Reference Architecture for Google Cloud’s Anthos 1.0 on Cisco HyperFlex Systems

Contents

Executive summary

Google Cloud's Anthos is a modern application-management and hybrid-cloud technology platform from Google Cloud. Anthos enables deployment of some of the main public-cloud capabilities in customers' own on-premises data centers. One of the core components of Anthos is the on-premises version of the popular Google Kubernetes Engine (GKE) container orchestrator, Anthos on VMware, which enables the development of modern applications based on microservices architecture. Customers have the flexibility to develop and test their workloads on-premises and later decide to deploy them: either on-premises or in the public cloud. Multiple different cloud providers support the use Kubernetes clusters. In addition to Anthos GKE, Anthos includes other software capabilities such as Anthos Service Mesh (service management), Anthos Config Management (policy management), Stackdriver (operation management), and Google Cloud Platform Marketplace and Cloud Run for Anthos (application development).

One of the main on-premises components required to deploy Anthos is an infrastructure platform. Cisco has partnered with Google Cloud to validate the Cisco HyperFlex™ hyperconverged infrastructure (HCI) platform for Anthos. Cisco HyperFlex systems with VMware are certified HCI servers with a comprehensive set of options configurable depending on application workload and business needs. Cisco HyperFlex HCI provides a common building block for Kubernetes-based applications as well as virtualized workloads. Cisco® infrastructure is optimized to run Anthos. When a Cisco HyperFlex or Cisco HyperFlex Edge system is used with Anthos, the technologies work together to provide a consistent experience whether in an on-premises environment, in Google Cloud, or in other public clouds (future). Anthos provides a ubiquitous platform that is consistent, secure, and reliable across environments.

This document provides a technical overview of the Anthos on-premises solution. The document describes the reference architecture for an Anthos on-premises deployment with Cisco HyperFlex HCI. It provides high-level implementation details.

The document also provides an example of a multitier application deployment to demonstrate the microservices architecture. The application used in the demonstration is composed of multiple microservices written in different languages that can talk to each other over gRPC, a popular open-source remote procedure call (RPC) developed by Google.

Business challenges

Organizations today want to accelerate innovation by developing and modernizing applications with microservices across the data center and remote, edge, and public-cloud environments to take advantage of emerging technologies. To achieve this, organizations need the flexibility to build a hybrid cloud that delivers agility and innovation. Configuring and managing Kubernetes across hybrid-cloud environments can be complicated, because it requires manual effort and multiple tools for clustering, networking, monitoring, security, etc.

The main technical challenges can be addressed with the following capabilities that this combined solution offers:

- Easily deploy Kubernetes-based applications in on-premises (including edge) environments using the Cisco HyperFlex hyperconverged platform.
- Establish a common and secure Kubernetes experience across the on-premises environment and public cloud.
- Provide a single administrative control plane for centralized policy and security across clouds.
- Extend existing on-premises applications and infrastructure into the public cloud without increasing risk and complexity.
- Increase application agility with access to industry-leading tools, technologies, and platforms from Google Cloud and Cisco.
- Enable access to a broader choice of applications from the cloud marketplace for on-demand consumption.

Technology overview

This section provides an overview of the hardware and software used for an Anthos on-premises deployment on Cisco HyperFlex infrastructure.

Cisco HyperFlex systems

Cisco HyperFlex systems are deployed as a preintegrated cluster with a unified pool of resources that can be quickly provisioned, adapted, scaled and managed. They are bundled with hybrid small-form-factor (SFF), large-form-factor (LFF), or all-flash storage configurations and a choice of management tools. Cisco HyperFlex systems include Cisco Unified Computing System™ (Cisco UCS®) M5 rack servers, based on second-generation Intel® Xeon® Scalable processors. These fifth-generation servers have faster processors, more cores, and faster and larger-capacity memory than previous-generation servers. In addition, they are ready for Intel 3D XPoint nonvolatile memory, which can be used as both storage and system memory, increasing your virtual server configuration options and flexibility for applications.

Physically, the system is delivered as a cluster of three or more Cisco HyperFlex HX240c M5 Nodes, HX240c M5 LFF Nodes, or HX240c M5 All Flash Nodes that are integrated into a single system by a pair of Cisco UCS 6200 or 6300 Series Fabric Interconnects.

Cisco HyperFlex HX240c M5 All Flash Node

This solution uses Cisco HyperFlex M5 all-flash servers. The Cisco HyperFlex HX240c M5 All Flash Node (Figure 1) is excellent for balanced-performance and capacity clusters.

Figure 1.
Cisco HyperFlex HX240c All Flash Node: Front with bezel

The HX240c M5 All Flash servers extend the capabilities of the Cisco HyperFlex portfolio in a 2-rack-unit (2RU) form factor with the addition of the second-generation Intel Xeon Scalable family processors, supporting up to 28 cores per CPU, 24 DIMM slots with configuration options ranging from 128 GB to 3 TB of DRAM (with 128-GB DIMMs), and an all-flash footprint of cache and capacity drives for highly available, high-performance storage.

For more information about the HX240c M5 All Flash Node, see https://www.cisco.com/c/dam/en/us/products/collateral/hyperconverged-infrastructure/hyperflex-hx-series/datasheet-c78-736784.pdf.

Cisco HyperFlex HX Data Platform software

Cisco HyperFlex HX Data Platform is a hyperconverged software appliance that transforms Cisco servers into a single pool of computing and storage resources. It eliminates the need for network storage and tightly integrates with VMware vSphere and its existing management application to provide a seamless data management experience. In addition, native compression and deduplication reduce storage space occupied by the virtual machines.

Cisco HyperFlex systems deliver a new generation of flexible, scalable, enterprise-class hyperconverged solutions. This solution also delivers storage efficiency features such as thin provisioning, data deduplication, and compression for greater capacity and enterprise-class performance. Additional operational efficiency is facilitated through features such as cloning and snapshots. The HX Data Platform can be administered through a VMware vSphere web client plug-in or through the HTML5-based native Cisco HyperFlex Connect management tool.

HX Data Platform consists of the following components:

- **Cisco HyperFlex HX Data Platform Installer:** Download this installer to a server connected to the storage cluster. The HX Data Platform Installer configures the service profiles and policies within Cisco UCS Manager, deploys the controller virtual machines, installs the software, creates the storage cluster, and updates the VMware vCenter plug-in.
- **Storage controller virtual machine:** The HX Data Platform Installer installs the storage controller virtual machine on each converged node in the managed storage cluster.
- **Cisco HyperFlex HX Data Platform Plug-in:** This integrated VMware vSphere interface monitors and manages the storage in your storage cluster.

Cisco Intersight platform for Cisco HyperFlex systems

The Cisco Intersight™ platform simplifies data center operations by delivering systems management as a service, alleviating the need to maintain islands of on-premises management infrastructure.

The Cisco Intersight platform provides an installation wizard to install, configure, and deploy Cisco HyperFlex clusters, attached with Cisco HyperFlex Edge and Cisco UCS fabric interconnects. The wizard constructs a preconfiguration definition of your cluster called a Cisco HyperFlex cluster profile. This definition is a logical representation of the Cisco HyperFlex HX-Series nodes in your Cisco HyperFlex cluster. It includes these features:

- **Security:** Credentials for the Cisco HyperFlex cluster such as the controller virtual machine password and the hypervisor user name and password
- **Configuration:** Server requirements, firmware, etc.
- **Connectivity:** Upstream network, virtual network, etc.

The main features and benefits of the Cisco Intersight platform include the following:

- Unified management
  - Simplify Cisco UCS and Cisco HyperFlex management with a single management platform.
  - Scale across data centers and remote locations without additional complexity.
- Configuration, provisioning, and server profiles
  - Create multiple server profiles with just a few clicks or through the available API, automating the provisioning process.
  - Create, deploy, and manage your Cisco HyperFlex configurations.
  - Help ensure consistency and eliminate configuration drift, maintaining standardization across many systems.
- Inventory information and status
  - Display and report inventory information for Cisco UCS and Cisco HyperFlex systems.
  - Monitor Cisco UCS and Cisco HyperFlex server alerts and health status across data centers and remote locations.

  - View your Cisco HyperFlex configurations.
  - Track and manage firmware versions across all connected Cisco UCS and Cisco HyperFlex systems.
  - Track and manage software versions and automated patch updates for all claimed Cisco UCS Director software installations.
- Enhanced support experience
  - Get automated alerts about failure notifications.
  - Automate the generation and forwarding of technical support files to the Cisco Technical Assistance Center (TAC) to accelerate the troubleshooting process.
- Open API
  - The representational state transfer (REST) API supports the Open API Specification (OAS) to provide full programmability and deep integration of systems.
  - The Python and PowerShell software development kits (SDKs) enable integration with DevOps and IT operations management (ITOM) tools.
- Seamless integration and upgrades
  - Upgrades are available for Cisco UCS, Cisco HyperFlex, and Cisco UCS Director systems software running supported firmware and software versions.
  - Upgrades to the Cisco Intersight platform are delivered automatically without disruption of your operations.

Cisco UCS Manager

Cisco UCS Manager is embedded software that resides on a pair of fabric interconnects and provides complete configuration and management capabilities for Cisco HyperFlex HX-Series servers. The most common way to access Cisco UCS Manager is to use a web browser to open the GUI. Cisco UCS Manager supports role-based access control (RBAC).

A critical benefit of Cisco UCS Manager is its application of stateless computing. Each node in an HX-Series cluster has no set configuration. MAC addresses, universally unique IDs (UUIDs), firmware, and BIOS settings, for example, are all configured on Cisco UCS Manager in a service profile and applied uniformly to all the HX-Series servers. This approach enables consistent configuration and configuration that can easily be reused.

Cisco UCS fabric interconnects

Cisco UCS fabric interconnects provide the management and communication backbone for the Cisco HyperFlex HX-Series rack-mount servers and Cisco UCS B-Series Blade Servers and Cisco UCS 5100 Series Blade Server Chassis.

Cisco UCS 6300 Series Fabric Interconnects support high-performance, low-latency, lossless, line-rate 40 Gigabit Ethernet, with up to 2.56 Tbps of switching capacity. Backward compatibility and scalability are assured with the capability to configure 40-Gbps Quad Small Form-Factor Pluggable (QSFP) ports as breakout ports using four 10 Gigabit Ethernet breakout cables. Existing Cisco UCS servers with 10 Gigabit Ethernet interfaces can be connected in this manner, although Cisco HyperFlex nodes must use a 40 Gigabit Ethernet virtual interface card (VIC) adapter to connect to a Cisco UCS 6300 Series Fabric Interconnect.

Cisco UCS virtual interface cards

The Cisco UCS VIC 1385 card works with Cisco Nexus® Family 40 and 10 Gigabit Ethernet switches for high-performance applications. The Cisco UCS VIC 1385 implements the Cisco Data Center Virtual Machine Fabric Extender (VM-FEX), which unifies virtual and physical networking into a single infrastructure. The extender provides virtual machine visibility from the physical network and a consistent network operations model for physical and virtual servers. The modular LAN-on-motherboard (mLOM) slot can be used to install a Cisco UCS VIC without consuming a PCI Express (PCIe) slot, thus providing greater I/O expandability.

The Cisco UCS VIC 1387 is a dual-port Enhanced QSFP (QSFP+) 40-Gbps Ethernet and Fibre Channel over Ethernet (FCoE)–capable PCIe mLOM adapter installed in the HX-Series rack servers. The VIC 1387 is used in conjunction with the Cisco UCS 6332 or 6332-16UP Fabric Interconnect.

Cisco Nexus Family switches

Cisco Nexus 9000 Series Switches can scale to up to 30 Tbps of nonblocking performance with latency of less than 5 microseconds, 1152 x 10-Gbps or 288 x 40-Gbps nonblocking Layer 2 and Layer 3 Ethernet ports, and wire-speed Virtual Extensible LAN (VXLAN) gateway, bridging, and routing capabilities.

VMware vCenter management

Cisco HyperFlex systems use VMware vCenter–based management. The vCenter Server is a data center management server application developed to monitor virtualized environments. The HX Data Platform is also accessed from the preconfigured vCenter Server to perform all storage tasks. vCenter supports shared storage features such as VMware vMotion, Distributed Resource Scheduler (DRS), High Availability (HA), and vSphere replication. More scalable, native HX Data Platform snapshots and clones replace VMware snapshots and cloning capabilities.

You must have vCenter installed on a separate server to access HX Data Platform. vCenter is accessed through the vSphere Client, which is installed on the administrator's laptop or PC.

F5 BIG-IP Virtual Edition

F5 BIG-IP Virtual Edition (VE) is a virtual application delivery controller (vADC) that can be deployed on all leading hypervisors and cloud platforms running on commodity servers. BIG-IP VE delivers all the same market-leading application delivery services—including advanced traffic management, acceleration, Domain Name System (DNS), firewall, and access management—that run on F5 purpose-built hardware. VE software images are downloadable and portable among on-premises virtualized data center, public-cloud, and private-cloud environments. With BIG-IP VE, you can rapidly provision consistent application services across the data center and into the cloud.

F5 BIG-IP Local Traffic Manager (LTM) helps you deliver your applications to your users in a reliable, secure, and optimized way. With BIG-IP LTM, you have the power to simplify, automate, and customize application services faster and more predictably. BIG-IP LTM enables you to control network traffic, selecting the right destination based on server performance, security, and availability.

F5 BIG-IP integrates with Anthos on-premises, making it the recommended choice (Figure 2). F5 application services can be readily applied to containers running within Anthos. Using F5 Container Ingress Services, which integrates natively with the Kubernetes orchestrator, and BIG-IP, customers can deliver F5's broad suite of traffic management and security services to their containers while easily orchestrating application services insertion. On-premises container workloads can now be easily moved and scaled across Anthos GKE, while F5 works with Kubernetes to help ensure that your on-premises applications receive the advanced services they require.

