Technical White Paper: **June 2023**

Architecting Kubernetes-as-a-Service Offering with VMware Cloud Director

Shady Ali ElMalatawey
Sr. Staff Multi-Cloud Solutions Architect, VCDX #249
June 2023

vmware®

Table of contents

Introduction

According to VMware's "The State of Kubernetes 2023" report, the organizations whose Kubernetes deployments are used for production prefer using either multiple public cloud vendors (53%) or a single public cloud vendor (36%) for their upcoming deployments within the next year[1]. With this increasing demand, the public cloud providers are competing to provide best-of-breed feature-rich Kubernetes-as-a-Service (KaaS) offerings.

VMware Cloud Service Providers has been leveraging VMware Cloud Director (VCD) for years to provide variety of service offerings in their platforms whether in a single data center or spanning multiple data centers. Today, they can leverage Tanzu Kubernetes Grid (TKG) with VMware Cloud Director to provide KaaS offering that meets customers demand.

The purpose of this document is to provide architectural guidance to build a KaaS service using VMware Cloud Director and Tanzu Kubernetes Grid to provide self-service Kubernetes offering in a multi-site cloud platform.

[1] The State of Kubernetes 2023 report, VMware Inc.: https://tanzu.vmware.com/content/ebooks/stateofkubernetes-2023

Components and Bill of Materials

A KaaS service offering consists of multiple components. In each cloud site, the provider needs to deploy the following:

- VMware Cloud Director Container Service Extension
- VMware Cloud Director Object Storage Extension
- VMware Cloud Director App Launchpad
- VMware Cloud Director Extension for Data Solutions

VMware Cloud Director Container Service Extension

VMware Cloud Director Container Service Extension (CSE) is a VCD extension which enables KaaS capabilities. CSE offers VMware supported multi-tenant and production ready Kubernetes services built on TKG. When using CSE, the provider administrators can enable KaaS for new and existing customer's VCD tenants. Tenants can use CSE to create and manage TKG clusters in their virtual data centers alongside their virtual machines.

Providers only need to deploy CSE if they wish to offer a basic KaaS service, which offers only TKG clusters to the tenants without any additional capabilities.

VMware Cloud Director Object Storage Extension

VMware Cloud Director Object Storage Extension (OSE) is a VCD extension that provides self-service object storage capabilities. With OSE, the provider can enable a multi-tenant object storage service on a supported object storage platform and assign object storage quota to existing VCD tenants. Tenants can use an S3-compatible API or the UI to consume their assigned quota. OSE provides backup and restore capabilities for TKG clusters managed by CSE using Velero.

OSE is an optional component to provide object storage and backup/restore capabilities for the TKG clusters.

VMware Cloud Director App Launchpad

VMware Cloud Director App Launchpad (ALP) is a VCD extension which provides an applications marketplace experience to tenants. ALP enables the provider to create and publish catalogs of deployment-ready applications. Tenants can then deploy the applications with a single click. ALP exposes the configuration parameters so that the tenants can fine-tune the advanced settings of the container applications during deployment.

ALP supports using the Bitnami applications catalog that is available in the VMware Marketplace. Providers can also deliver their own managed application marketplace using managed Helm Chart repositories.

ALP is an optional component to provide a containerized applications marketplace experience.

VMware Cloud Director Extension for Data Solutions

VMware Cloud Director Extension for Data Solutions (Data Solutions Extension or DSE) is a plugin for VCD that enables the provider to deliver a portfolio of on-demand data services at scale. With DSE, the provider administrators can offer tenants enterprise-ready VMware Tanzu Data Solutions for building and running more powerful and modernized applications. Tenants can use the DSE self-service user interface (UI) for the life cycle management of the data solution instances. It also provides a single view across multiple instances, and URLs for individual instances for service specific management.

DSE is an optional component to provide data solutions instances on top of TKG clusters.

Note: Currently, DSE 1.1 supports VMware Tanzu RabbitMQ, VMware SQL with Postgres for Kubernetes and VMware SQL with MySQL for Kubernetes. More VMware Tanzu Data Solutions will be supported in the future.

Bill of Materials

The following table details the required components for the provider to introduce a KaaS offering on an existing Infrastructure-as-a-Service (IaaS) platform powered by VCD.

