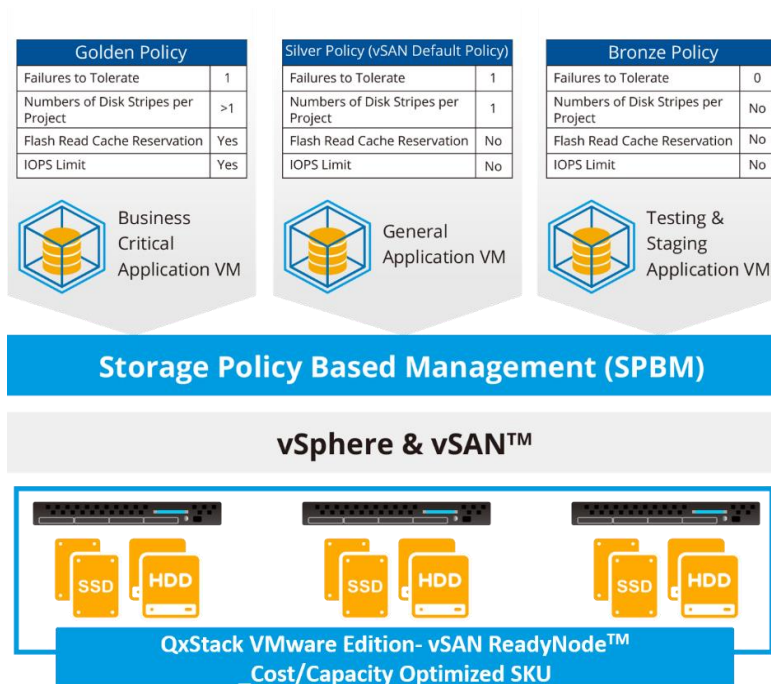

Figure 16. Customized policies including Golden, Silver, and Bronze Based on SPBM.

Use Case 2.

In this use case, the capability of vSAN™ is validated for different application workloads. HCIbench is used as a benchmark tool to simulate different I/O loads on Hyper-Converged Infrastructure (HCI). HCIbench benchmark tool is based on the VDbench open source to fully automate the test process. The measured values are input/output operations per second (IOPS), latency, and throughput. HCIbench can be used to generate synthetic I/O loads and simulate different application I/O behaviors. With the I/O loads, we can measure storage performance in a virtual environment and diagnose storage performance.

Five different types of I/O loads with the workload behaviors will be the targets in this case, as shown in Table 2. IOPS, throughput, and latency generated from the five different types of I/O loads are recorded. The specification of I/O loads is illustrated below.

Table 2. Specification of I/O Load Access.

Block Size [kB]	Workload Type	Justification
4	100% Random, 100% Read	Popular test, close to Web File Server and File Server workloads
4	100% Random, 70%/30% Read/Write	Most Popular Virtualized Application Behavior
8	100% Random, 70%/30% Read/Write	Database/Online Transaction Processing, Operating System Drive
64	100% Sequential, 100% Write	SQL Server Logging
512	100% Sequential, 100% Read	Table Scan

- 4K Block Size－All Read Workload: Each VDBench worker thread is configured to access random reads across the entire volume. The workload can be used to evaluate the maximum random read IOPS a storage solution can deliver.

- 4K Block Size－Mixed Read/Write (Mixed R/W) Workload: Each VDBench worker thread is configured to execute a mixed R/W access with a 70%/30% ratio. Since most applications adopt a mixed read and write workload, this trace comes the closest to represent the performance of a commercial application deployed in a vSAN™ cluster.

- 8K Block Size－Mixed Read/Write (Mixed R/W) Workload: Each VDBench worker thread is configured to execute a mixed R/W access with a 70%/30% ratio. The I/O loads can simulate the Database Online Transaction Processing (OLTP) and Operating System Drive.

- 64K Block Size－Sequential Write: Each VDBench worker thread is configured to

execute a sequential write across the entire volume. The I/O loads can simulate the SQL Server logging process.

- 512K Block Size—Sequential Read: Each VDBench worker thread is configured to execute a sequential read across the entire volume. The I/O loads can simulate the Table Scan.

Hybrid vSAN™ Cluster with I/O Loads

The five types of workloads are divided based on different I/O workloads. In each VM, the IOPS and the latency are measured for each I/O request for a 30-minute steady-state duration on each running instance. The measured values are added across each node in a cluster to get cumulative cluster-wide IOPS, throughput, and latency. The values can help administrators understand and characterize the vSAN™ performance, as shown in Table 3.

