WHITE PAPER – **MAY 2017**

RUNNING CONTAINERS AT SCALE WITH VMWARE PHOTON PLATFORM

A Secure, Multitenant Architecture for Cloud-Native Applications

vmware®

Table of Contents

USE CASES

- Kubernetes as a service: Deploy and resize highly available Kubernetes clusters to develop, test, and automate applications.
- Infrastructure as a service: Provide self-service access to VMs, disks, and subnets to support high-churn workloads.
- Platform as a service: Integrate with Pivotal Cloud Foundry to build and scale applications for the cloud.
- Continuous Integration and Delivery: Improve the CI/CD pipeline with uniformity and reusability.
- API-managed on-premises cloud: Deploy a vast cloud of virtual machines and automate their management through an API.

Introduction

Enterprises are increasingly adopting container technology. A recent survey by 451 Research revealed a profile of impressive implementation for an emerging ecosystem.[1]

For early adopters, containers are on the cusp of moving from development and testing to production. As container technology matures and early adopters seek to put containerized applications into production, such operational concerns as security and management come to the fore.

Meanwhile, other organizations are beginning to experiment with containers. Some see the technology as a way to modernize their traditional applications while implementing the services and infrastructure to develop new applications with container technology.

Both early adopters and new experimenters see the potential of containers to help them adapt to the changes in the marketplace brought about by the digital transformation. Companies are under pressure to rapidly create and deploy innovative software that engages their customers—a demand best fulfilled by shifting to containers, microservices, distributed frameworks, orchestration tools, and other technology used to build and run cloud-native applications.

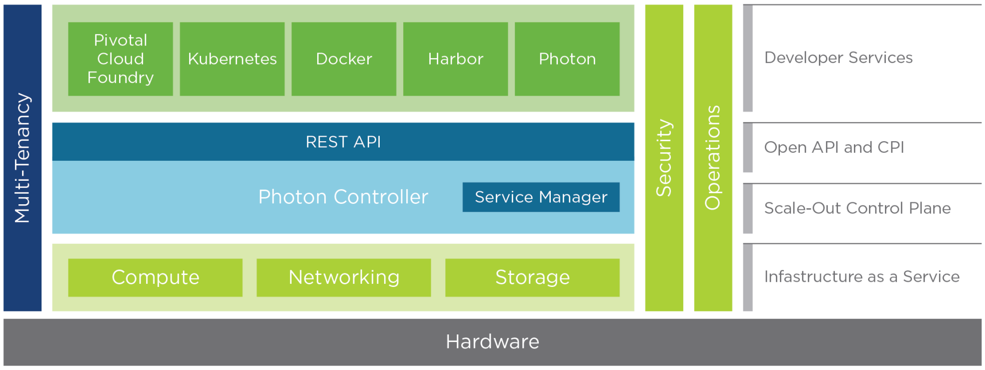
VMware Photon™ Platform delivers compute, storage, and networking infrastructure that is optimized for running cloud-native applications. The platform combines VMware Photon Controller with Project Lightwave™, VMware NSX®, and VMware vSAN™. Photon Controller manages virtualized infrastructure and supplies Kubernetes as a service to securely run containerized workloads at scale. Lightwave secures Photon Platform, NSX provides software-defined virtualized networking, and vSAN establishes a virtual storage area network.

This technical overview describes Photon Platform's components and illustrates how they fit together to form an elastic, distributed architecture for running containerized applications at scale.

The High-Level Architecture of Photon Platform

The following diagram shows how the components of Photon Platform come together to deliver Kubernetes as a service, infrastructure as a service, and platform as a service.

VMware Photon Platform



[1] "Application containers will be a $2.7bn market by 2020, representing a small but high-growth segment of the Cloud-Enabling Technologies market," 451 Research, Jan. 10, 2017. https://451research.com/blog/1351-application-containers-will-be-a-$2-7bn-market-by-2020,-representing-a-small-but-high-growth-segment-of-the-cloud-enabling-technologies-market

PHOTON PLATFORM MAIN COMPONENTS

- Photon Controller
- Lightwave security services
- ESXi
- Load balancer
- NSX
- vSAN
- Photon OS
- Pivotal Cloud Foundry (optional)

The main components of Photon Platform are Photon Controller, VMware ESXi™ hypervisors on commodity hardware, vSAN distributed storage, NSX virtual networking, and the Lightwave security suite. Project Photon OS, a minimalist Linux operating system from VMware, is freely available on demand as stackable, replaceable hosts for running Docker containers.