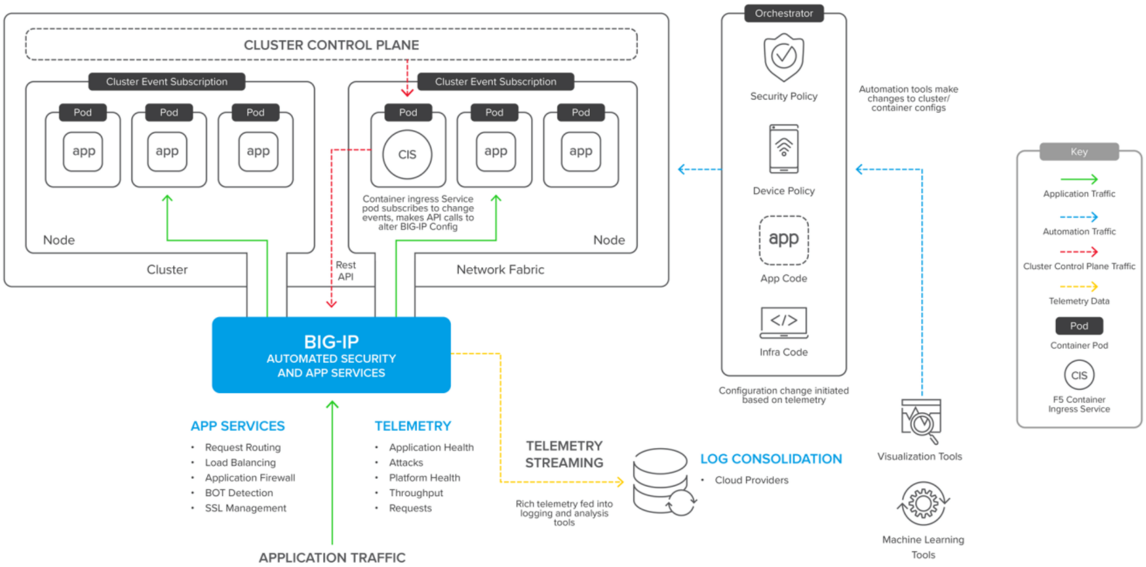


Figure 2.
F5 BIG-IP for Anthos on VMware

Anthos on VMware

Anthos on VMware is hybrid-cloud software that brings GKE to on-premises data centers. With Anthos on VMware, you can create, manage, and upgrade Kubernetes clusters in your on-premises environment through the Anthos cloud-based management plane.

Anthos helps accelerate application development by bringing your code into production reliably, securely, and consistently with low risk, helping enable your business strategically.

Figure 3 shows the main components of Anthos GKE. In the figure you can see that the components running on Google Public Cloud are mainly the same as those that are running on the on-premises cloud. Hence, the Anthos clusters running in the on-premises data center are essentially an extension of the public cloud. This is why Anthos’ approach is unique and the best in its class. As soon as the Anthos on-premises cluster is deployed and connected to Google Cloud Platform, the hybrid cloud is ready for operation.

GKE Connect connects the on-premises Kubernetes clusters or the Kubernetes clusters running on other public clouds with the Google Cloud. GKE Connect uses an encrypted connection between the Kubernetes clusters and Google Cloud Platform. It enables authorized users to log in to clusters; access details about their resources, projects, and clusters; and manage cluster infrastructure and workloads independent of the hardware they are running. The GKE Connect Agent is installed in the on-premises Kubernetes cluster. It is authenticated and then the encrypted connection with Google Cloud Platform is established without any public IP address. The Google Cloud Platform Virtual Private (VPC) and Google Cloud Interconnect provide secure connectivity and controls to enable Google Cloud Platform to be an extension of the on-premise cloud.

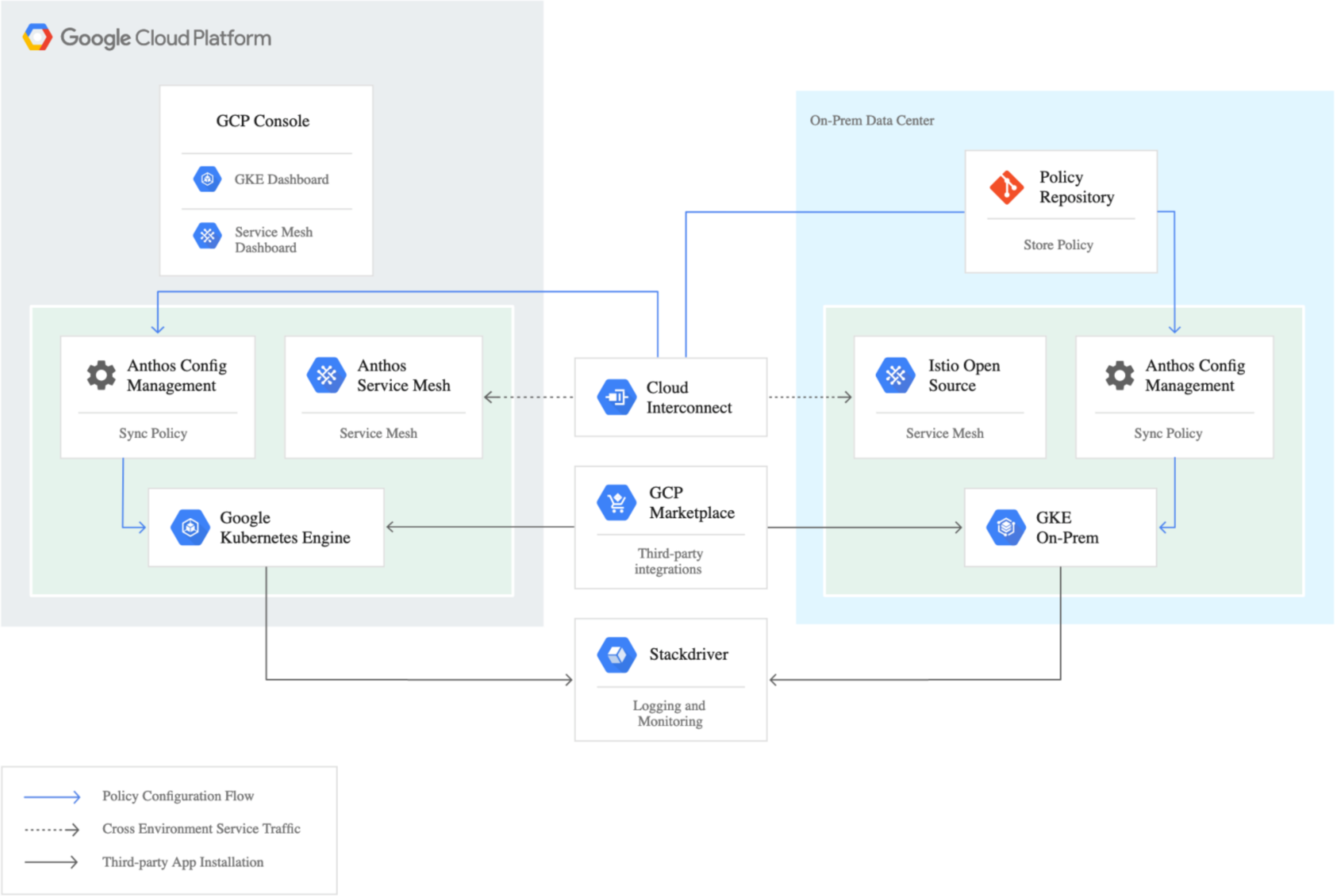


Figure 3.
Anthos GKE architecture

Solution overview

The primary computing environment for Anthos relies on Anthos GKE to manage Kubernetes installations in both cloud and on-premises environments in which you intend to deploy your applications. These offerings bundle upstream Kubernetes releases and provide management capabilities for creating, scaling, and upgrading conformant Kubernetes clusters. It is a software-based service that enables developers to run and manage their containerized applications in both hybrid and multicloud environments, even if those environments include cloud solutions from Microsoft or Amazon. Because Anthos focuses primarily on enabling containerized workloads running on Kubernetes, a hybrid cloud implemented with Anthos enables cross-cloud orchestration of containers and microservices across the Kubernetes clusters. Migration of container workloads between on-premises and GKE clusters running in other clouds can be achieved through a public container registry such as Google Container Registry (GCR), or through your own private registries that are secured with your central identity and access control to allow only authenticated users or service accounts to push or pull container images in the registry. You do not need to convert container images running on-premises to run on GKE. Anthos, by contrast, is more of a blank canvas on which you can install anything that runs in a container, including applications from the Google Cloud Platform marketplace. Figure 4 illustrates the Anthos platform.

Hybrid-cloud implementations tend to be complex because of the integration required between the on-premises data center and public-cloud data centers. Concerns may arise about network and Internet security, complex bidirectional traffic routing, data access, provisioning requirements, etc. Hence, hybrid-cloud implementations tend to require several third-party tools and services depending on the specific capabilities expected. Anthos simplifies hybrid clouds by providing the necessary tools and services for a secure and scalable hybrid-cloud implementation. Hybrid clouds enable organizations to take advantage of public-cloud capabilities while retaining some parts of their application landscape and data within their data centers. Anthos allows organizations to selectively migrate workloads to public-cloud providers such as Google Cloud Platform, Amazon Web Services (AWS), and Microsoft Azure.

Another important benefit of the Anthos technology is the capability to see and manage applications across multiple cloud providers, as well as internally, on any private clouds through a single pane. Again, this capability provides more flexibility to organizations that are still working through their hybrid-cloud and multicloud strategies.

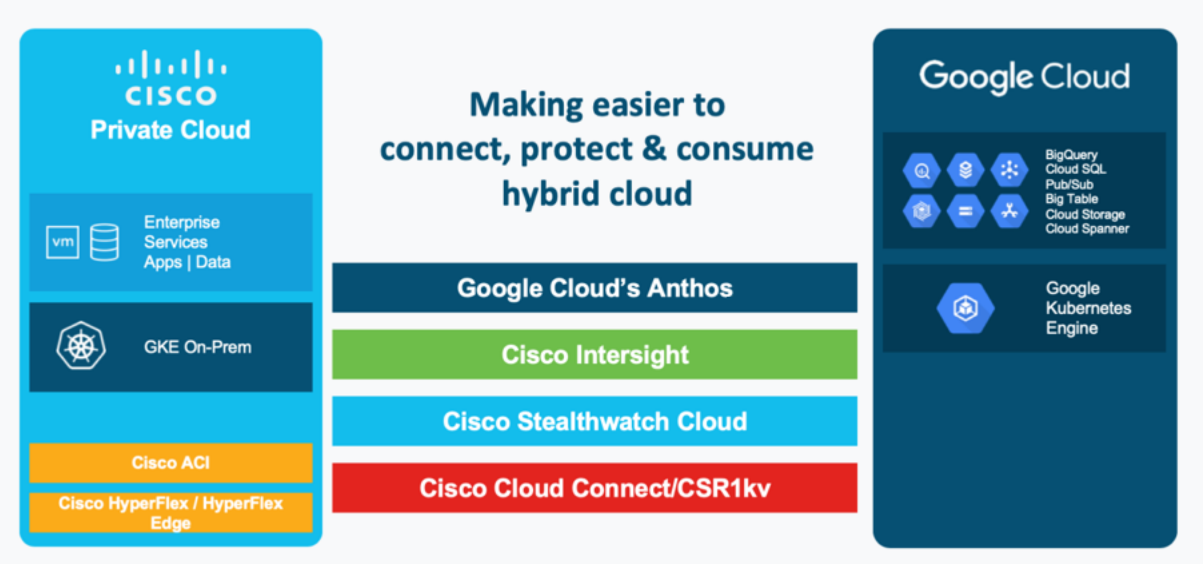


Figure 4.
Anthos overview

Anthos on Cisco HyperFlex systems reference architecture

The solution discussed here uses a Cisco HyperFlex cluster consisting of four Cisco HyperFlex HX240c M5SX All Flash Nodes. This Cisco HyperFlex cluster is a fully contained virtual server platform with computing and memory resources, integrated networking connectivity, a distributed high-performance log-based file system for virtual machine storage, and hypervisor software for running the virtualized servers, all within a single Cisco UCS management domain.

For upstream network connections, also referred to as northbound network connections, Cisco Nexus 9000 Series Switches are used. These are connected to the customer data center network at the time of installation.

Table 1 lists the components used in this solution.

Table 1. Cisco HyperFlex system components

Components	Required hardware
Fabric interconnects	2 Cisco UCS 6332-16UP Fabric Interconnects
Servers	4 Cisco HyperFlex HX240c M5SX All Flash rack-mount servers
Upstream switches	2 Cisco Nexus 9396PX Switches

For server specifications and more information, see the Cisco HyperFlex HX240c M5SX All Flash Node specification sheet, at https://www.cisco.com/c/dam/en/us/products/collateral/hyperconverged-infrastructure/hyperflex-hx-series/hxaf-240c-m5-specsheet.pdf.

Table 2 lists the hardware options for one Cisco HyperFlex HX240c M5SX All Flash server.

Table 2. Cisco HyperFlex HX240c M5SX All Flash Node server configuration

Cisco HyperFlex HX240c M5SX All Flash Node options	Hardware required
Processors	2 Intel Xeon Gold 6240 CPUs with 18 cores each
Memory	12 x 16 GB = 192 GB
Disk controllers	Cisco 12-Gbps modular SAS controller
Solid-state disks (SSDs)	• 1 x 240-GB 6-Gbps SATA SSD for housekeeping tasks • 6 x 960-GB 6-Gbps SATA SSDs for capacity tier • 1 x 400-GB SAS SSD for caching tier
Network	• 1 x Cisco UCS VIC 1387 mLOM • 1 x Cisco UCS VIC 1385 PCIe card
Boot device	1 x 240-GB M.2 form-factor 6-Gbps SATA SSD

Table 3 lists the software components and the versions required for the Cisco HyperFlex system.

Table 3. Software components

Component	Software required
VMware hypervisor	VMware ESXi 6.5.0 U3 13932383
VMware vCenter management server	VMware vCenter Server 6.5 8307201
Cisco HyperFlex HX Data Platform	Cisco HyperFlex HX Data Platform 4.0.1b
Cisco UCS firmware	Cisco UCS infrastructure: Cisco UCS C-Series bundle 4.0.4d
Anthos on VMware - Admin Appliance	gke-on-prem-admin-appliance-vsphere-1.0.11.ova
Anthos on VMware - VMware Installer	gke-onprem-vsphere-1.0.10-full.tgz
GKE bundle	GKE bundle 1.12.7-gke.19
F5 load balancer	F5 BIG-IP VE 13.1.3-0.0.6.ALL-scsi.ova

Figure 5 shows the physical topology of Anthos on a Cisco HyperFlex system.