Table 3. Measured Values in Five Types of Workloads.

Block Size [kB]	Workload Type	IOPS	Throughput [MB/s]	Latency [ms]
4	100% Random, 100% Read	434,500	1697	1.7
4	100% Random, 70%/30% Read/Write	275,736	1077	2.7
8	100% Random, 70%/30% Read/Write	232,287	1814	3.3
64	100% Sequential, 100% Write	27877	1742	3.4
512	100% Sequential, 100% Read	9844	4922	10

Quanta CLOUD TECHNOLOGY

4.2 Centralized Backup and Management

In this scenario, QxStack VMware Edition- vSAN ReadyNode™ Cost/Capacity Optimized SKU can be used as a centralized backup storage for backup server or backup appliance. VMware vSphere® Data Protection™ is used as a backup appliance to design an architecture which can be easily applied to the third-party backup appliance. VMware vSphere® Data Protection™ is a backup and recovery solution provided by VMware. When users deploy the vSphere® Data Protection™ appliance, the virtual machine disk (VMDK) files for the vSphere® Data Protection™ virtual appliance can be provisioned on vSANDatastore. Therefore, vSAN™ can provide additional protection to the backup data storage, as shown in Fig. 17.
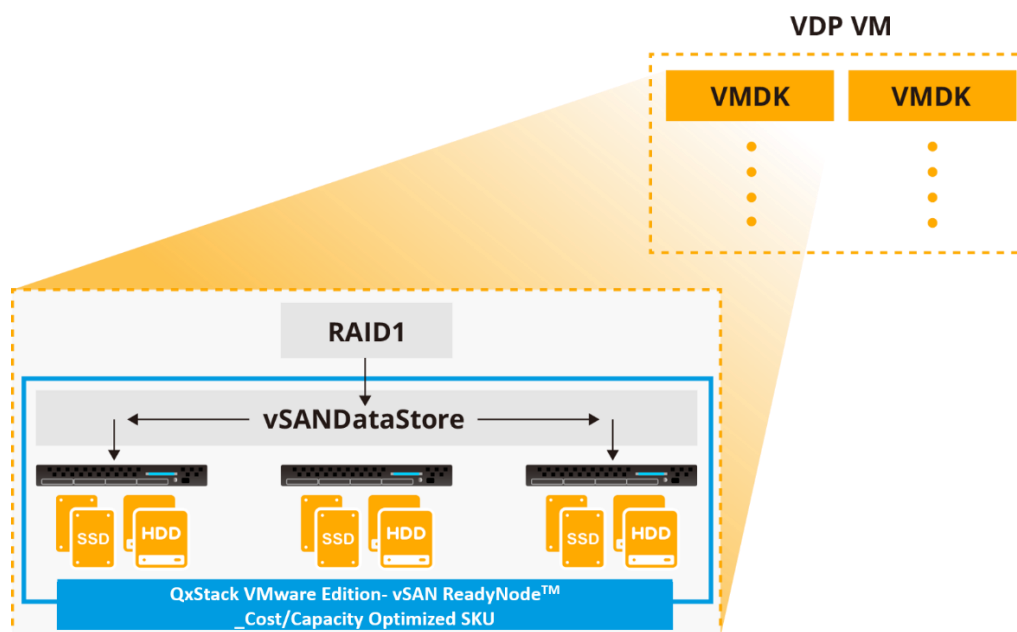

Figure 17. vSphere® Data Protection™ Provisioned on vSANDatastore.

Use Case 1

Backup can be created under the same vCenter® (site) or can be replicated between different vCenters (site) using vSphere® Data Protection™. This capability is useful when administrators move backup data to an off-site location or replicate the backup data from one site to another. For example, administrators can replicate data from production site to disaster recovery (DR) site for data backup or archive, as shown in Fig. 18.
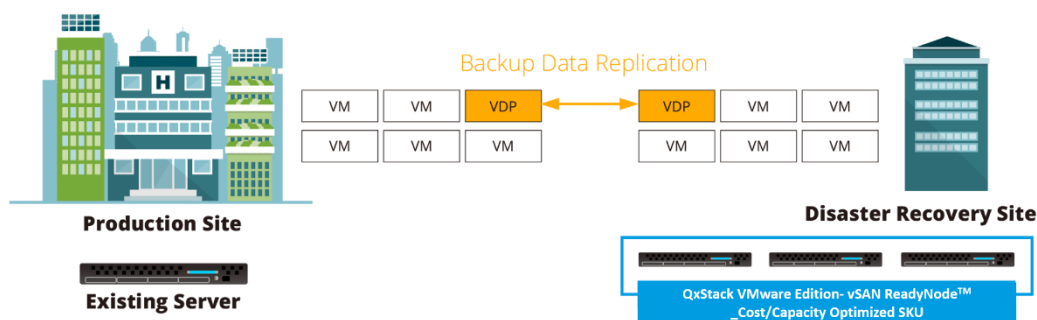
Figure 18. Backup Data Replication between Production Site and Disaster Recovery Site.