The architecture of Photon Platform revolves around Photon Controller, which provides the core services that enable system administrators, DevOps, and developers to create the following kinds of user accounts and infrastructure elements:

- Tenants
- Projects and project users
- Virtual machines
- Disks for virtual machines and persistent storage
- Virtual subnets
- Docker containers
- Kubernetes clusters

The ESXi hosts and the virtual machines in a Photon cluster are powered by NSX, which establishes a virtual network of routers and switches. The virtual network connects the cluster to external networks, segments the cluster into subnets, and routes traffic. NSX also connects the Photon cluster to the virtual storage area network provided by VMware vSAN.

The Lightwave security system secures Photon Platform. An open source project published by VMware on GitHub, Lightwave authenticates and authorizes users as they work with the system and manage Kubernetes clusters. System administrators and DevOps can create security groups in Lightwave to control access to Photon Platform.

This architecture produces enterprise container infrastructure for IT administrators and development teams, infrastructure that is designed to securely run cloud-native applications in your own data center. Cloud-native applications are typically composed of microservices, packaged in containers, and scheduled to run dynamically in nodes on distributed cloud infrastructure.

Transforming ESXi Hosts into a Cluster for Running Containers

VMware ESXi is the hypervisor that is the foundation of a Photon Platform deployment. Photon Platform turns a set of ESXi hypervisors into a private cloud infrastructure that includes compute, storage, and networking. The cluster of hypervisors provides an elastic compute platform for running virtual machines, Docker containers, and Kubernetes clusters.

The Lightwave security service also runs on an ESXi hypervisor. Similarly, for storage, vSAN runs on one or more ESXi hypervisors, and NSX coordinates the vSwitches in a grouping of ESXi hypervisors to create virtual networks for Photon Platform.

Load Balancer

The Photon Platform installer automatically sets up HAProxy load balancers. A high-performance, reliable TCP/HTTP load balancer, HAProxy routes requests, API calls, user logins, and other connections among the Photon Controller management VMs as well as VMs running Lightwave. The load balancer ensures the high availability of Photon Platform's distributed components.

Photon Controller

Photon Controller furnishes the platform's distributed compute system. The infrastructure of Photon Controller contains three main components:

- A management plane that runs as virtual machines on ESXi hosts. The management plane administers tenants, allocates resources, controls projects, processes API requests, schedules VMs deployments from images, and creates Kubernetes clusters.
- An agent that handles communication between the Photon Controller management plane and ESXi hosts.
- A cloud host that resides on dedicated ESXi hosts to store VM images and run users' VMs.

The Management Plane

The load-balanced servers of the management plane share a persistent database, called the CloudStore, that holds state information and metadata for all objects—containers, clusters, VMs, disks, networks, physical hosts, and so forth.

The management plane contains the following subsystems:

- A distributed scheduler that places workloads according to their requirements. When a user creates a new VM, for example, the scheduler uses the API to request placement scores from ESXi hosts that meet certain restraints, such as the right availability zone. The scheduler then places the VM on the host with the best match for such resources as RAM, CPUs, and disks.
- An image management and replication service that helps ensure self-healing of the infrastructure.
- A decentralized control plane that provides a multi-version replicated document store so Photon can implement highly scalable components built as a durable collection of microservices.
- A RESTful API that lets you programmatically interact with the platform and automate operations. All the platform's functionality is exposed through the API. The API conforms with the OpenAPI standard, frequently referred to as Swagger. API calls can automate the provisioning of tenants, projects, virtual machines, Kubernetes clusters, and other resources.
- A web interface lets you connect to the load-balancer of the Photon Controller management plane to manage cluster resources, create tenants, allocate quotas, set up projects, provision virtual machines, and deploy Kubernetes clusters. Lightwave also includes a web interface for creating security groups and managing users.
- A command-line interface that you install on Linux, Mac, or Windows workstations to securely connect to the platform.