Table 1 - Bill of Materials	
Component	Version
VCD	10.3.1 and above
NSX Data Center	3.x and above (any supported version with the VCD version used)
NSX Advanced Load Balancer	21.1.x and above (any supported version with the VCD version used)
vSphere	vSphere 7.0 and above (any supported version with the VCD version used)
VMware Chargeback	8.10
CSE	4.0.3
OSE	2.2.x
ALP	2.1.2
DSE	1.1
TKG	1.6.1
vCloud Usage Meter	4.6
Cloudian HyperStore	7.4.x and above

Cloud Platform Building Blocks

For this architecture, we need to define the building blocks of the proposed multi-site cloud platform. We will define two main building blocks:

- Region
- Availability Zone (AZ)

Region

Regions are groups of data centers which are fully isolated from each other. Regions are in dispersed geographical locations to mitigate large-scale risks, whether human-made, like wars and conflicts, or natural disasters, like tsunamis and hurricanes. This enables providers to offer the maximum required availability for their tenants. Usually, not all workloads are replicated between different regions due to the low bandwidth between them. A region usually consists of one or more Availability Zones (AZs).

For this document, we assume that each region of the provider's cloud platform will provide its own KaaS offering. We also assume that there is not a single KaaS offering that spans different regions.

Availability Zone

Availability Zones (AZs) are isolated data centers in different geographical locations. However, they are within proximity to allow for low-latency, high-bandwidth links to connect them. This would allow replication of workloads across these AZs for disaster recovery purposes. Sometimes a single AZ can consist of a group of data centers within a single campus. On other occasions, a group of AZs can be co-located in the same geographical area, however, these AZs are totally isolated from each other to achieve the required availability. An AZ must always be an atomic self-sufficient building block of the provider's cloud platform that is totally isolated from other AZs.

Each AZ consists of, at least, the following components:

- VMware Cloud Director
    - All VCD instances in all AZs are configured into a single multi-site association. For more information about multi-site VCD configuration, refer to VMware Cloud Director 10.3 Multi-site Architecture with Global Load Balancing Solution.
    - A solutions organization deployed in each VCD instance in each AZ. Refer to Solutions Organization.
- Advanced Message Queuing Protocol (AMQP) Broker
    - Highly available RabbitMQ instance integrated with the local VCD instance.
- VMware Infrastructure
    - VMware NSX
    - VMware NSX Advanced Load Balancer (Formerly known as AVI Load Balancer)
    - VMware vSphere
- Supported Object Storage Platform (Cloudian HyperStore, Dell EMC ECS or AWS Account with AWS S3 storage service enabled).
- One or more vCloud Usage Meter appliances
- VMware Chargeback Appliance
    - This requires a backend VMware Aria Operations cluster (not shown through this document).
- Supporting Infrastructure
    - DNS

- NTP
- Identity Provider (IdP)
- Physical Load Balancer for the Object Storage Platform
- Global Load Balancing (GLB) Solution to provide a unified URL for all VCD instances across all AZs.
- (Optional) Private ChartMuseum Helm Chart Repository
- (Optional) Private Open Container Initiative (OCI) registry

For this document, we assume that each AZ of the provider's cloud platform will provide its own KaaS offering. Tenants can deploy TKG clusters in all AZs that they have access to and can perform backup/restore of TKG clusters between AZs for disaster recovery purposes.

Note: For this document, Cloudian HyperStore will be used as the object storage platform. Other supported object storage platforms are not in the scope of this document.