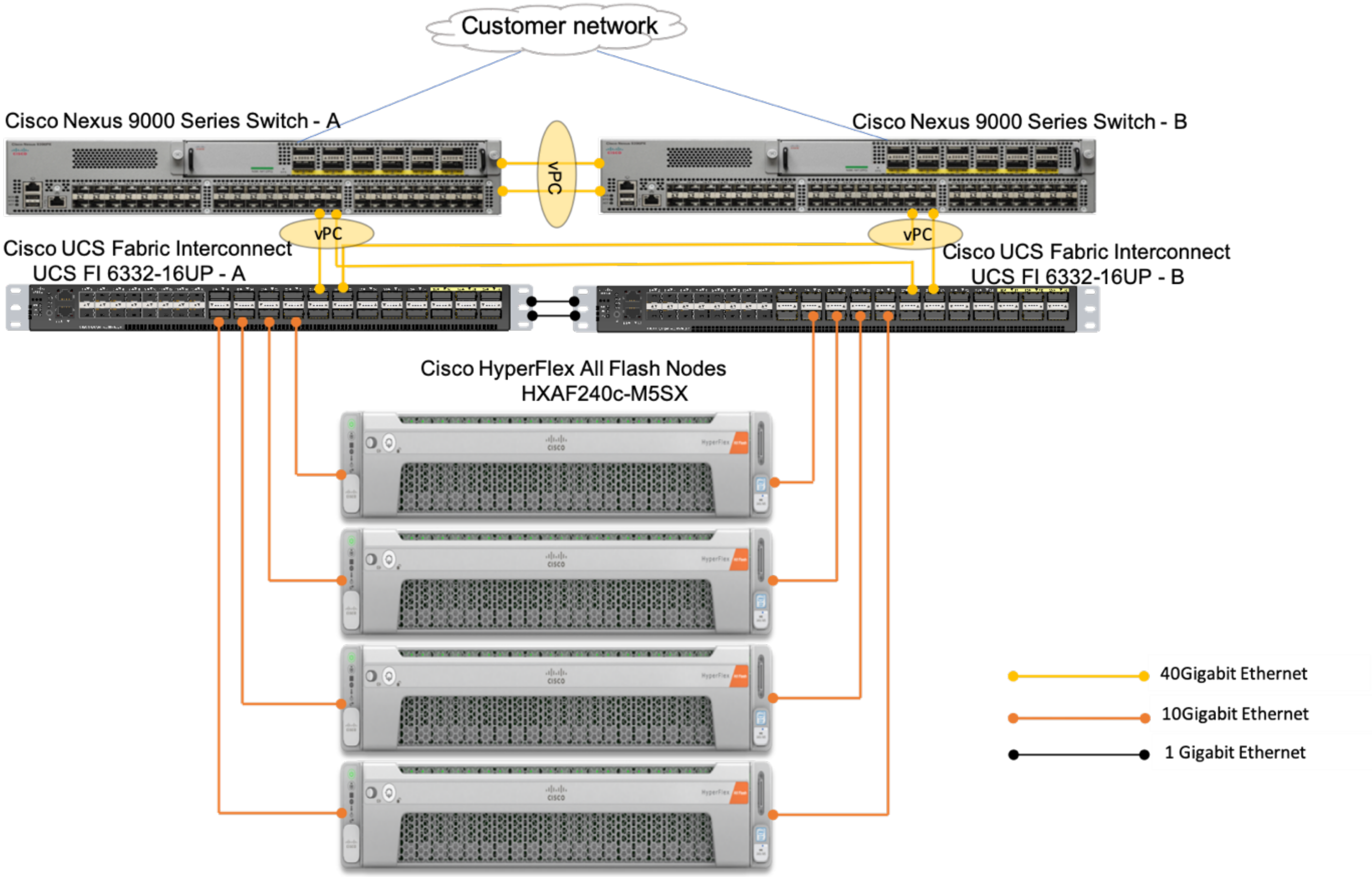

Customer network
Cisco Nexus 9000 Series Switch - A
Cisco Nexus 9000 Series Switch - B
vPC
vPC
vPC
Cisco UCS Fabric Interconnect
UCS FI 6332-16UP - A
Cisco UCS Fabric Interconnect
UCS FI 6332-16UP - B
Cisco HyperFlex All Flash Nodes
HXAF240c-M5SX
40Gigabit Ethernet
10Gigabit Ethernet
1 Gigabit Ethernet

Figure 5.
Physical topology of Anthos on Cisco HyperFlex system

From a logical perspective, traffic flows into and out of the on-premises system using an F5 BIG-IP LTM virtual appliance. The F5 BIG-IP LTM creates dynamic connections between the computing nodes and the external network interfaces. Currently, three special-purpose networks are used for management, internal, and external traffic. Figure 6 shows several different networks configured and connected for Anthos on the Cisco HyperFlex platform:

- Internal management network
- Internal virtual machine network
- External virtual machine network

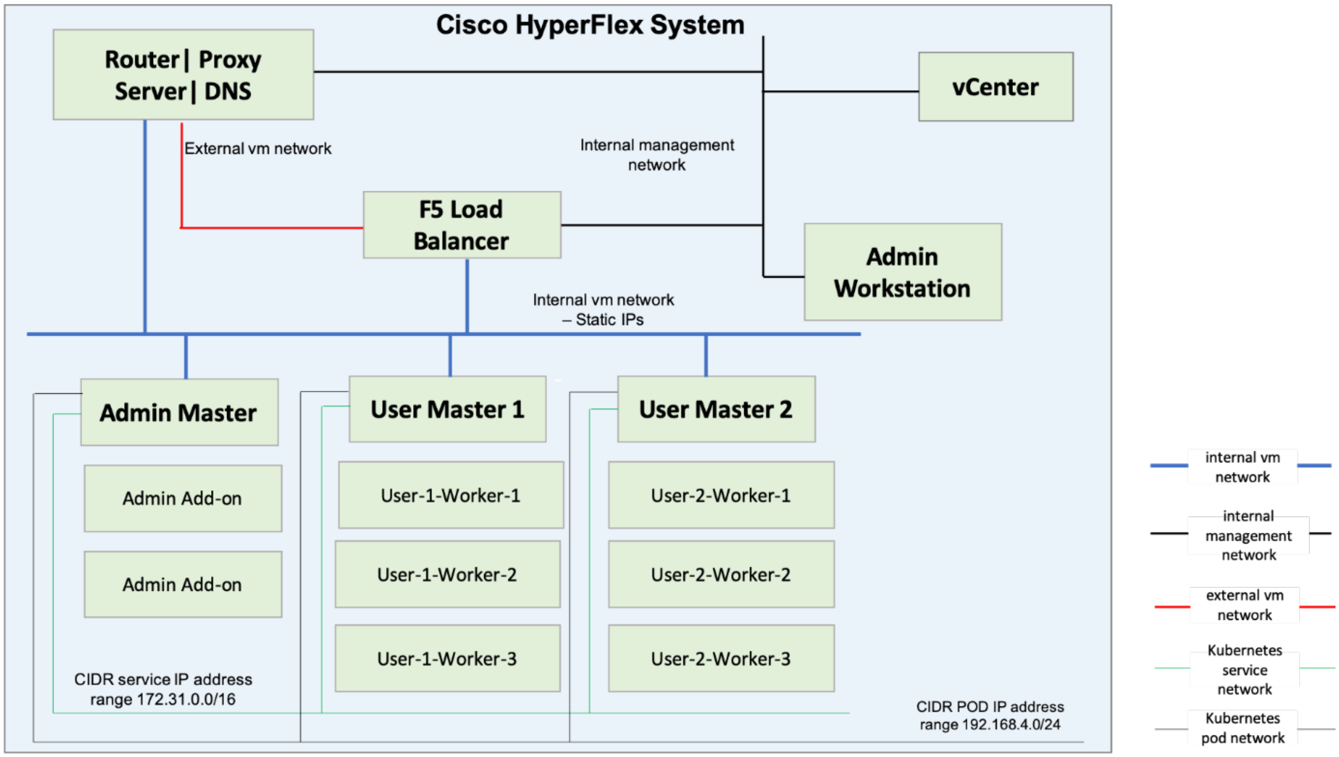


Figure 6.
Logical network topology of Anthos on Cisco HyperFlex system

Create a new computing cluster with at least one computing server and one resource pool in vCenter to host the workloads in vCenter. This cluster requires DRS as well as one resource pool.

The vCenter Server hosts multiple virtual machines in the new resource pool. These comprise a virtual Anthos on VMware computing cluster. The application workloads are processes that run on one of the Anthos on VMware computing cluster virtual machines. When an application is deployed to the Anthos on VMware computing cluster, it runs on the Anthos on VMware cluster virtual machines. If additional workload capacity is required, the Anthos on VMware computing cluster is expanded using the gkectl command-line utility or the Kubernetes cluster API.

Note: This solution assumes that you have VMware vCenter Server installed or have deployed the VMware vCenter Server Appliance.

Here is a summary of the correlation between vSphere virtual machines and Anthos on VMware cluster servers:

- Anthos on VMware cluster is a collection of vSphere 6.5 virtual machines.
- Anthos on VMware cluster's performance profile (RAM, CPU, storage, etc.) is the sum of all vSphere 6.5 virtual machines running in the Anthos on VMware cluster.
- Multiple Anthos on VMware clusters can exist in a single vCenter deployment.

Prerequisites for deploying Anthos on VMware on Cisco HyperFlex system

To deploy Anthos on the Cisco HyperFlex system, your setup must meet the following prerequisites:

- Fully configured and working Cisco HyperFlex infrastructure
- Google account, Google Cloud Platform service account, and a billing-enabled Google Cloud Platform project
- VMware vSphere ESXi 6.5 (Update 3) and VMware vCenter 6.5
- One VMware Virtual Machine File System (VMFS) datastore with 2-TB capacity
- Capability to create required DNS entries
- Anthos on VMware Admin Appliance (OVA file) Version gke-on-prem-admin-appliance-vsphere-1.1.0-gke.6.ova
- Official Anthos on VMware Installer (OVA file) Version 1.0.1-gke.5
- GKE bundle Version 1.12.7-gke.19 downloaded
- F5 BIG-IP VE Layer 4 LTM version 13.1.3 virtual appliance installed and licensed reserve IP addresses (Dynamic Host Configuration Protocol [DHCP] or static) in three networks
- Network access to Google Cloud (googleapis.com)

Before configuring the F5 BIG-IP load balancer, you need to reserve IP addresses for the Anthos on VMware admin control plane and each of the user cluster control planes. Figure 6 earlier in this document shows the master and user cluster details.

The following are required for a Anthos on VMware installation:

- The admin cluster needs N + 3 IP address pools for the control plane, control-plane add-ons, and control-plane ingress, where N is the number of user clusters.
- Each user cluster control plane needs three IP addresses for control-plane, ingress-controller, and cluster upgrades.
- Reserve Classless Inter-Domain Routing (CIDR) blocks for the pod IP address range and service IP address range.

Anthos on VMware deployed on Cisco HyperFlex system

This document does not discuss the full deployment of Anthos. However, deployment can be broken down into the following high-level steps, discussed here:

1. Bring up the Cisco HyperFlex cluster.
2. Deploy and configure F5 BIG-IP LTM.
3. Deploy Anthos on VMware admin workstation virtual machine on the Cisco HyperFlex system.
4. Create the Anthos on VMware admin master cluster and user cluster on the Cisco HyperFlex system.
5. Install Anthos on VMware on the Cisco HyperFlex system.
6. Create a second user cluster on the Cisco HyperFlex system (as needed).
7. Register the Anthos on VMware user cluster with a Google Cloud Platform project.

Installing and launching Cisco HyperFlex system

Installing the Cisco HyperFlex system is done primarily through a deployable Cisco HyperFlex installer virtual machine, available for download as an OVA file at Cisco.com. The installer virtual machine performs most of the Cisco UCS configuration work, and you can use it to simplify the installation of VMware ESXi on the Cisco HyperFlex hosts. The installer virtual machine also performs significant portions of the ESXi configuration. You also can use the installer virtual machine to install the Cisco HyperFlex HX Data Platform software and create the Cisco HyperFlex cluster.

For details about installing the Cisco HyperFlex system, see https://www.cisco.com/c/en/us/td/docs/hyperconverged_systems/HyperFlex_HX_DataPlatformSoftware/Installation_VMWare_ESXi/4_0/b_HyperFlexSystems_Installation_Guide_for_VMware_ESXi_4_0.html.

Cisco HyperFlex system post-installation task: Checking Cisco HyperFlex Connect HTML 5 management webpage

After you have installed the Cisco HyperFlex system, you can use a new HTML 5–based web user interface as the primary management tool for Cisco HyperFlex systems (Figure 7). Through this centralized point of control for the cluster, administrators can create volumes, monitor data platform health, and manage resource use. Administrators also can use this data to predict when the cluster needs to be scaled. To use the Cisco HyperFlex Connect user interface, connect using a web browser to the Cisco HyperFlex cluster IP address: http://<hx controller cluster ip>.