To extend these subsystems to all the ESXi hypervisors in the platform's cluster, the management plane communicates with the Photon Controller agent.

The Photon Agent

The Photon agent binds all the ESXi hosts into a cluster. The agent is installed on each ESXi hypervisor to handle inter-host communication and communication between the Photon Controller management plane and ESXi hosts.

KEY FEATURES OF PHOTON CONTROLLER

- API-driven
- Highly available
- Cloud scale
- Multitenant
- Secure
- Availability zones to isolate resources
- Flexibility for various workloads
- Kubernetes management

More specifically, the agent furnishes the API that Photon Platform uses to perform basic tasks on ESXi, such as creating a VM, uploading an image, deleting images, listing networks on ESXi, and supporting NSX. The agent also collects statistics and gives Photon Controller a placement score of its ESXi host so that Photon's scheduler can determine how to schedule and place resources.

The communication between Photon Controller and the agent is encrypted with SSL. When Photon is installed, Lightwave joins each ESXi host to the Lightwave domain and generates certificates for the ESXi hosts. The management plane then uses the certificates to authenticate the agent's calls.

The Cloud Host and the Image Datastore

Photon Platform uses two related concepts to describe where virtual machines are placed and where images are stored:

- The cloud host
- The image datastore

The Photon cloud host is a set of ESXi hypervisors dedicated to hosting virtual machines. You can, for example, tag an ESXi host as a cloud host and then specify that a given virtual machine run on it.

The Photon image datastore holds images for creating virtual machines, Kubernetes clusters, and other resources. The Photon image datastore uses native ESXi datastores or VMware vSAN as its storage system. Photon Controller's image management service replicates images across the datastores so that the images are highly available to eliminate downtime and to ensure performance in creating VMs and clusters. You can set up different image datastores and associate them with cloud hosts.

A datastore can, for example, contain an image for Ubuntu 14.04, and Photon Controller can use the Ubuntu image to fulfill a request for a virtual machine. The image datastore can also contain images for other Linux operating systems, such as Photon OS.

When you upload an image in Photon Controller, you mark it as either ON_DEMAND or EAGER. Photon Controller replicates an eager image to each image datastore when it is uploaded, making it immediately available across the system. In contrast, Photon Controller replicates an on-demand image to other cloud hosts' datastores only when a tenant creates a VM from it. An eager image produces VMs faster but consumes more storage space; an on-demand image produces VMs slower but takes less storage space.

Multitenancy

Photon Platform's hierarchical model of multitenancy uses tenants to segment business units, teams, and individuals. Multitenancy lets Photon Platform dynamically allocate resources for transient workloads, regulate the consumption of those resources, and control access to the API and administrative commands. Tenancy, however, does not isolate the infrastructure within a cluster, a function that is handled by availability zones, which are discussed later.

Photon Controller's multitenant model controls access and resources with the following types of users:

TYPE OF USER	ROLE
System administrators	The Photon Controller system administrator has rights to perform any action. They create tenants and establish the security group for each tenant.
Tenant administrators	A system administrator assigns at least one tenant administrator to each tenant. A tenant administrator can manage quotas, projects, and other objects under a tenant. A tenant administrator can also manage the security groups associated with the tenant.
Project users	Project users can view and modify project resources, including Kubernetes clusters, VMs, and disks. After a Photon Controller tenant administrator or system administrator creates a security group for a project, the group members are granted project user rights. Project users fall under a tenant.