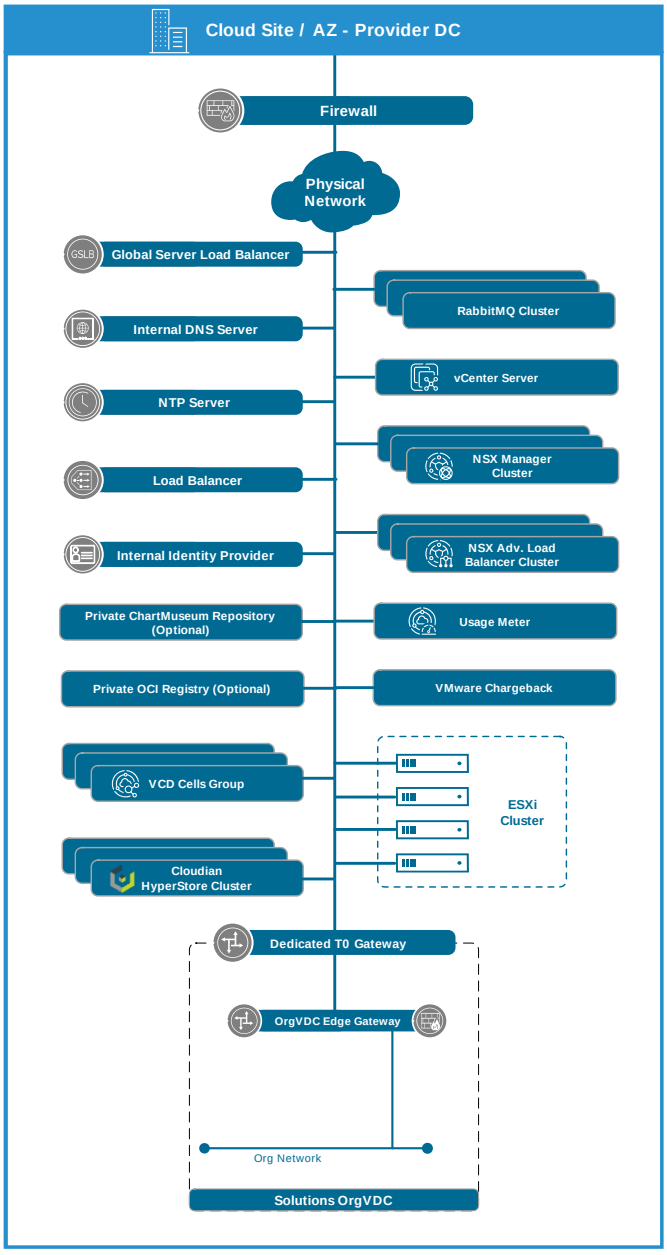

Figure 1: Provider Availability Zone

Solutions Organization

This organization is a Cloud Provider managed organization which will host CSE appliance(s) in an Organization Virtual Data Center (OrgVDC). It will also host a catalog of TKG templates which will be the basis for tenants' TKG clusters.

This organization will provide network connectivity for CSE appliance(s) to the VCD public API endpoint.

To achieve this, the provider must create an OrgVDC with the desired allocation model and set the allocation settings to provide enough resources to host the anticipated number of CSE appliance(s). The provider must consider the required storage resources as well for the TKG templates. It is recommended to guarantee all the assigned resources for the OrgVDC to prevent any performance issues. The following table shows the required resources for a CSE appliance and the required storage resources for the templates.

Table 2 - Required Resources for CSE Appliance and TKG Templates

Attribute	Specifications
CSE Appliance CPU	2 vCPUs
CSE Appliance Memory	4 GB
CSE Appliance Storage	51 GB
Ubuntu 2004 Kubernetes v1.22.9 Template Storage	20 GB
Ubuntu 2004 Kubernetes v1.21.11 Template Storage	20 GB
Ubuntu 2004 Kubernetes v1.20.15 Template Storage	20 GB

The next step is to create an Org Edge Gateway which will be connected to a dedicated T0 Gateway or a dedicated T0 VRF. This Org Edge Gateway will provide connectivity to the local VCD instance public API endpoint to the OrgVDC network where the CSE appliance(s) will be deployed. The Org Edge Gateway will also provide connectivity to the internal provider-managed DNS servers so CSE appliance(s) can resolve the VCD public API endpoint FQDN.

The provider must create an OrgVDC network, connect it to this Org Edge Gateway and configure the OrgVDC network with the required static IP pool and DNS settings to allow CSE appliance(s) to resolve and communicate with the local VCD public API endpoint. Figure 2 below shows the OrgVDC networking design.

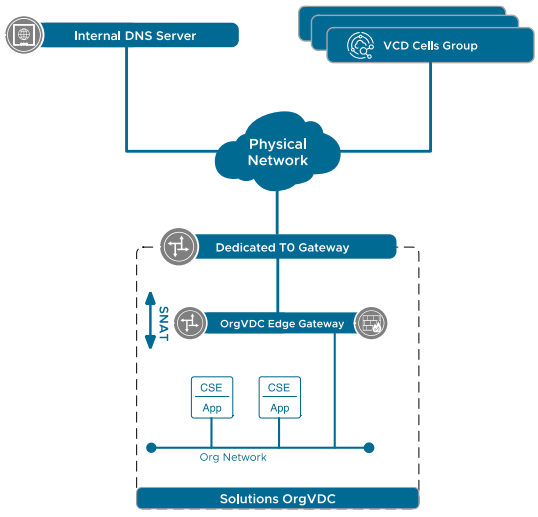

Figure 2: Solutions OrgVDC Networking Design

The final step is to create a catalog in the OrgVDC which will host all TKG templates. This catalog must be shared across all tenants that will use CSE to deploy TKG templates.