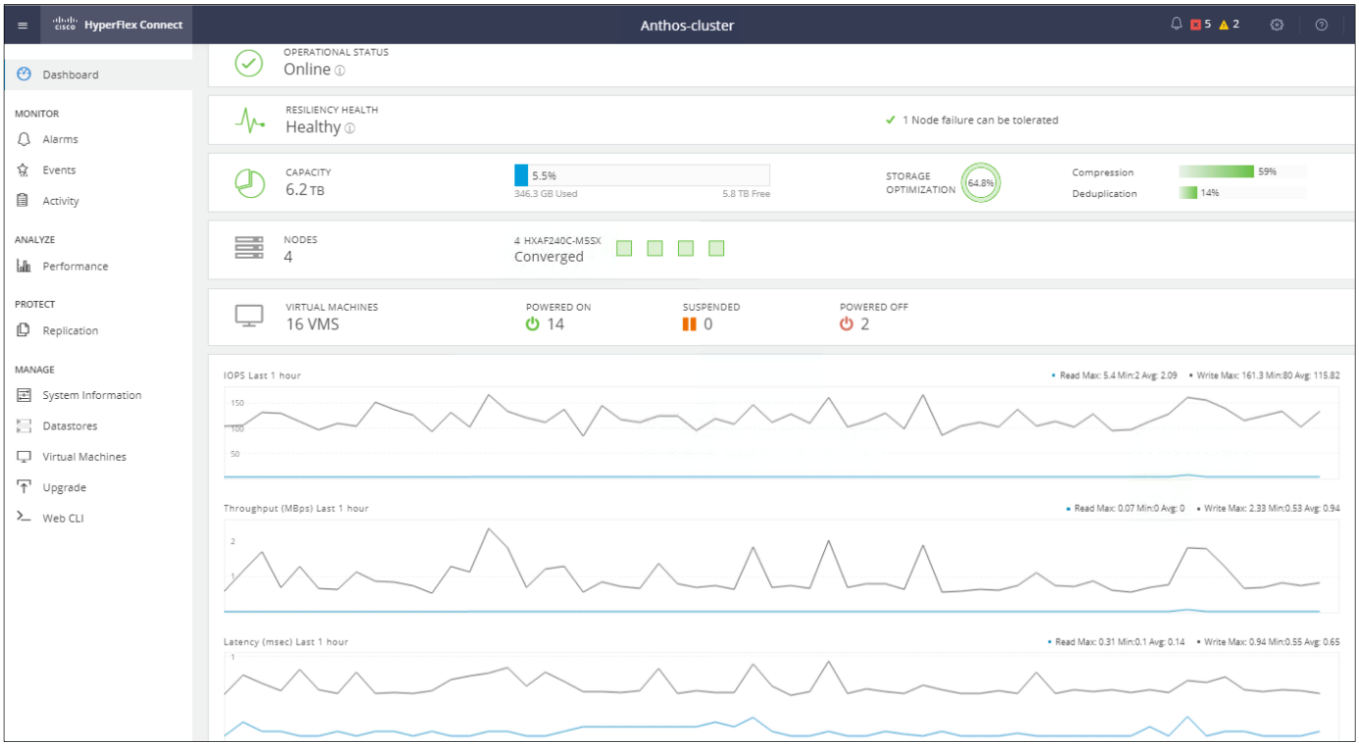
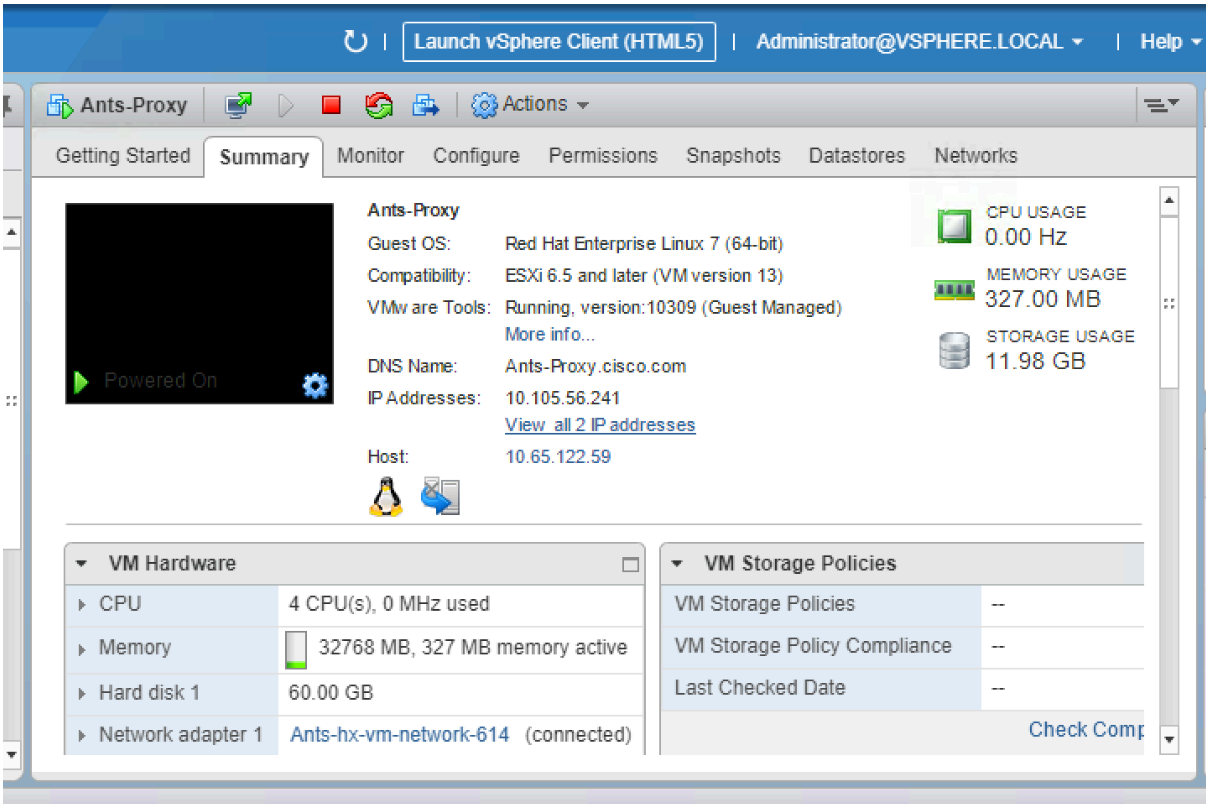


Figure 7.
Cisco HyperFlex Connect dashboard

1. After the Cisco HyperFlex system is installed and accessible, log in to the VMware vSphere web client to access vCenter on the Cisco HyperFlex cluster: https://<vSphere IP address>.
2. Click Home and the select Hosts & Clusters icon. Hosts & Clusters displays the connected Cisco HyperFlex and ESX nodes and the storage controller virtual machines running on the nodes. Never modify these storage controller virtual machine configurations because otherwise cluster support may be lost.
3. Right-click the cluster and choose New Resource Pool.
4. Right-click the HyperFlex Cluster and choose New Virtual Machine to create the virtual machine.
5. On the screen that appears, select Create New Virtual Machine and click Next at the bottom of the screen. This virtual machine provides DNS services to Anthos GKE. Also, this solution uses the same virtual machine for running Squid proxy services.
6. On the next screen, enter the name of the virtual machine and choose the vCenter data center that was created. Then click Next.

7. On the next screen, choose the computing resource and click Next.
8. Select the datastore for the virtual machine that was created with the Cisco HyperFlex installation. Click Next.
9. Choose the compatibility option “ESXi 6.5 and later” and click Next.
10. For the OS family choose Linux, and for the OS version choose RHEL 7.6. Click Next.
11. Choose the appropriate CPU, memory resources, and disk space.
12. Choose the network—for example, internal vm network—used to allow virtual machine accessibility in the network.
13. Select Datastore ISO File and choose the RHEL 7.6 ISO file from the drop-down menu.
14. Select the Connect check box.
15. Review the settings carefully and make sure the entries are correct to create a virtual machine.

When the OS installation is complete, you will see the virtual machine configured as shown in Figures 8 and 9.



Ants-Proxy

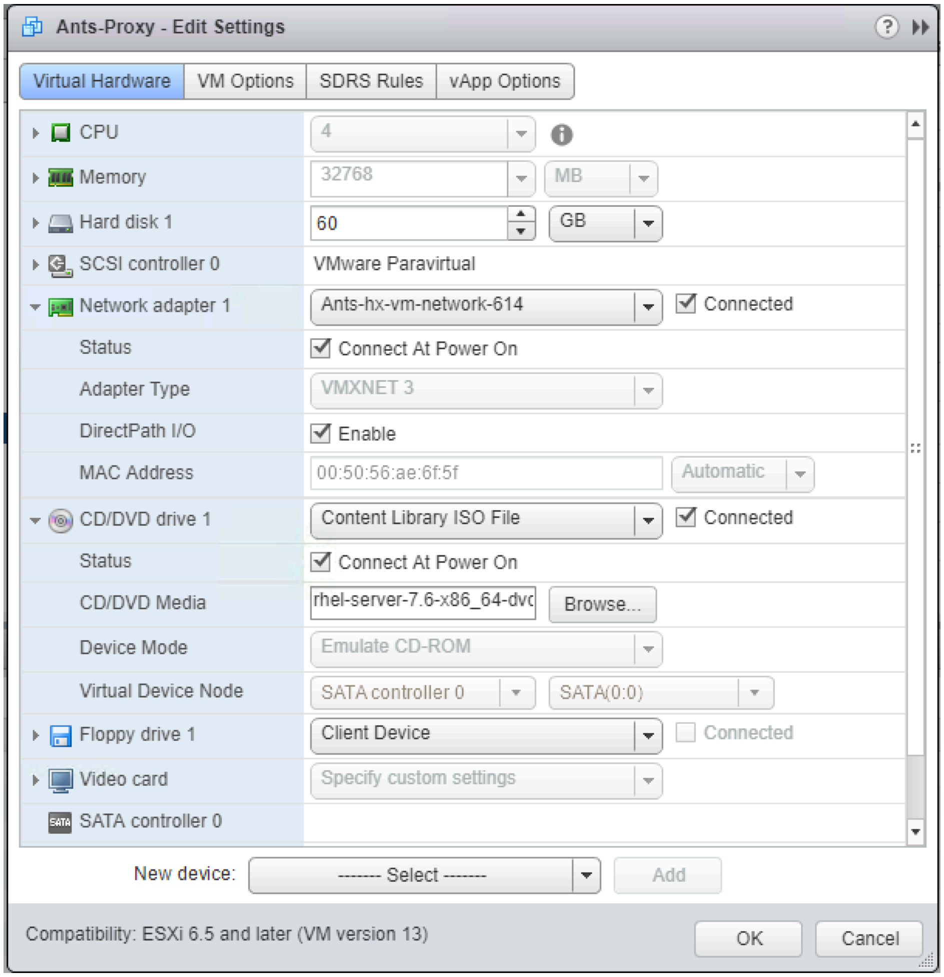
Guest OS:	Red Hat Enterprise Linux 7 (64-bit)
Compatibility:	ESXi 6.5 and later (VM version 13)
VMware Tools:	Running, version:10309 (Guest Managed)
DNS Name:	Ants-Proxy.cisco.com
IP Addresses:	10.105.56.241
Host:	10.65.122.59

CPU USAGE 0.00 Hz
MEMORY USAGE 327.00 MB
STORAGE USAGE 11.98 GB

VM Hardware	
CPU	4 CPU(s), 0 MHz used
Memory	32768 MB, 327 MB memory active
Hard disk 1	60.00 GB
Network adapter 1	Ants-hx-vm-network-614 (connected)

VM Storage Policies	
VM Storage Policies	--
VM Storage Policy Compliance	--
Last Checked Date	--

Figure 8.
Virtual machine configuration (1)



Ants-Proxy - Edit Settings

Virtual Hardware | VM Options | SDRS Rules | vApp Options

CPU	4	
Memory	32768	MB
Hard disk 1	60	GB
SCSI controller 0	VMware Paravirtual	
Network adapter 1	Ants-hx-vm-network-614	☑ Connected
Status	☑ Connect At Power On	
Adapter Type	VMXNET 3	
DirectPath I/O	☑ Enable	
MAC Address	00:50:56:ae:6f:5f	Automatic
CD/DVD drive 1	Content Library ISO File	☑ Connected
Status	☑ Connect At Power On	
CD/DVD Media	rhel-server-7.6-x86_64-dvd	Browse...
Device Mode	Emulate CD-ROM	
Virtual Device Node	SATA controller 0	SATA(0:0)
Floppy drive 1	Client Device	□ Connected
Video card	Specify custom settings	
SATA controller 0		

New device: ------- Select ------- Add

Compatibility: ESXi 6.5 and later (VM version 13)

OK Cancel

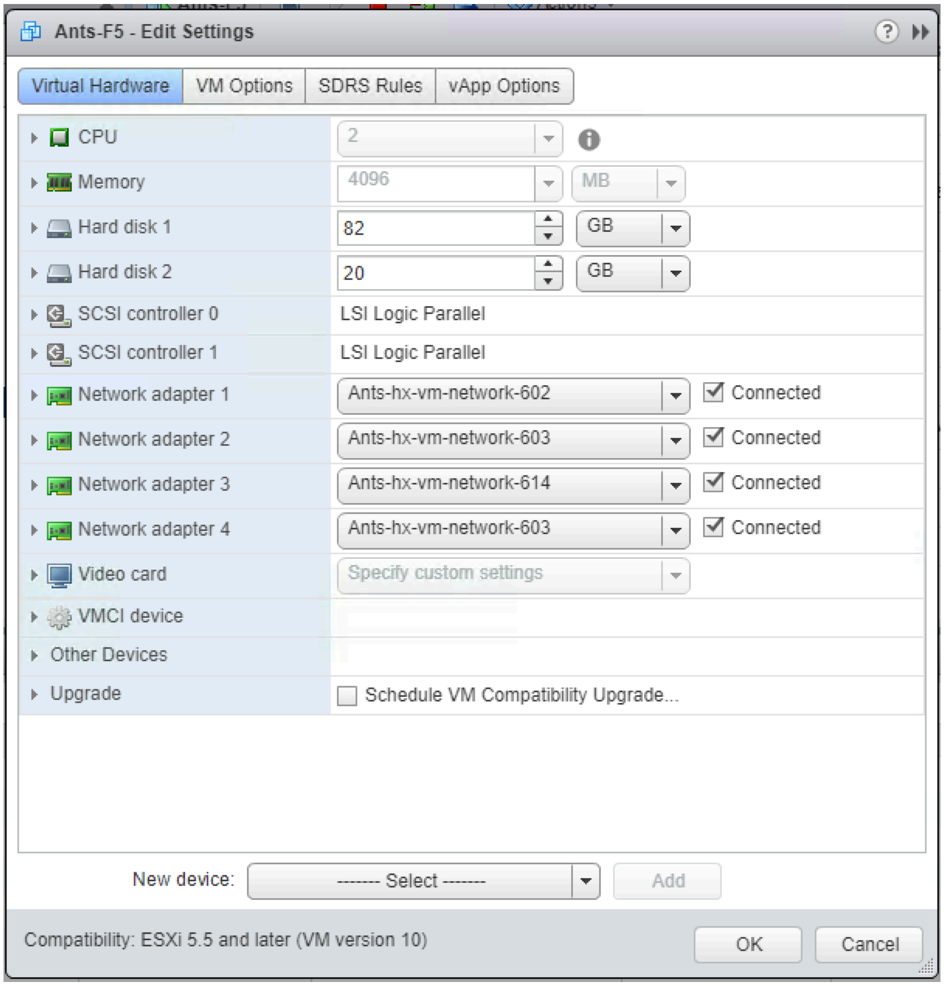
Figure 9.
Virtual machine configuration (2)

Note: This virtual machine acts as a DNS and proxy server to the Anthos on VMware cluster.