Primitives in Photon Controller

As a multitenant system, Photon Controller abstracts physical resources so that it can allocate them across the system on behalf of users. Photon contains the following primitives for creating and managing infrastructure:

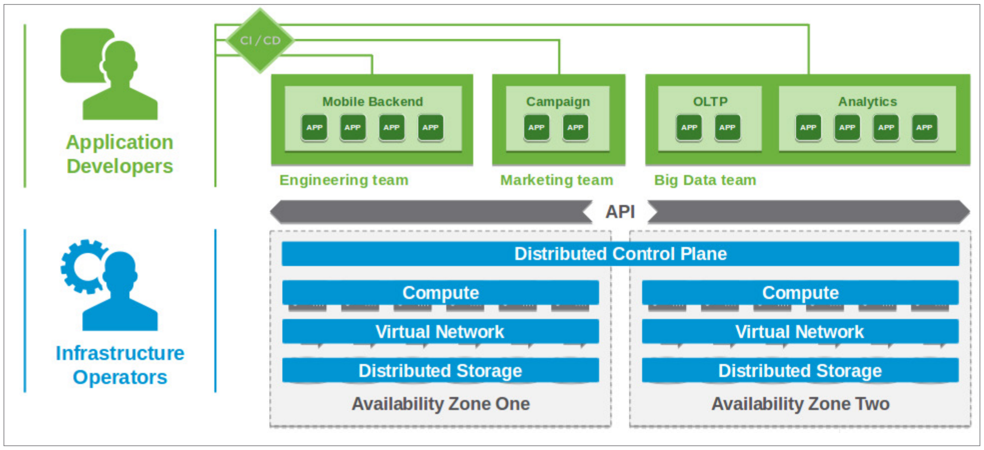
PRIMITIVE	DESCRIPTION
Tenants	When you create a tenant, you give it a pool of resources that the tenant's projects can consume. The pool is allocated through a quota.
Projects	A project is a way to allocate the quota assigned to a tenant. Tenant administrators create projects based on a quota and can impose project-specific limits on resource consumption. Project users can create VMs, disks, Kubernetes clusters, and other resources. A project falls under a tenant.
Quotas	A quota sets aside a limited amount of resources, such as CPUs, RAM, and disks, for use by a tenant or a project.
Flavors	A flavor defines how much a virtual piece of hardware costs. When you create a virtual machine, for example, you must select a flavor for it, and the flavor defines how much is subtracted for the VM and its disks from the allocation of resources specified by a quota. Flavors, in effect, provide self-created, standardized templates for managing resources at scale without bogging down project users in the details of hardware consumption. Photon Controller includes default flavors for creating Kubernetes clusters and other infrastructure.
Persistent Disks	Persistent disks endow virtual machines with resources for a project. You can attach a persistent disk to a VM, but after you discard the VM, the disk persists with the project.
Virtual machines	You create a virtual machine by specifying an image, a name, a flavor, and an ephemeral boot disk. Unlike persistent disks, which must be created before they can be attached to a VM, ephemeral disks are created along with a VM.

Availability Zones

Availability zones group ESXi hosts to isolate an application running on virtual machines, to provide data locality or physical affinity, to improve performance, or to ensure high availability.

A set of VMs running an application can be segmented into an availability zone to ensure that the VMs reside on the same physical rack or behind a switch for low latency. Inversely, you can place VMs on separate physical infrastructure to provide high availability. Similarly, availability zones can locate VMs on the hardware that provides the right level of performance for the application.

The following diagram illustrates the role of availability zones in the context of Photon Platform's model of multitenancy:



Lightwave Security Services

Photon Platform integrates with VMware Lightwave to provide security services for the platform and for Kubernetes clusters running on it. Lightwave is a cloud-scale, distributed security system for identity management, authentication, and certificate management. Lightwave comprises the following key elements:

- Directory service. This standards-based, multi-tenant, multi-master, highly scalable LDAP v3 directory service enables an enterprise's infrastructure to be used by the most-demanding applications as well as by multiple teams.
- Certificate authority. This directory-integrated certificate authority helps simplify certificate operations and key management across the infrastructure.
- Certificate store. This endpoint certificate store holds certificate credentials.
- Authentication. This cloud authentication service supports Kerberos, OAuth 2.0/ OpenID Connect, SAML, and WSTrust for interoperability with other standards-based technologies in the data center.

SECURITY SERVICES IN LIGHTWAVE

- Authentication and authorization.
- LDAP directory service.
- Certificate authority.
- Certificate store.

Controlling Access to Kubernetes Clusters

Lightwave enforces security on Kubernetes clusters by implementing OpenID Connect for authentication. Lightwave helps ensure that only authorized users can connect to Kubernetes clusters.