The replication can be a many-to-one topology which indicates that the virtual machines are backupped from Remote Office A and Remote Office B to the Headquarter-Primary Data Center, as shown in Fig. 19.
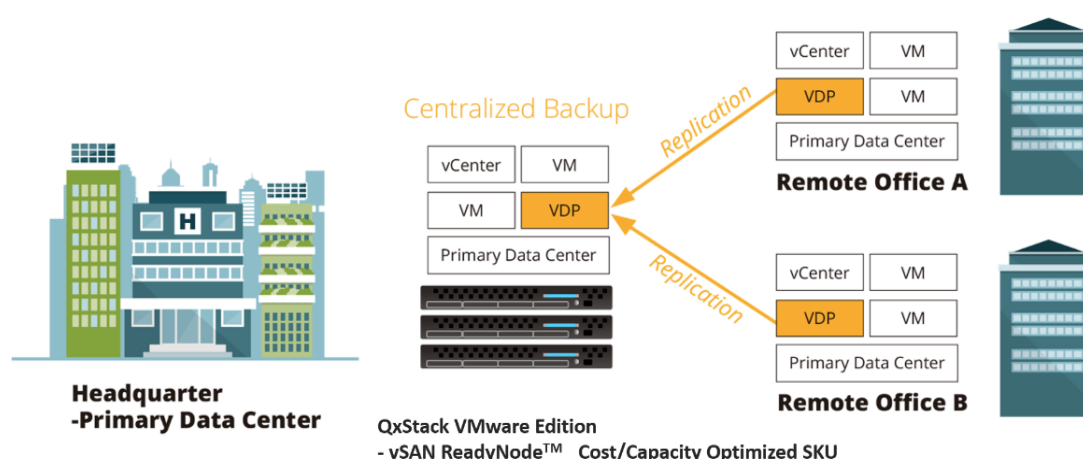

Figure 19. Replication from Remote Office to Primary Data Center.

Use Case 2

With vSAN™ 6.5, vSANDatastore can be extended to several physical servers through iSCSI target service. It can easily create a LUN through vSphere Web Client by several clicks. LUNs in vSANDatastore are similar to other objects which can be managed through Storage Policy Based Management (SPBM). Administrators can change the storage policy, including performance and availability to adapt the service level requirements on the specific LUN. In this use case, when physical server or backup server needs a storage for data backup or archive, QxStack VMware Edition-vSAN ReadyNode™ Cost/Capacity Optimized SKU can provide an iSCSI protocol to let the physical server access the volume, and simultaneously provide the performance and availability SLA to the volume, as shown in Fig. 20.
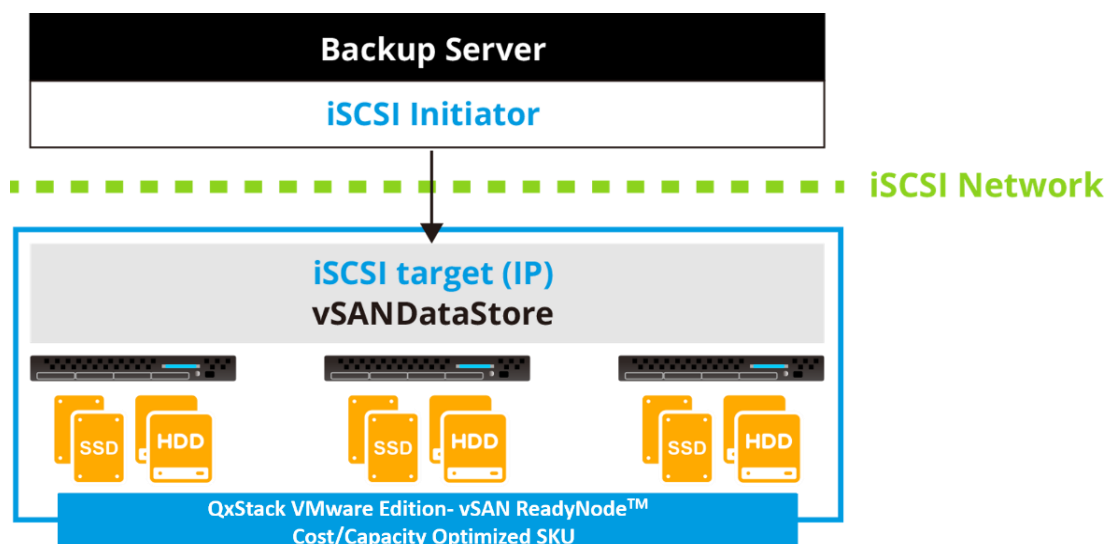
Figure 20. Backup Storage Provisioned on vSAN™ iSCSI Target.