Deploying and configuring F5 BIG-IP load balancer

By default, Anthos on VMware integrates with F5 BIG-IP Versions 12.0 and 13.0. Download the F5 BIG IP VE 13.1.3 OVA file from the F5 Download site. Refer to the section “Downloading the installation image” at: https://support.f5.com/csp/article/K13117 - download.

The downloaded OVA file has multiple network interface card (NIC) templates by default. Using this OVA file, create a virtual machine (follow the steps to create a virtual machine as described in the previous section) and enter the configuration details through the wizard. Make sure that your network is configured with management, internal, and external IP addresses. Your F5 virtual machine configuration should look similar to Figure 10.



Ants-F5 - Edit Settings

Virtual Hardware | VM Options | SDRS Rules | vApp Options

CPU	2	
Memory	4096	MB
Hard disk 1	82	GB
Hard disk 2	20	GB
SCSI controller 0	LSI Logic Parallel	
SCSI controller 1	LSI Logic Parallel	
Network adapter 1	Ants-hx-vm-network-602	☑ Connected
Network adapter 2	Ants-hx-vm-network-603	☑ Connected
Network adapter 3	Ants-hx-vm-network-614	☑ Connected
Network adapter 4	Ants-hx-vm-network-603	☑ Connected
Video card	Specify custom settings	
VMCI device		
Other Devices		
Upgrade	□ Schedule VM Compatibility Upgrade...	

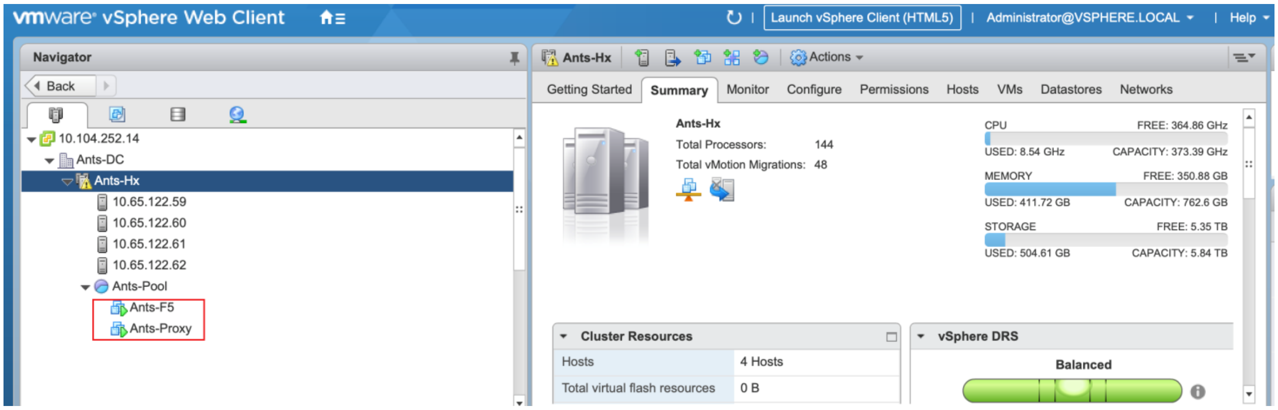
New device: ------- Select ------- Add

Compatibility: ESXi 5.5 and later (VM version 10)

Figure 10.
F5 virtual machine configuration

For details about configuring the F5 load balancer, see https://cloud.google.com/solutions/partners/installing-f5-big-ip-adc-for-gke-on-prem

You will now see two virtual machines in vCenter: one for DNS and proxy service (in this example, Ants-Proxy), and another for the F5 load balancer (in this example, Ants-F5), as shown in Figure 11.



- 10.104.252.14
  - Ants-DC
    - Ants-Hx
      - 10.65.122.59
      - 10.65.122.60
      - 10.65.122.61
      - 10.65.122.62
      - Ants-Pool
        - Ants-F5
        - Ants-Proxy

Ants-Hx
Total Processors: 144
Total vMotion Migrations: 48

CPU FREE: 364.86 GHz
USED: 8.54 GHz CAPACITY: 373.39 GHz
MEMORY FREE: 350.88 GB
USED: 411.72 GB CAPACITY: 762.6 GB
STORAGE FREE: 5.35 TB
USED: 504.61 GB CAPACITY: 5.84 TB

Cluster Resources

Hosts	4 Hosts
Total virtual flash resources	0 B

vSphere DRS
Balanced

Figure 11.
Two virtual machines created

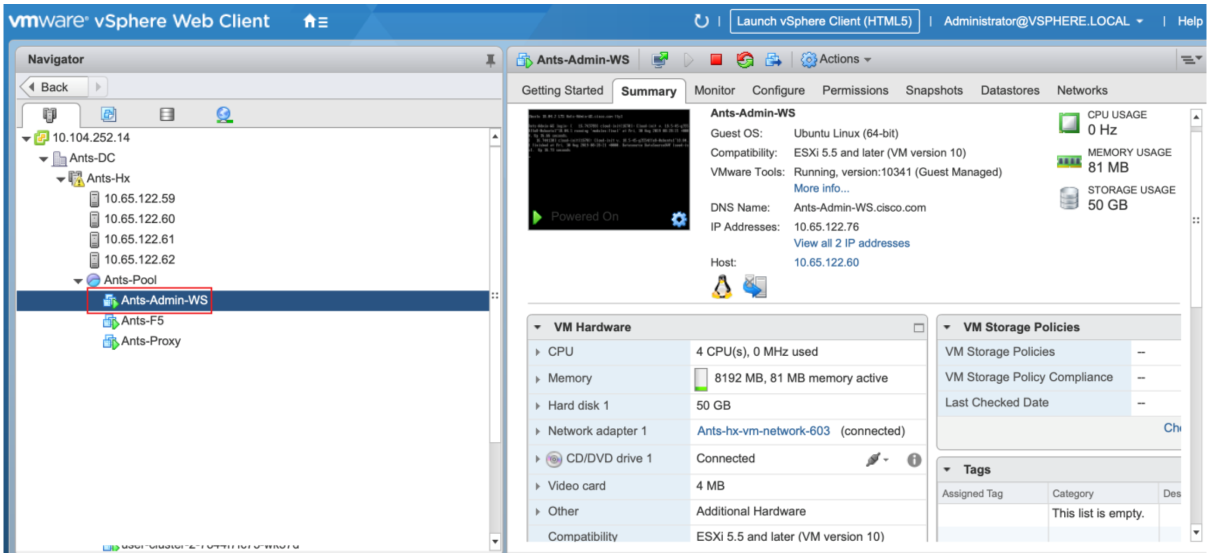
Deploying Anthos on VMware admin workstation virtual machine on Cisco HyperFlex system

The admin workstation virtual machine acts as the deployment host for all GKE clusters on the premises.

Refer to the section “Deploy the admin OVA” in the Google Anthos on VMware doc at https://cloud.google.com/gke-on-prem/docs/how-to/installation/admin-workstation.

High-level configuration steps for setting up the admin workstation virtual machine include installation of the Google Cloud SDK, govc (the VMware vSphere command-line interface [CLI], and HashiCorp Terraform 0.11. After downloading the admin workstation OVA file, use Terraform to deploy the virtual machine to vSphere.

After the installation is complete, you will see admin workstation virtual machine on vSphere Web Client as shown in Figure 12.



Ants-Admin-WS

Guest OS:	Ubuntu Linux (64-bit)
Compatibility:	ESXi 5.5 and later (VM version 10)
VMware Tools:	Running, version:10341 (Guest Managed) More info...
DNS Name:	Ants-Admin-WS.cisco.com
IP Addresses:	10.65.122.76 View all 2 IP addresses
Host:	10.65.122.60

CPU USAGE 0 Hz
MEMORY USAGE 81 MB
STORAGE USAGE 50 GB

VM Hardware

CPU	4 CPU(s), 0 MHz used
Memory	8192 MB, 81 MB memory active
Hard disk 1	50 GB
Network adapter 1	Ants-hx-vm-network-603 (connected)
CD/DVD drive 1	Connected
Video card	4 MB
Other	Additional Hardware
Compatibility	ESXi 5.5 and later (VM version 10)

VM Storage Policies

VM Storage Policies	--
VM Storage Policy Compliance	--
Last Checked Date	--

Tags

Assigned Tag	Category	Des
	This list is empty.	

Figure 12.
Admin workstation virtual machine configuration

Preparing admin workstation for Anthos on VMware installation on Cisco HyperFlex system

The admin cluster provides the administrative control plane for all on-premises clusters and for connectivity to Google Cloud. In contrast, the user cluster serves as the Kubernetes cluster for running user workloads.

Before proceeding with Anthos on VMware cluster deployment, complete the following tasks.

Set up private registry

If your setup is behind a web proxy, you must configure a private registry for storing images locally. The Anthos on vmware virtual machines that are created as part of Anthos on VMware deployment do not receive the proxy details, and hence images required for Anthos on VMware deployment cannot be pulled externally into these virtual machines. Therefore, you need to prepopulate the private registry with the required images.

This task requires you to configure Docker deamon environment variables using the steps at https://docs.docker.com/config/daemon/systemd/ - httphttps-proxy.

Note: Be sure that you change the Docker systemd service file (not the demon.json file).

Also, Anthos on VMware does not support insecure Docker registries. When you start your Docker registry, you must provide a certificate and a key. The certificate can be signed by a public certificate authority (CA), or it can be self-signed.

To set up a private registry, refer to the section “Creating Docker Registry” at https://cloud.google.com/gke-on-prem/docs/how-to/installation/private-registry.

Configure static IP addresses

If you want to use static IP addresses to assign IP addresses to Anthos on VMware virtual machines, you need to create two separate YAML files on the admin workstation:

- One hostconfig file for the admin cluster
- One hostconfig file for the user cluster

Each host file should have its own separate set of unique IP addresses, using the formula N + 4, where N is the number of user clusters that you plan to create.

For more information about creating static IP addresses, refer to https://cloud.google.com/gke-on-prem/docs/how-to/installation/static-ips.

Installing Anthos on VMware on Cisco HyperFlex system

The instructions in this section guide you through the process of creating an admin cluster and one user cluster.

Google Cloud Platform projects form the basis for creating, enabling, and using all Google Cloud Platform services, including managing APIs, enabling billing, adding and removing collaborators, and managing permissions for Google Cloud Platform resources.

To whitelist your google account, create a project ID and Google Cloud Platform service account. For details, see https://cloud.google.com/resource-manager/docs/creating-managing-projects - creating_a_project. When this process is complete, you should see a project space created for your name and ID in Google Cloud Platform.

Make sure that you provide roles appropriately on Google Cloud Platform. A role is a group of permissions that you can assign to members. See https://cloud.google.com/gke-on-prem/docs/how-to/installation/getting-started.

Tables 4 and 5 list the hardware requirements for the virtual machines that are created as part of the GK On-Prem installation.

Table 4. Admin master cluster minimum requirements

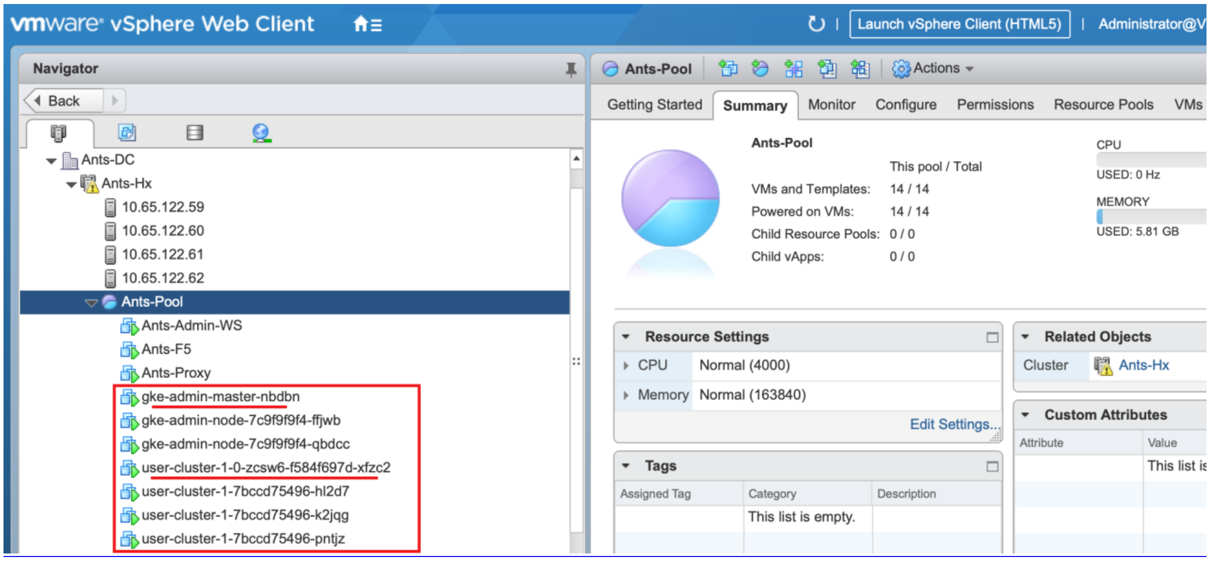
Name	Specifications	Purpose
Admin master cluster	• 4 virtual CPUs (vCPUs) • 16,384 MB of RAM • 40 GB of hard-disk space	Runs the admin control plane
Add-on virtual machines	2 virtual machines running with the following specifications: • 4 vCPUs • 16,384 MB of RAM • 40 GB of hard-disk space	Runs the admin control plane add-ons
User master cluster	• 4 vCPUs • 8192 MB of RAM • 40 GB of hard-disk space	Each user cluster has its own control plane. User control plane virtual machines run in the admin cluster.