NSX

VMware NSX forms an integral part of Photon Platform. NSX delivers on-demand network virtualization without requiring reconfiguration of the underlying physical network. NSX not only streamlines network operations but also heightens app-dev agility through rapid, automated network provisioning.

Problems with Traditional Networking

Traditional networking services constrain the progress of software-defined approaches to servers and storage by requiring manual provisioning tied to established topologies and vendor-specific hardware. As a result, it hinders the development and deployment of a new application.

The problems associated with traditional networking plague the software-defined data center:

- Slow IT response times to new business requirements
- Downtime because of misconfiguration or security breaches
- A lack of workload mobility
- VLAN and firewall rule sprawl
- Performance bottlenecks associated with routing traffic to essential network services

Solving Old Problems with Network Virtualization

Network virtualization solves these problems. VMware NSX is a software-defined network that runs on hypervisors to reproduce traditional Layer 2 through Layer 7 networking services as logical switches, routers, and firewalls. In a Photon Platform cluster, NSX coordinates the vSwitches in ESXi hypervisors to create a virtual network. Because NSX is a non-disruptive solution that works with your existing physical network infrastructure—requiring only IP packet forwarding—you can programmatically assemble networking services to produce a virtual network that meets your requirements.

NSX Overview

NSX has three planes: management, control, and data. The planes run as distributed software modules on three kinds of nodes or clusters of nodes: manager, controller, and Edge transport nodes. The manager hosts various services. The controller deploys virtual networks across the NSX architecture. And the NSX Edge transport node connects virtual machines to logical switches by routing IP addresses and performing other IP services. On a transport node, NSX creates a hostswitch that binds logical router uplinks and downlinks to physical NICs.

A collection of transport nodes forms a transport zone—a fabric of logical switches that connects VMs to hypervisors. In NSX, a transport zone determines the extent of a Layer 2 network. An NSX Edge transport node can belong to one overlay transport zone and one or more VLAN transport zones. The overlay transport zone is for internal NSX tunneling between transport nodes. The VLAN transport zones are for the VLAN uplinks to the external physical network.

NSX Network Topology for Photon Platform

For Photon Controller, a single overlay transport zone is shared by all Photon Controller cloud hosts as well as the Edge node. The overlay transport zone provides L2 connectivity between VMs through a logical switch.

A single VLAN transport zone is also created. The purpose of this VLAN transport zone is to connect the logical network with the external, physical network. The Edge node joins both the overlay transport zone and VLAN transport zone.

The benefits of deploying Photon Platform with NSX include rapid application deployment with automated subnet provisioning. Combined with Lightwave security and Photon's model of multitenancy, NSX isolates development, test, and production environments on the same underlying physical infrastructure.

vSAN

Combining VMware vSAN with Photon Controller, Lightwave, and ESXi creates a powerful platform for cloud-native applications. vSAN establishes a hyper-converged, software-defined shared storage cluster that transforms the local physical resources of commodity hardware running ESXi into a distributed pool of storage for Photon Controller. The pool of storage can be seen as a distributed RAID cluster using either striping or mirroring.

The virtual storage network can support the storage needs of all the ESXi hosts and virtual machines running on Photon Platform, including the Photon management VMs, cloud hosts, and image stores. The Photon management VMs can reside inside or outside the vSAN cluster.

Optimizing Storage for Cloud-Native Workloads

vSAN is natively integrated with ESXi to eliminate dependencies on an external shared file system, such as NAS. With vSAN, the virtual machines running on Photon Platform can be assigned policies to establish predictable storage performance and availability for cloud-native workloads.

The flash-based storage and SSD caching of vSAN enhances performance by handling all the writes and nearly all the reads. Active data remains in flash storage to furnish a performance tier tuned for virtualized cloud-native workloads. In addition, vSAN lets you set up fault domains of ESXi hosts to protect against rack failure.

Integrating vSAN with Photon Platform

A vSAN storage cluster that is integrated with Photon Platform contains three or more x86 physical hosts with either a combination of magnetic disks and flash devices or all flash devices. The devices contribute cache and storage capacity to the distributed vSAN datastore to support as many as 6,400 VMs with more than 8 petabytes of storage.