VMware Cloud Director Containers Service Extension

To offer a basic KaaS service, which offers only TKG clusters to the tenants without any additional capabilities, the provider must install and configure a CSE instance in each cloud site. The following sections detail the design and best practices for CSE architecture.

Deployment Specifications

CSE is distributed as an Open Virtualization Appliance (OVA) which the provider can deploy in the Solutions Organization. For high availability, the provider must install two or more CSE appliances in each cloud site or AZ and configure CSE to connect to the local VCD instance. Each CSE appliance requires with the following resources:

- 2 vCPU
- 4 GB RAM
- 51 GB Disk

For more information about the exact deployment and configuration steps, refer to the VMware Official Documentation.

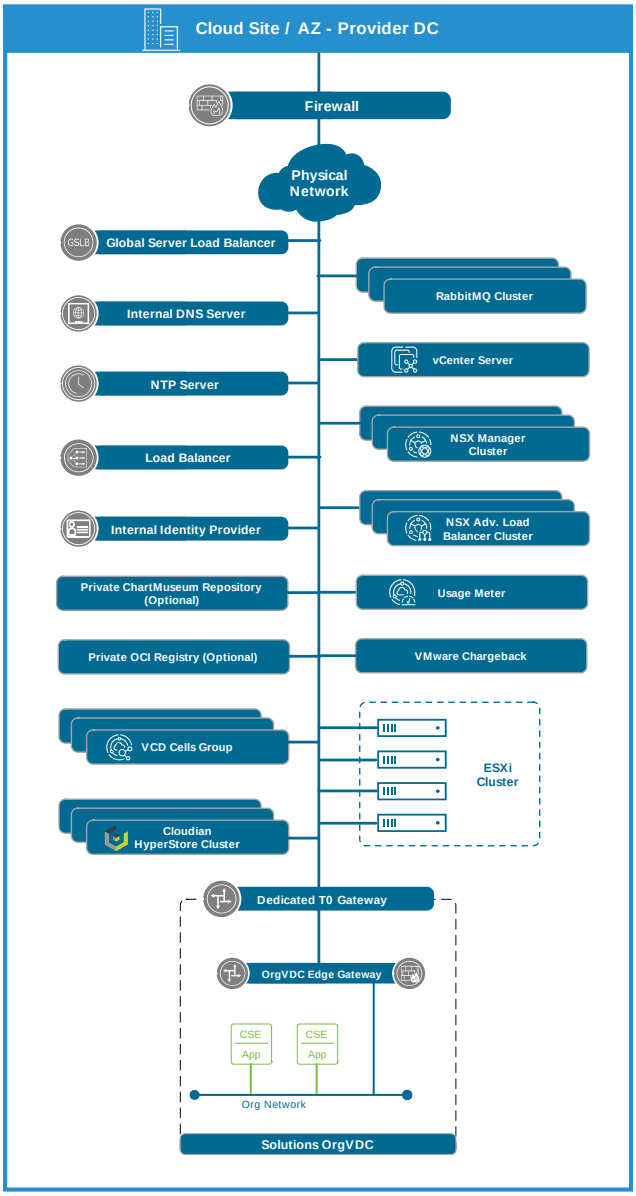

Figure 3: CSE Deployment in each AZ (new components marked in green)

Network Design

Basic Networking

In each cloud site or AZ, CSE appliances must be able to connect to the local VCD instance. CSE must be able to resolve the VCD public address endpoint FQDN to an internal or private IP address for integration between CSE and the local VCD instance. Refer to Traffic Flow and Communication Matrix for the complete CSE communication matrix.

Load Balancing

Although CSE itself does not require load balancing, NSX Advanced Load Balancer, that is integrated with VCD, is a prerequisite for CSE. This is required to support the Services and Ingress resources that are deployed on TKG clusters for cluster services, like Cluster API service, or the user's applications.

External Accessibility

A tenants' TKG clusters must be deployed on a routed OrgVDC network with internet access. During TKG cluster deployment or deletion, an ephemeral VM is created to boot-strap the TKG cluster. That ephemeral VM is deployed to the same location and network as the TKG cluster and must have internet access to download the Tanzu packages required for cluster operations from GitHub and other online depots. TKG nodes also must have internet access to download the required packages for other Day 1 and Day 2 operations like cluster resizing or upgrades. Please refer to the Traffic Flow and Communication Matrix for the complete CSE communication matrix.