Table 5. User master cluster minimum requirements

Name	Specifications	Purpose
User cluster worker nodes	• 4 vCPUs • 8192 MB of RAM • 40 GB of hard-disk space	A user cluster node is a virtual machine on which workloads run. When you create a user cluster, you decide how many nodes it should run. The configuration required for each node depends on the workloads you run.

Follow the installation procedure described at https://cloud.google.com/gke-on-prem/docs/how-to/installation/install.

After the Anthos on VMware installation is complete, you will see the master cluster virtual machines and user cluster virtual machines in vCenter (Figure 13)



- Ants-DC
  - Ants-Hx
    - 10.65.122.59
    - 10.65.122.60
    - 10.65.122.61
    - 10.65.122.62
    - Ants-Pool
      - Ants-Admin-WS
      - Ants-F5
      - Ants-Proxy
      - gke-admin-master-nbdbn
      - gke-admin-node-7c9f9f9f4-ffjwb
      - gke-admin-node-7c9f9f9f4-qbdcc
      - user-cluster-1-0-zcsw6-f584f697d-xfzc2
      - user-cluster-1-7bccd75496-hl2d7
      - user-cluster-1-7bccd75496-k2jqg
      - user-cluster-1-7bccd75496-pntjz

Ants-Pool

	This pool / Total
VMs and Templates:	14 / 14
Powered on VMs:	14 / 14
Child Resource Pools:	0 / 0
Child vApps:	0 / 0

CPU
USED: 0 Hz
MEMORY
USED: 5.81 GB

Resource Settings	
CPU	Normal (4000)
Memory	Normal (163840)

Related Objects	
Cluster	Ants-Hx

Attribute	Value
	This list is

Assigned Tag	Category	Description
	This list is empty.	

Figure 13.
Master and user cluster virtual machines in VMware vCenter

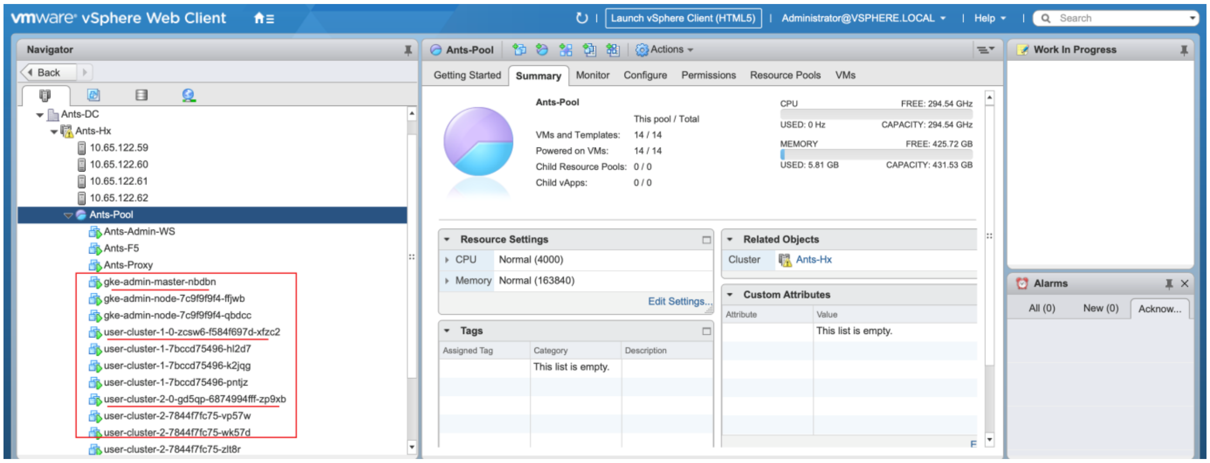
Creating a second user cluster on Cisco HyperFlex system

To create additional user clusters, you make a copy of the Anthos on VMware configuration file used to deploy your clusters. You modify the copied file to meet your expectations for the new user clusters, and then you use the file to create the cluster.

You need to copy and modify a Anthos on VMware configuration file for each additional user cluster you want to create. Make sure that you delete the admin master cluster block from the configuration file when you want to add a new user cluster.

To add a new user cluster, follow the procedure described at https://cloud.google.com/gke-on-prem/docs/how-to/administration/creating-additional-user-clusters.

After the Anthos on VMware installation is complete and the user clusters are ready, you will see all the virtual machines in vCenter (Figure 14).



Ants-DC
Ants-Hx
10.65.122.59
10.65.122.60
10.65.122.61
10.65.122.62
Ants-Pool
Ants-Admin-WS
Ants-F5
Ants-Proxy
gke-admin-master-nbdbn
gke-admin-node-7c9f9f9f4-ffjwb
gke-admin-node-7c9f9f9f4-qbdcc
user-cluster-1-0-zcsw6-f584f697d-xfzc2
user-cluster-1-7bccd75496-hl2d7
user-cluster-1-7bccd75496-k2jqg
user-cluster-1-7bccd75496-pntjz
user-cluster-2-0-gd5qp-6874994fff-zp9xb
user-cluster-2-7844f7fc75-vp57w
user-cluster-2-7844f7fc75-wk57d
user-cluster-2-7844f7fc75-zlt8r

Ants-Pool

	This pool / Total
VMs and Templates:	14 / 14
Powered on VMs:	14 / 14
Child Resource Pools:	0 / 0
Child vApps:	0 / 0

CPU FREE: 294.54 GHz
USED: 0 Hz CAPACITY: 294.54 GHz
MEMORY FREE: 425.72 GB
USED: 5.81 GB CAPACITY: 431.53 GB

Resource Settings

CPU	Normal (4000)
Memory	Normal (163840)

Related Objects

Cluster	Ants-Hx

Tags

Assigned Tag	Category	Description
	This list is empty.	

Custom Attributes

Attribute	Value
	This list is empty.

Figure 14.
Virtual machines in VMware vCenter

Registering your user clusters in Google Cloud Platform

The Anthos on VMware installation automatically deploys the GKE Connect Agent and registers the cluster with Google Cloud Platform. GKE Connect Agent is a Kubernetes deployment that establishes a long-lived connection between the cluster and Google Cloud Platform. Connect Agent makes it possible to log in to clusters from Google Cloud Platform. After the clusters are registered automatically, you need to create and assign the node-reader role to see the details of the cluster nodes. To create and assign the node-reader role, you need to create a YAML file with the cluster role. For details, see: https://cloud.google.com/anthos/multicluster-management/console/logging-in

After the Anthos on VMware cluster has been registered with Google Cloud Platform, the Kubernetes cluster nodes are visible from the Google Cloud Platform console. Navigate to Kubernetes Engine > Clusters to view the clusters registered Figures 15 and 16).

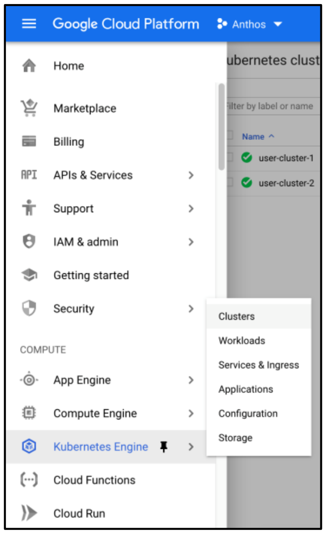
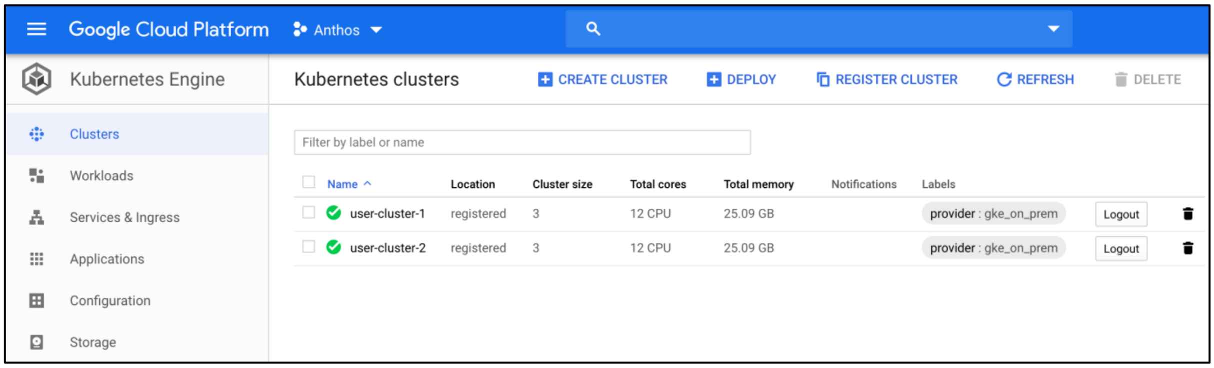


Figure 15.
Navigation path to view registered clusters



Name	Location	Cluster size	Total cores	Total memory	Notifications	Labels		
✓ user-cluster-1	registered	3	12 CPU	25.09 GB		provider : gke_on_prem	Logout	🗑
✓ user-cluster-2	registered	3	12 CPU	25.09 GB		provider : gke_on_prem	Logout	🗑

Figure 16.
View the registered clusters

Deployment configuration example

After the Anthos on VMware cluster has been deployed and registered, you can easily deploy applications from Google Cloud Platform Marketplace or any other applications.

Hipster Shop multitier application deployment

As an example, you will see how to deploy a web-based e-commerce application called Hipster Shop that users can use to browse items, add them to a cart, and purchase them (Figure 17). This application works on any Kubernetes cluster, as well as GKE. It is easy to deploy, with little to no configuration. This application is developed by Google and is used to demonstrate technologies such as GKE, Istio, Stackdriver, and others.

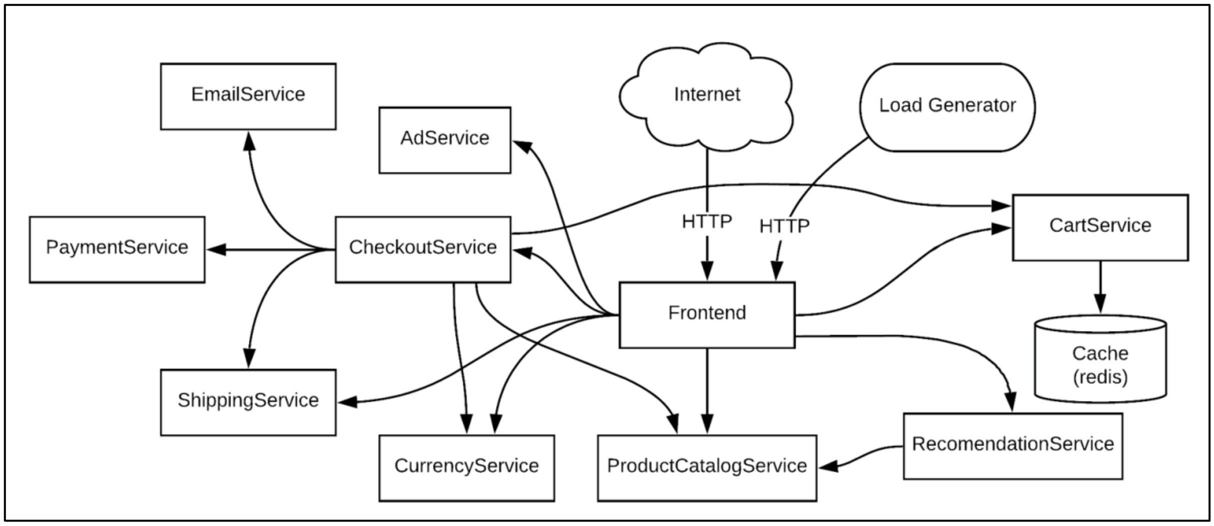


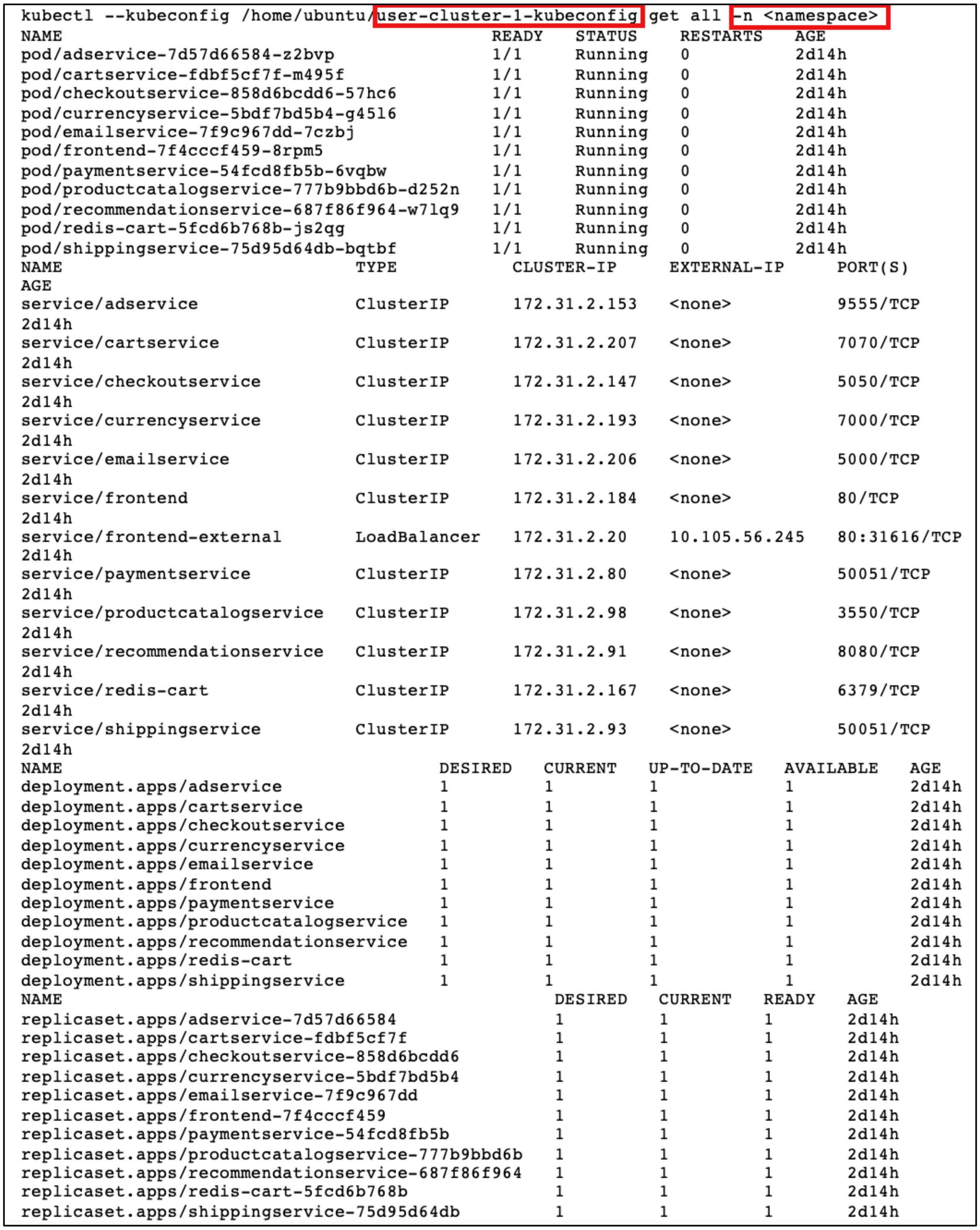
Figure 17.
Microservices in Hipster Shop

To deploy Hipster Shop download the application from Github at https://github.com/GoogleCloudPlatform/microservices-demo.git.