The features of vSAN include efficient near-line deduplication, compression, and data-protecting erasure codes. IOPS limits help guarantee quality of service. For details, see VMware vSAN.

ESXI DATASTORE FORMATS ON PHOTON PLATFORM

- The default Virtual Machine File System (VMFS), a file system optimized for storing virtual machines.
- The Network File System (NFS), a file system on a network-attached storage system (NAS).
- VMware Virtual SAN (vSAN)—the preferred datastore for production deployments.

BENEFITS OF USING PHOTON OS

- Enables you to run Docker containers with ease.
- Improves resource-efficiency by using smaller server builds.
- Reduces security risks by removing vulnerable components.
- Decreases management effort by having fewer components to update.

Photon OS

Photon OS plays a role in Photon Platform. Photon OS is an open-source minimalist Linux operating system from VMware optimized for running containers. A central motivation behind the development of Photon OS is producing a Linux container host that delivers performance and security. The performance gains are acute when Photon OS runs on ESXi in Photon Platform.

Moving Containers from Development to Production

As part of VMware's cloud-native stack, Photon OS is a key element in a strategy to rapidly build and deploy containerized applications while continuing to fulfill such IT requirements as cost-effectiveness, security, and isolation.

Photon OS segregates and streamlines the process of moving containers from development to production by coming in a minimal version and a full version. Each version contains only the elements necessary to fulfill its use case for containers.

The minimal version of Photon OS is a fast, lightweight container host best suited to managing and hosting containers. The full version of Photon OS includes additional packages to help you create containerized applications; it is targeted at helping you develop, test, and package an application that runs in a container or in the cloud.

Curated Packages

VMware curates the packages in the Photon OS repositories. The packages of key clustering technologies—Kubernetes and Docker—are updated with the latest versions on a regular basis and signed with PGP signatures. The curation yields two results: An operating system primed with the latest open source projects for building cloud-native applications, and a suite of common components maintained to a secure standard.

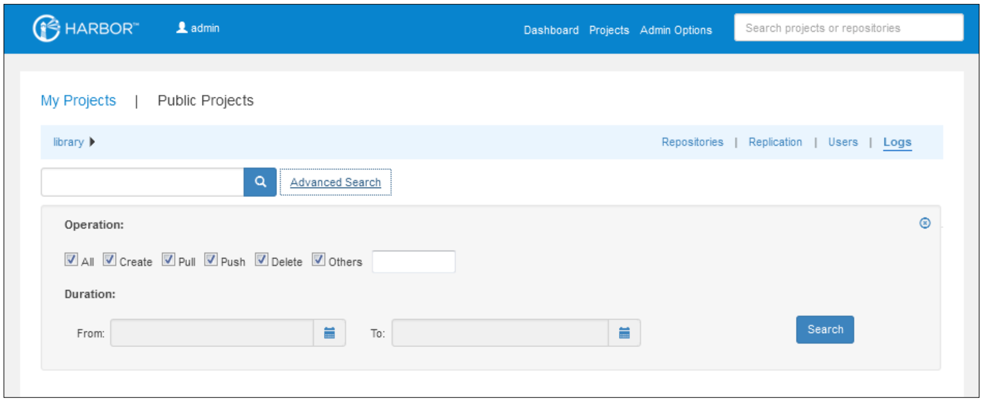
Docker Containers

Photon OS includes the open source version of Docker. With Docker and a hypervisor running on a laptop, a developer can replicate a cluster of virtual machines to build containerized applications with microservices on Photon OS. A microservice is an architectural pattern that isolates the code performing a business function into an independently deployable service. For these developers, a common use case is running Docker to accelerate the development and testing of their code. Photon OS expedites the work of getting a Docker engine running in a hypervisor.

By being optimized to run on VMware vSphere®, which often forms the basis of the production environment operated by IT, Photon OS paves the way to put containerized applications into production, especially when it runs as a virtual machine in Photon Platform.