Alternatively, the provider can configure CSE to use a proxy server. CSE pushes proxy server settings to all TKG clusters being deployed. These TKG clusters will use the proxy settings to access the internet. The OrgVDC network where the TKG clusters are deployed must be able to reach the proxy server only. The provider can also exclude certain domains, IP addresses or URLs from being accessed using the proxy server, for example an internal provider-managed registry.

It is also recommended for the provider to configure CSE with a GitHub API token to remove any rate limits imposed by GitHub on packages downloads.

Internal Accessibility and Integrations

CSE appliances must be able also to resolve the VCD public address endpoint FQDN to an internal or private IP address for integration between CSE and the local VCD instance.

Tenants' TKG clusters must be deployed on routed OrgVDC networks with accessibility to internal provider-managed DNS servers. Tenants' TKG clusters must be able to resolve the public address endpoint FQDN of the local VCD instance to poll the VCD instance for any resizing and update operations using CSE.

Traffic Flow and Communication Matrix

Figure 4 below shows the complete traffic flows for CSE. Table 3 shows the firewall ports that need to be opened on Tenant and Provider firewalls. Table 4 shows the required NAT rules, and Table 5 shows the URLs that the CSE appliances need to be able to reach to download Tanzu components.

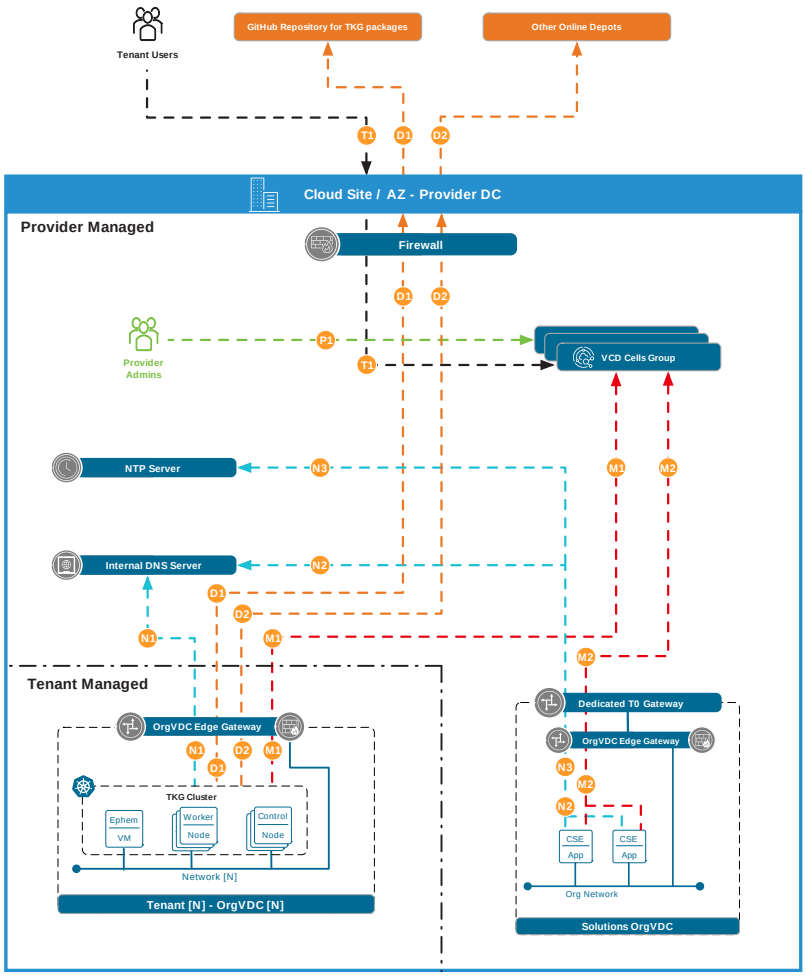

Figure 4: CSE Traffic Flows

Table 3 - CSE Traffic Flows				
Legend	Source	Destination	Protocol / Port	Description
T1	Tenant Users	VCD	TCP 443	Used for access CSE Tenant Portal
A1	Provider Admins	VCD	TCP 443	Used for access CSE Provider Portal