This example uses the prebuild container images type of installation for running on GKE. Follow the steps provided at the Github link to deploy the application.

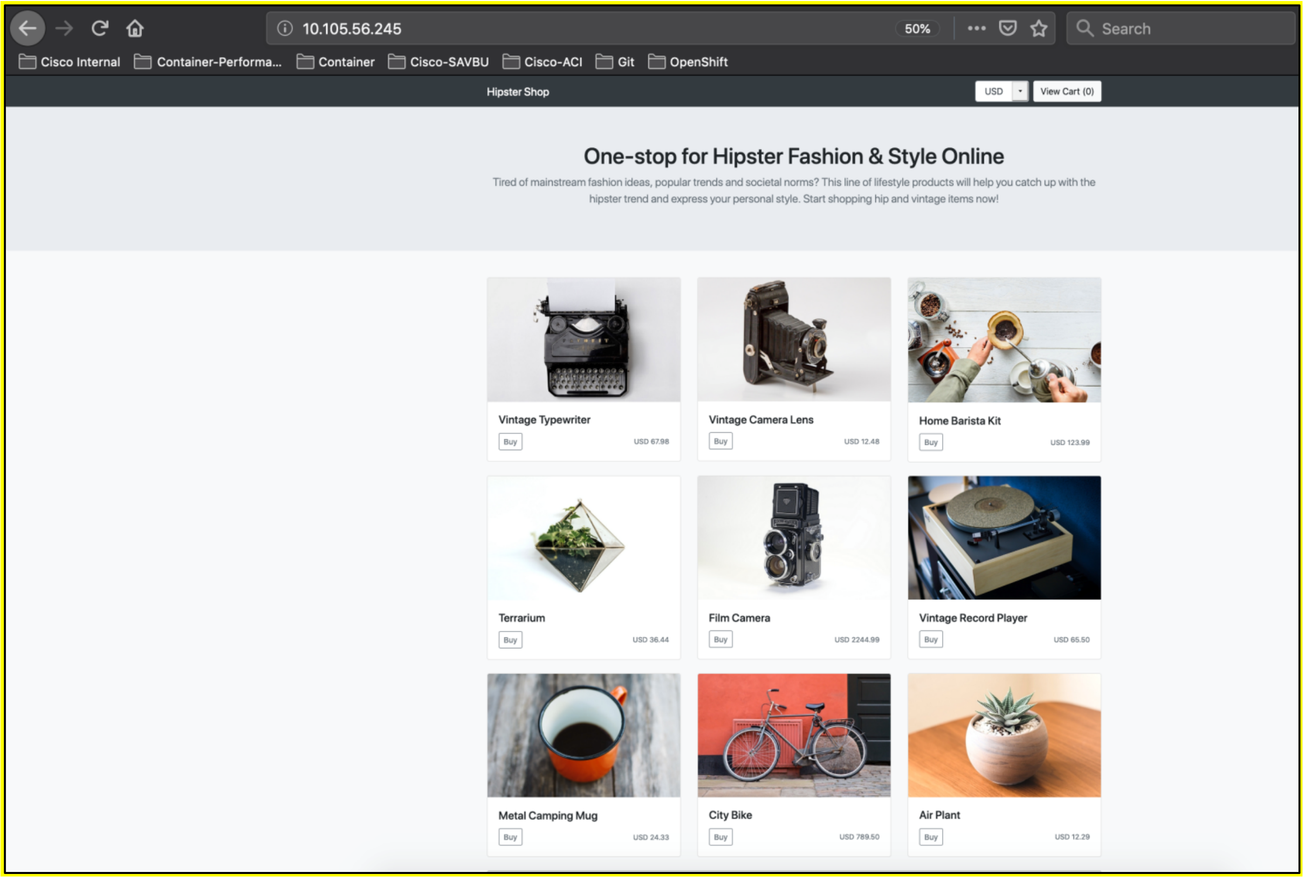
You may have to change the Kubernetes manifest and Istio manifest (if used; this deployment does not use Istio). The change primarily is made to expose your on-premises load balancer to the manifest to make your application aware of the configured load balancer. If you rely on Istio, then you do not have to expose the load balancer because the Istio manifest will handle the load-balancing task.

You can deploy this application on any of your user clusters. Create a name space in any user cluster of your choice to run this application. After the application is deployed, you will see numerous applications running and their services being created (Figures 18 and 19).



```
kubectl --kubeconfig /home/ubuntu/user-cluster-1-kubeconfig get all -n <namespace>
NAME                                         READY   STATUS    RESTARTS   AGE
pod/adservice-7d57d66584-z2bvp               1/1     Running   0          2d14h
pod/cartservice-fdbf5cf7f-m495f              1/1     Running   0          2d14h
pod/checkoutservice-858d6bcdd6-57hc6         1/1     Running   0          2d14h
pod/currencyservice-5bdf7bd5b4-g4516         1/1     Running   0          2d14h
pod/emailservice-7f9c967dd-7czbj             1/1     Running   0          2d14h
pod/frontend-7f4cccf459-8rpm5                1/1     Running   0          2d14h
pod/paymentservice-54fcd8fb5b-6vqbw          1/1     Running   0          2d14h
pod/productcatalogservice-777b9bbd6b-d252n   1/1     Running   0          2d14h
pod/recommendationservice-687f86f964-w7lq9   1/1     Running   0          2d14h
pod/redis-cart-5fcd6b768b-js2qg              1/1     Running   0          2d14h
pod/shippingservice-75d95d64db-bqtbf         1/1     Running   0          2d14h
NAME                             TYPE           CLUSTER-IP      EXTERNAL-IP     PORT(S)
AGE
service/adservice                ClusterIP      172.31.2.153    <none>          9555/TCP
2d14h
service/cartservice              ClusterIP      172.31.2.207    <none>          7070/TCP
2d14h
service/checkoutservice          ClusterIP      172.31.2.147    <none>          5050/TCP
2d14h
service/currencyservice          ClusterIP      172.31.2.193    <none>          7000/TCP
2d14h
service/emailservice             ClusterIP      172.31.2.206    <none>          5000/TCP
2d14h
service/frontend                 ClusterIP      172.31.2.184    <none>          80/TCP
2d14h
service/frontend-external        LoadBalancer   172.31.2.20     10.105.56.245   80:31616/TCP
2d14h
service/paymentservice           ClusterIP      172.31.2.80     <none>          50051/TCP
2d14h
service/productcatalogservice    ClusterIP      172.31.2.98     <none>          3550/TCP
2d14h
service/recommendationservice    ClusterIP      172.31.2.91     <none>          8080/TCP
2d14h
service/redis-cart               ClusterIP      172.31.2.167    <none>          6379/TCP
2d14h
service/shippingservice          ClusterIP      172.31.2.93     <none>          50051/TCP
2d14h
NAME                                    DESIRED   CURRENT   UP-TO-DATE   AVAILABLE   AGE
deployment.apps/adservice               1         1         1            1           2d14h
deployment.apps/cartservice             1         1         1            1           2d14h
deployment.apps/checkoutservice         1         1         1            1           2d14h
deployment.apps/currencyservice         1         1         1            1           2d14h
deployment.apps/emailservice            1         1         1            1           2d14h
deployment.apps/frontend                1         1         1            1           2d14h
deployment.apps/paymentservice          1         1         1            1           2d14h
deployment.apps/productcatalogservice   1         1         1            1           2d14h
deployment.apps/recommendationservice   1         1         1            1           2d14h
deployment.apps/redis-cart              1         1         1            1           2d14h
deployment.apps/shippingservice         1         1         1            1           2d14h
NAME                                               DESIRED   CURRENT   READY   AGE
replicaset.apps/adservice-7d57d66584               1         1         1       2d14h
replicaset.apps/cartservice-fdbf5cf7f              1         1         1       2d14h
replicaset.apps/checkoutservice-858d6bcdd6         1         1         1       2d14h
replicaset.apps/currencyservice-5bdf7bd5b4         1         1         1       2d14h
replicaset.apps/emailservice-7f9c967dd             1         1         1       2d14h
replicaset.apps/frontend-7f4cccf459                1         1         1       2d14h
replicaset.apps/paymentservice-54fcd8fb5b          1         1         1       2d14h
replicaset.apps/productcatalogservice-777b9bbd6b   1         1         1       2d14h
replicaset.apps/recommendationservice-687f86f964   1         1         1       2d14h
replicaset.apps/redis-cart-5fcd6b768b              1         1         1       2d14h
replicaset.apps/shippingservice-75d95d64db         1         1         1       2d14h
```

Figure 18.
Applications running and services being created



One-stop for Hipster Fashion & Style Online

Tired of mainstream fashion ideas, popular trends and societal norms? This line of lifestyle products will help you catch up with the hipster trend and express your personal style. Start shopping hip and vintage items now!

Vintage Typewriter
Buy USD 67.98

Vintage Camera Lens
Buy USD 12.48

Home Barista Kit
Buy USD 123.99

Terrarium
Buy USD 36.44

Film Camera
Buy USD 2244.99

Vintage Record Player
Buy USD 65.50

Metal Camping Mug
Buy USD 24.33

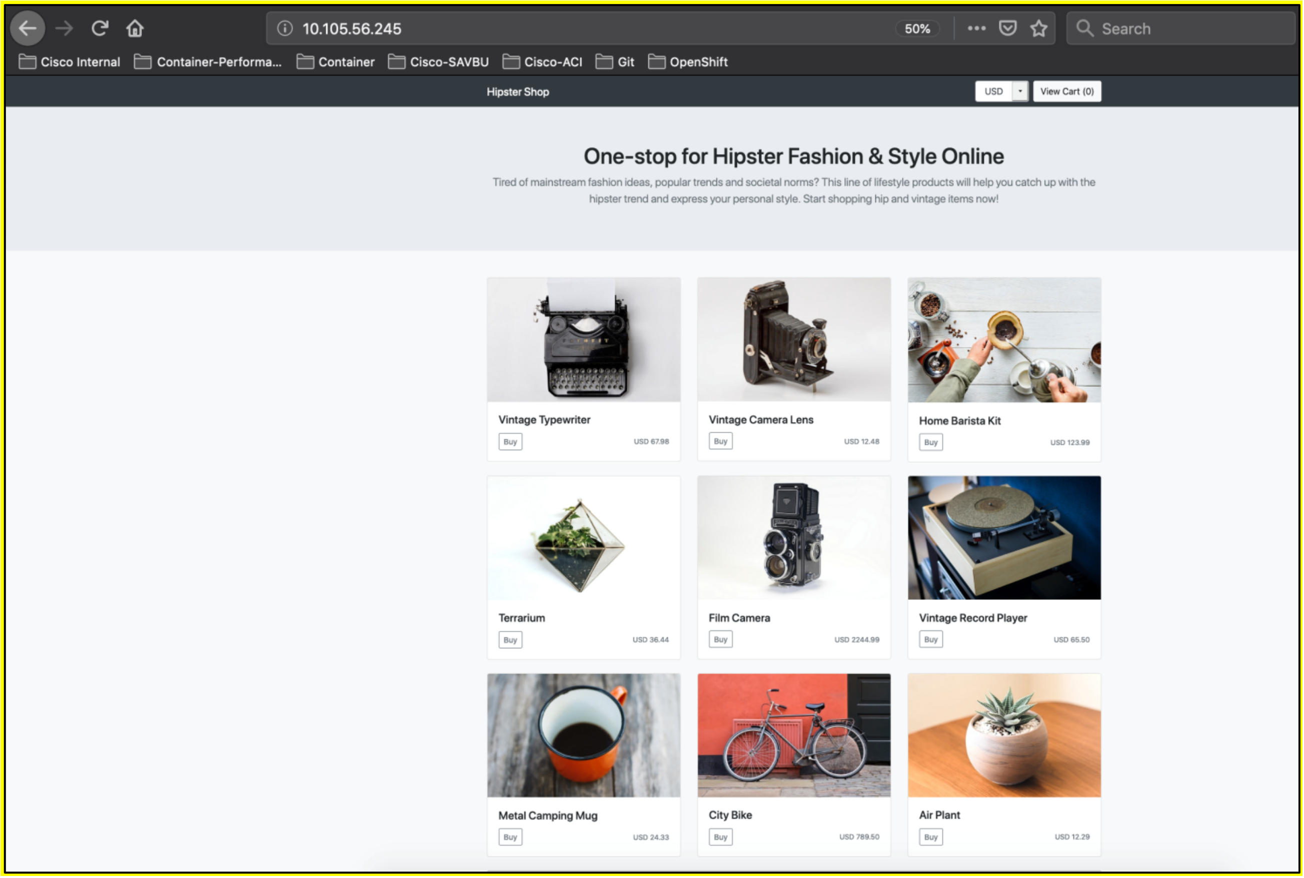
City Bike
Buy USD 789.50

Air Plant
Buy USD 12.29

Figure 19.
Hipster Shop application

Managing the workload from the Google Cloud Platform console

After the application is up and running, you can manage your workload by monitoring on Google Cloud Platform. You can also view and manage the configuration, storage, cluster status, and so on (Figure 20).