Harbor Registry

Project Harbor is an open source, enterprise-class registry server from VMware that stores and distributes Docker images in a private registry behind your firewall. Harbor extends the open source Docker distribution by adding such functionality as security and management. Harbor can be set up with multiple registries, and images can be replicated across the registries.


HARBOR™
admin
Dashboard
Projects
Admin Options
Search projects or repositories
My Projects | Public Projects
library
Repositories | Replication | Users | Logs
Advanced Search
Operation:
☑ All ☑ Create ☑ Pull ☑ Push ☑ Delete ☑ Others
Duration:
From:
To:
Search

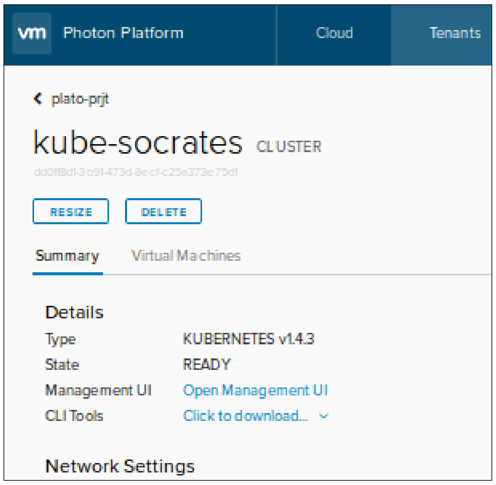
The graphical user interface of the Harbor portal.

Harbor provides a graphical portal, shown in the image above, and a RESTful API for managing repositories. Images are protected with role-based access control.

Harbor integrates with Lightwave to authenticate users. Harbor also tracks user interactions for auditing. On Photon Platform, Harbor can be deployed on Kubernetes.

Kubernetes

Photon Platform includes Kubernetes to deploy and manage the containers running a distributed application. You can horizontally scale cloud-native applications and orchestrate their containers to meet industrial-strength requirements. Photon Platform uses OpenID Connect for authentication with Kubernetes.



kube-socrates CLUSTER

dd0118d1-3b91-473d-8ecf-c25a373e75d1

Details

Type	KUBERNETES v1.4.3
State	READY
Management UI	Open Management UI
CLI Tools	Click to download...

Network Settings

The Photon Platform user interface showing a Kubernetes cluster.

For a demonstration of an application running on a Kubernetes cluster on Photon Platform, see Creating a Kubernetes Cluster.

Pivotal Cloud Foundry

Photon Platform is part of the Pivotal-VMware Cloud-Native Stack, an integrated platform-as-a-service solution that pairs Pivotal Cloud Foundry with Photon Platform to deliver a complete cloud-native software stack. The solution combines Pivotal's continuous innovation cloud platform with Photon's cloud-native infrastructure to accelerate development and deployment with built-in application and infrastructure management. For more information, see Pivotal-VMware Cloud-Native Stack.

Conclusion

The unique, hyper-converged architecture of Photon Platform transforms your virtualized infrastructure into an enterprise-ready container platform so that you can cost-effectively fulfill several high-level use cases in-house:

- Kubernetes as a service: Developers can deploy and resize highly available Kubernetes clusters to develop, test, and automate containerized applications.
- Infrastructure as a service: Photon Platform endows IT, developers, and DevOps with self-service access to such resources as VMs, disks, and networks so they can support high-churn workloads on demand.
- Platform as a service: Photon Platform integrates with Pivotal Cloud Foundry to build and scale applications for the cloud.
- Continuous Integration and Delivery. The simplicity of using Photon Platform for IaaS improves the CI/CD pipeline with uniformity and reusability, especially in environments with high churn.
- API-managed on-premises cloud: IT can deploy a vast cloud of virtual machines and automate their management through a RESTful API.
- Security: The Lightwave security service protects applications, virtual machines, Kubernetes clusters, and Photon Platform's own components.

The hyper-converged infrastructure of Photon Platform empowers you to run new applications with Docker containers and Kubernetes clusters while maintaining the security and performance of your data center.

VMware, Inc. 3401 Hillview Avenue Palo Alto CA 94304 USA Tel 877-486-9273 Fax 650-427-5001 www.vmware.com