Use Case 3

VMware provides a single portal vCenter® to manage compute and storage resources. By deploying and configuring vSAN™ service, there is no need to create LUNs or volumes which can minimize the setup and management efforts to the storage. vSAN™ service can be simply enabled to provide a shared storage for virtual machines to be provisioned by merely several clicks on vCenter®. vCenter® also features vSAN™ health monitoring and alert. The health service check plugin checks the hardware compatibility, networking configuration, advanced vSAN™ configuration options, storage device health, and virtual machine object health, as shown in Fig. 21. It can provide administrators an overall view of the vSAN™ deployment and real-time alert when the failure of vSAN™ services occurs. This is radically helpful for administrator to troubleshoot and highly reduce the resolution time.

Figure 21. vSAN™ Health Service Check Plugin.

The performance metric for vSAN™ is also useful information for administrators to decide if vSAN™ encounters the performance issue. The performance metrics provide the information such as IOPS, latency, and throughput from cluster level down to each VMDK level, as shown in Fig. 22.

Figure 22. Performance Metrics for vSAN™.

5   References

1. **VMware vSAN™ Customer Amounts**
    https://www.vmware.com/products/vSAN™/customers.html
2. **V**Mware® vSAN™ Design and Sizing Guide 6.5**
    https://www.google.com.tw/url?sa=t&rct=j&q=&esrc=s&source=web&cd=1
    &cad=rja&uact=8&ved=0ahUKEwiryPOjudjVAhUKvbwKHTGOBpkQFggpMAA
    &url=https%3A%2F%2Fstoragehub.vmware.com%2Fexport_to_pdf%2Fvmwa
    re-r-virtual-san-tm-design-and-sizing-
    guide&usg=AFQjCNHSBusC8pMJnrcP0XVX_dXBceQ4bQ
3. **HCIBench Benchmark Tool**
    https://labs.vmware.com/flings/hcibench
4. **VMware vSAN Health Check Guide**
    https://www.vmware.com/techpapers/2015/vmware-virtual-san-health-
    check-guide-10505.html

Quanta CLOUD TECHNOLOGY

About QCT

Quanta Cloud Technology (QCT) is a global datacenter solution provider. We combine the efficiency of hyperscale hardware with infrastructure software from a diversity of industry leaders to solve next-generation datacenter design and operation challenges. QCT serves cloud service providers, telecoms and enterprises running public, hybrid and private clouds.

Product lines include hyper-converged and software-defined datacenter solutions as well as servers, storage, switches, integrated racks with a diverse ecosystem of hardware component and software partners. QCT designs, manufactures, integrates and services cutting edge offerings via its own global network. The parent of QCT is Quanta Computer, Inc., a Fortune Global 500 corporation.

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 号润城中心 2 号楼
Tower No.2, Run Cheng Center, No.12, East Bridge Rd.,
Chaoyang District, Beijing, China
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
Room 303，Building No.4，ZheShang Wealth Center
No. 83 GuDun Road, Xihu District, Hangzhou, Zhejiang , China
TEL: +86-571-2819-8650

Japan
Quanta Cloud Technology Japan 株式会社
日本国東京都港区芝大門二丁目五番八号
牧田ビル 3 階
Makita Building 3F, 2-5-8, Shibadaimon ,
Minato-ku, Tokyo 105-0012, Japan
TEL: +81-3-5777-0818
FAX: +81-3-5777-0819

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
TEL: +886-3-286-0707
FAX: +886-3-327-0001

Germany
Quanta Cloud Technology Germany GmbH
Hamborner Str. 55, 40472 Düsseldorf , Germany
TEL: + 492405-4083-1300

Other regions
Quanta Cloud Technology
No. 211 Wenhua 2nd Rd., Guishan Dist., Taoyuan
City 33377, Taiwan
TEL: +886-3-327-2345
FAX: +886-3-397-4770