D1	- TKG Cluster - Ephemeral VM	GitHub Repository for TKG	- TCP 443 - TCP 6443	Used for download TKG Packages
D2	- TKG Cluster - Ephemeral VM	Other Online Depots	- TCP 443 - TCP 6443	Used for download TKG Packages
M1	TKG Cluster	VCD	TCP 443	Used for polling VCD for changes in TKG Cluster specifications through resizing or upgrade operations
M2	CSE Appliances	VCD	TCP 443	Used for polling VCD for any TKG cluster creation or deletion operations
N1	- TKG Cluster - Ephemeral VM	DNS Server	TCP/UDP 53	Used for name resolution
N2	CSE Appliances	DNS Server	TCP/UDP 53	Used for name resolution

N3	CSE Appliances	NTP Server	UDP 123	Used for time synchronization

Table 4 – Required NAT Rules for CSE					
Legend	Type	Internal IP	Internal Port	External IP	External port
M2 / N2/ N3	SNAT	CSE Appliances	Any	Public IP	Any
D1 / D2 / M1 / N1	SNAT	TKG Cluster	Any	Public IP	Any

Lifecycle Management

CSE builds upon the following components:

- VCD
- NSX
- NSX Advanced Load Balancer
- ALP
- OSE
- DSE

To upgrade CSE, make sure that all dependencies are upgraded, if needed, before upgrading CSE. Refer to VMware Interoperability Matrix before commencing with the upgrade.

Each new version of CSE brings support for newer versions of TKG clusters. It is recommended to check the official CSE Release Notes and VMware Interoperability Matrix to know which TKG versions are supported.

Multi-site / Multi-Availability Zone Considerations

Each cloud site in a Multi-Site or Multi-AZ configuration will have one or more dedicated CSE appliances deployed that connects to the local VCD instance. Tenants will use the local CSE instance(s) to deploy TKG clusters in the same cloud site or AZ. CSE does not currently support deploying TKG clusters across different VCD instances in a Multi-site Association.