One-stop for Hipster Fashion & Style Online

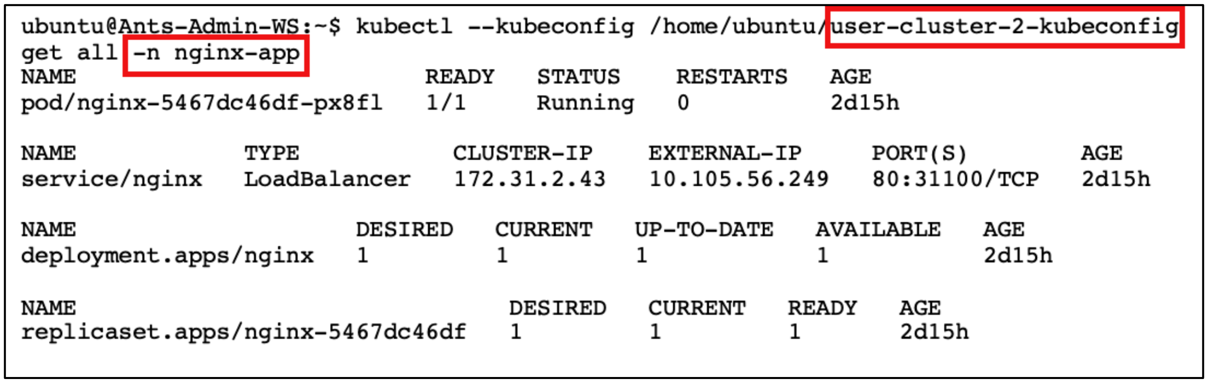
Tired of mainstream fashion ideas, popular trends and societal norms? This line of lifestyle products will help you catch up with the hipster trend and express your personal style. Start shopping hip and vintage items now!

Figure 20.
Managing the application on Google Cloud Platform console

Nginx web server deployment

This example uses the nginx web server application to demonstrate how the deployment is performed.

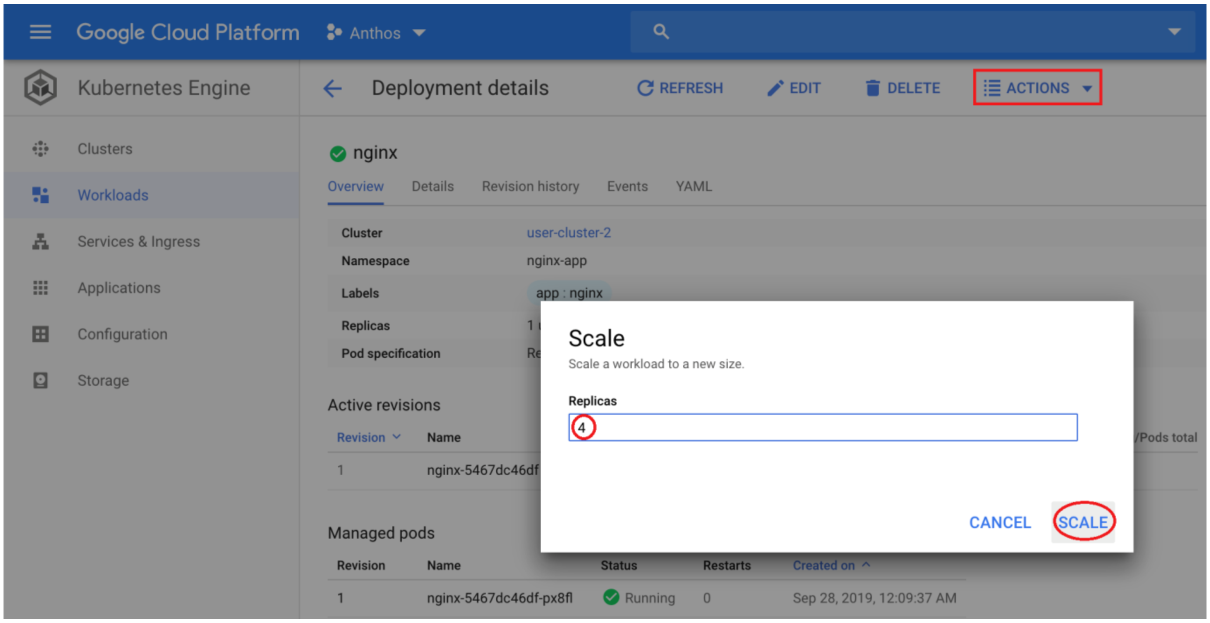
Currently, you see that one instance of the nginx web server is running in the user cluster (Figure 21).



```
ubuntu@Ants-Admin-WS:~$ kubectl --kubeconfig /home/ubuntu/user-cluster-2-kubeconfig
get all -n nginx-app
NAME                          READY   STATUS    RESTARTS   AGE
pod/nginx-5467dc46df-px8fl    1/1     Running   0          2d15h

NAME             TYPE           CLUSTER-IP    EXTERNAL-IP     PORT(S)         AGE
service/nginx    LoadBalancer   172.31.2.43   10.105.56.249   80:31100/TCP    2d15h

NAME                     DESIRED   CURRENT   UP-TO-DATE   AVAILABLE   AGE
deployment.apps/nginx    1         1         1            1           2d15h

NAME                               DESIRED   CURRENT   READY   AGE
replicaset.apps/nginx-5467dc46df   1         1         1       2d15h
```

Figure 21.
One instance of nginx web server running in the user cluster

You can scale the deployed application using the CLI or the Google Cloud Platform console (Figure 22). Figure 23 shows the deployment details after scaling.



Cluster	user-cluster-2
Namespace	nginx-app
Labels	app : nginx
Replicas	1
Pod specification	Re

Active revisions

Revision	Name	/Pods total
1	nginx-5467dc46df	

Scale

Scale a workload to a new size.

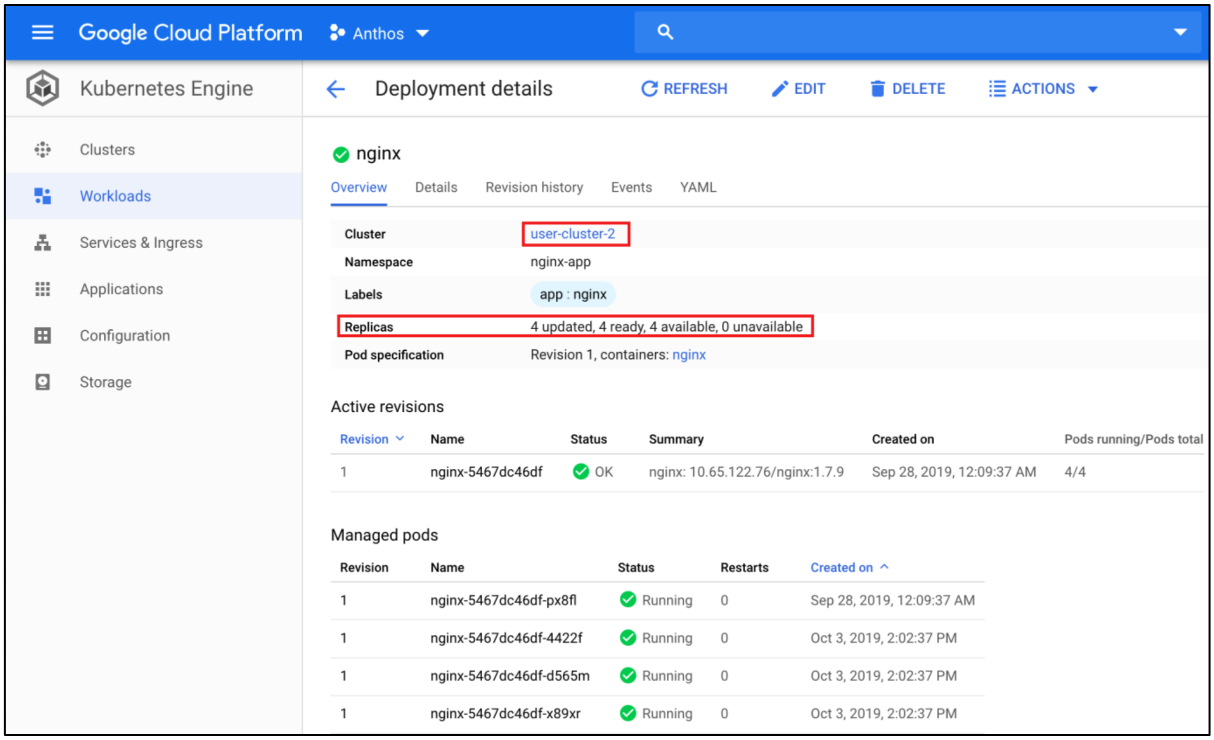
Replicas

4

Managed pods

Revision	Name	Status	Restarts	Created on
1	nginx-5467dc46df-px8fl	Running	0	Sep 28, 2019, 12:09:37 AM

Figure 22.
Scaling the application



Cluster	user-cluster-2
Namespace	nginx-app
Labels	app : nginx
Replicas	4 updated, 4 ready, 4 available, 0 unavailable
Pod specification	Revision 1, containers: nginx

Active revisions

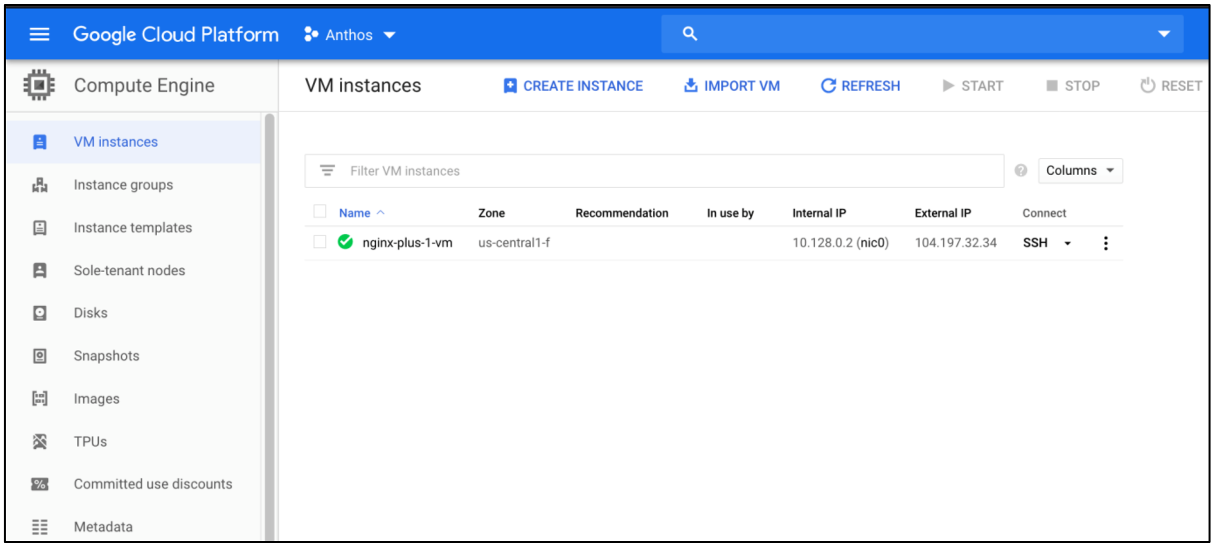
Revision	Name	Status	Summary	Created on	Pods running/Pods total
1	nginx-5467dc46df	✓ OK	nginx: 10.65.122.76/nginx:1.7.9	Sep 28, 2019, 12:09:37 AM	4/4

Managed pods

Revision	Name	Status	Restarts	Created on
1	nginx-5467dc46df-px8fl	✓ Running	0	Sep 28, 2019, 12:09:37 AM
1	nginx-5467dc46df-4422f	✓ Running	0	Oct 3, 2019, 2:02:37 PM
1	nginx-5467dc46df-d565m	✓ Running	0	Oct 3, 2019, 2:02:37 PM
1	nginx-5467dc46df-x89xr	✓ Running	0	Oct 3, 2019, 2:02:37 PM

Figure 23.
Scaling details

You can choose the application from the Google Cloud Platform Marketplace and run your workload on computing resources that would be instantaneously created on Google Cloud Platform. Figure 24 shows nginx-plus being deployed on Google Cloud Platform.



Name	Zone	Recommendation	In use by	Internal IP	External IP	Connect
✓ nginx-plus-1-vm	us-central1-f			10.128.0.2 (nic0)	104.197.32.34	SSH

Figure 24.
Nginx-plus deployed on Google Cloud Platform

Conclusion

Powered by Kubernetes and other open-source technologies, Anthos is the only software-based hybrid platform available today that lets you run your applications unmodified on

existing on-premises hardware investments or in the public cloud. The Cisco HyperFlex system, a hyperconverged infrastructure, lets you add storage and computing resources in real time, making it the best choice for enterprises to achieve cloud-like scale on-premises with Anthos GKE scale-out capabilities for container runtimes. Anthos simplifies your operations because you can use the same Kubernetes tools on-premises and in the cloud. In addition, you can configure your own private registry to maintain application container images between the two environments.

For more information

- Technical Overview on Anthos on VMware: https://cloud.google.com/gke-on-prem/docs/overview
- Anthos on VMware concepts: https://cloud.google.com/gke-on-prem/docs/concepts/
- Anthos on VMware concepts: https://cloud.google.com/kubernetes-engine/docs/concepts/
- Anthos on VMware cheatsheet: https://cloud.google.com/gke-on-prem/docs/reference/cheatsheet
- Anthos on VMware troubleshooting:
  https://cloud.google.com/gke-on-prem/docs/support/troubleshooting

Americas Headquarters
Cisco Systems, Inc.
San Jose, CA

Asia Pacific Headquarters
Cisco Systems (USA) Pte. Ltd.
Singapore

Europe Headquarters
Cisco Systems International BV Amsterdam,
The Netherlands

Cisco has more than 200 offices worldwide. Addresses, phone numbers, and fax numbers are listed on the Cisco Website at **https://www.cisco.com/go/offices**.

Printed in USA 219376.6 11/19