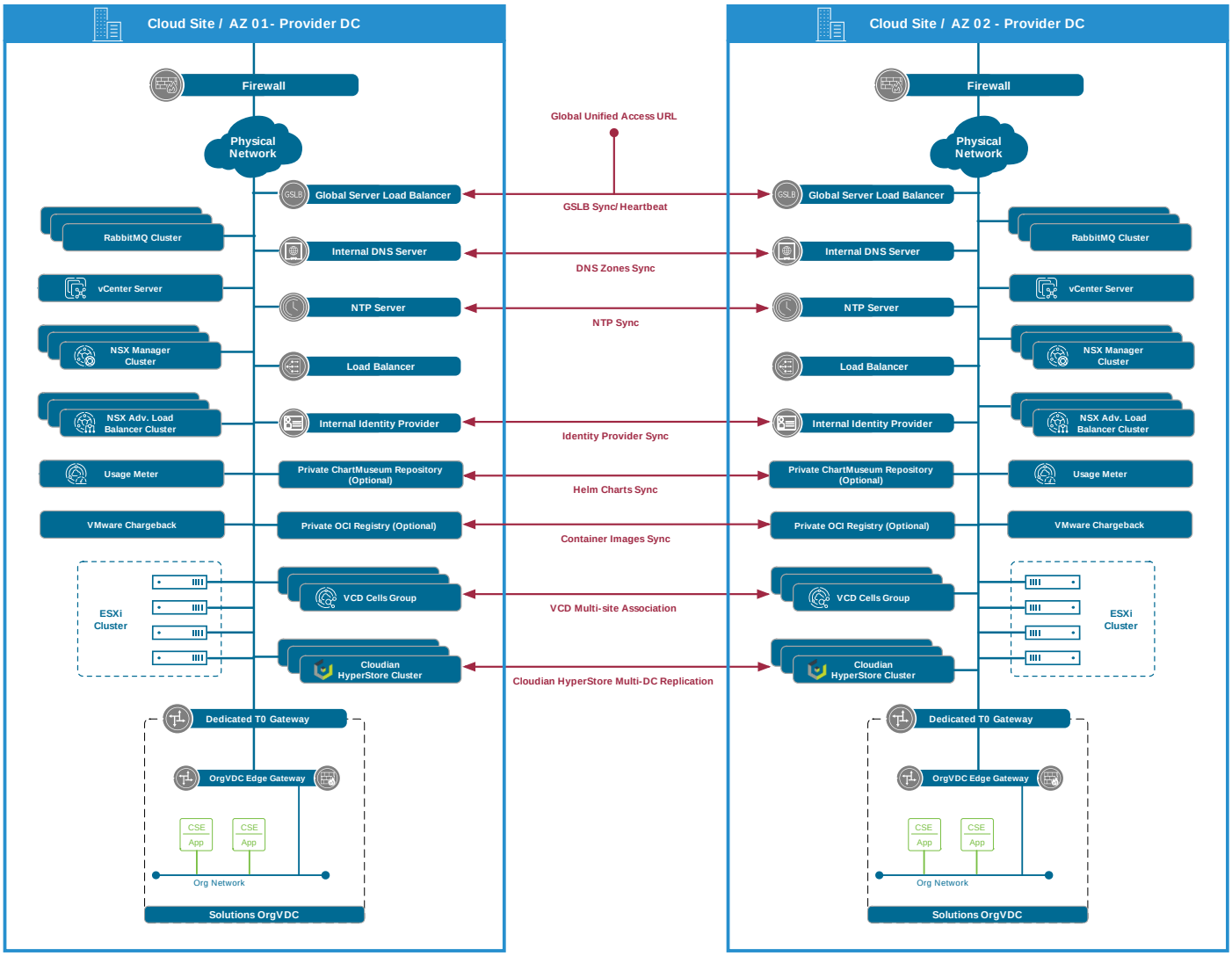

Figure 5: CSE Deployment in Multi-Site / Multi-AZ Configuration

VMware Cloud Director Object Storage Extension

To add self-service backup and restore capabilities for TKG clusters, the provider should install and configure an OSE instance in each cloud site. The following sections detail the design and best practices for OSE architecture.

Deployment Specifications

OSE is distributed as an RPM package and can be installed on the following Linux distributions:

- CentOS Linux 7
- CentOS Linux 8
- Red Hat Enterprise Linux 7 or later
- Photon OS 3 or later
- Oracle Linux 7 or later
- Ubuntu 18 or later
- Debian 10 or later.

OSE requires a Java 8 or later JRE and a backend PostgreSQL database to store metadata. For high availability, it is recommended to use a PostgreSQL database cluster.

For high availability and maximum performance and scale, the provider will install OSE on two or more Linux VMs in each cloud site or AZ and configure OSE to connect to the local VCD instance. OSE VMs must be of large size and configured with the following resources:

- 12 vCPU
- 12 GB RAM
- 120 GB Disk

For other deployment sizes, refer to the official VMware Cloud Director Object Storage Extension Documentation and VMware Cloud Director Object Storage Extension 2.2 – Reference Design.

The next step is to configure OSE nodes to connect the local Cloudian HyperStore cluster. For more information, refer to VMware Official Documentation.

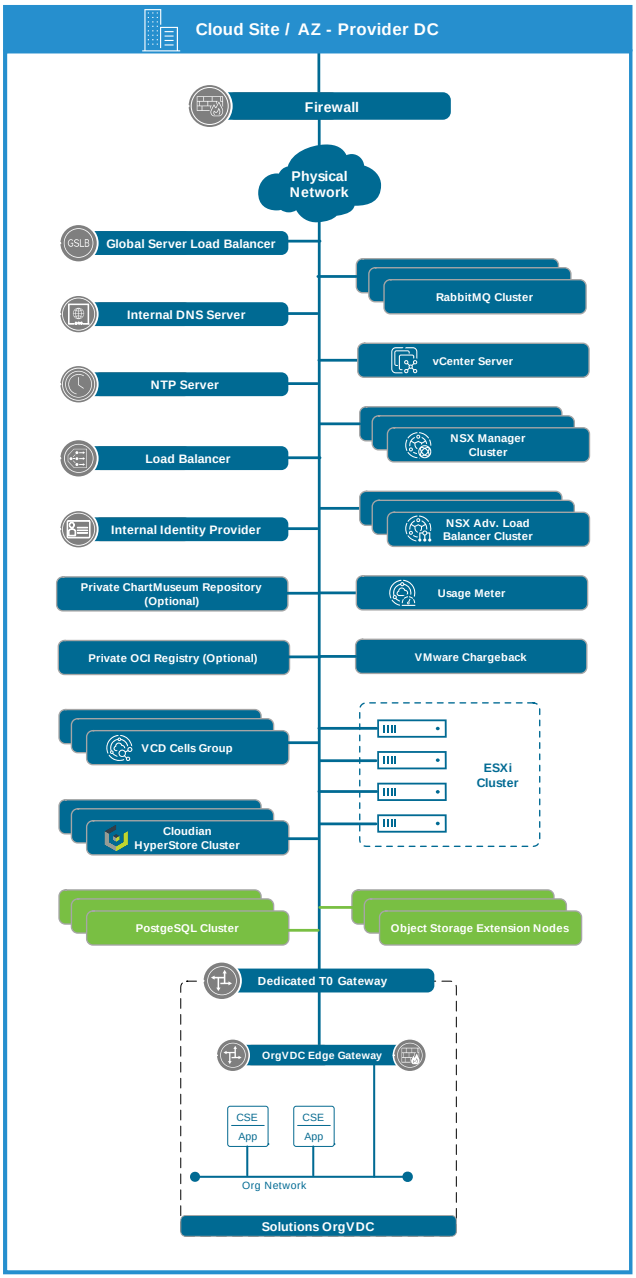

Figure 6: OSE Deployment in each AZ (new components marked in green)

Network Design

Basic Networking

In each cloud site or AZ, OSE nodes must be able to connect to the local VCD instance as well as the local Cloudian HyperStore cluster. OSE nodes will connect to different services on the Cloudian HyperStore cluster. It is recommended to use an external hardware load balancer in front of the Cloudian cluster which exposes the required services to ensure high availability and maximum performance with scalability. Refer to Traffic Flow and Communication Matrix for the complete OSE communication matrix.

Load Balancing

For scaling purposes, the provider will install two or more OSE nodes behind a load balancer in each cloud site or AZ. An external hardware load balancer is recommended for maximum performance. For other load balancing options, refer to VMware Cloud Director Object Storage Extension 2.2 – Reference Design.

The following tables indicate the recommended load balancing configurations.

Table 5 - Recommended Configuration for OSE Load Balancer Virtual Service	
Attribute	Specifications
Type	L4 TCP
Port	443
Persistence	Disabled
Load Balancing Algorithm	Least Connection
Idle Timeout	1,800 seconds

Table 6 - Recommended Configuration for OSE Load Balancer Service Monitoring	
Attribute	Specifications
Type	Active Monitoring
Protocol	HTTPS
Port	443
HTTP Request	HTTP OPTIONS /api/v1/core HTTP Version 1.1
HTTP Response	Response Code 200

External Accessibility

For the tenants external accessibility, the OSE load balancer virtual service must be externally accessible. The provider must assign a public IP address and a public FQDN, which resolves to the public IP address for the OSE virtual service in each cloud site or AZ.

To support Virtual Hosted-style S3 API access, public DNS servers must be able to map subdomains of that public FQDN to the same public IP address of OSE virtual service. Sub-domains are in the format of *s3.ose-FQDN* and **.s3.ose-FQDN*.

For secure communication, a public certificate with multiple Subject Alternative Names (SANs) can be used. The certificate SANs must include the OSE FQDN and the sub-domains mentioned above to support Virtual Hosted-style S3 API access.

Internal Accessibility and Integrations

Internal provider-managed DNS servers must be able to resolve the public FQDN of OSE virtual service to the internal load balancer virtual service IP address. All sub-domains of *s3.ose-FQDN* and **.s3.ose-FQDN* must be resolved internally to the same internal IP address. All Cloudian HyperStore services exposed through the load balancer must have internal DNS records which OSE can resolve and connect to. OSE must be able also to resolve the VCD public address endpoint FQDN to an internal IP address for integration between OSE and the local VCD instance.

Backup/Restore of Kubernetes Clusters

To perform backup and restore operations on tenants' TKG clusters, OSE nodes must be able to access tenants' TKG cluster API endpoints. Tenants' TKG clusters must be exposed to an external network which OSE nodes can communicate with. Refer to the Traffic Flow and Communication Matrix for the complete OSE communication matrix.

In addition, TKG clusters must be able to download the Velero / Kopia package from the public GitHub registry. Alternatively, the provider can configure OSE to instruct TKG clusters to download Velero / Kopia package from an internal provider managed OCI registry.

Traffic Flow and Communication Matrix

Figure 7 below shows the complete traffic flows for OSE. Table 7 shows the firewall ports that need to be opened on Tenant and Provider firewalls. Table 8 shows the required NAT rules, and Table 9 shows the URLs that need to be allowed to download Velero / Kopia packages.

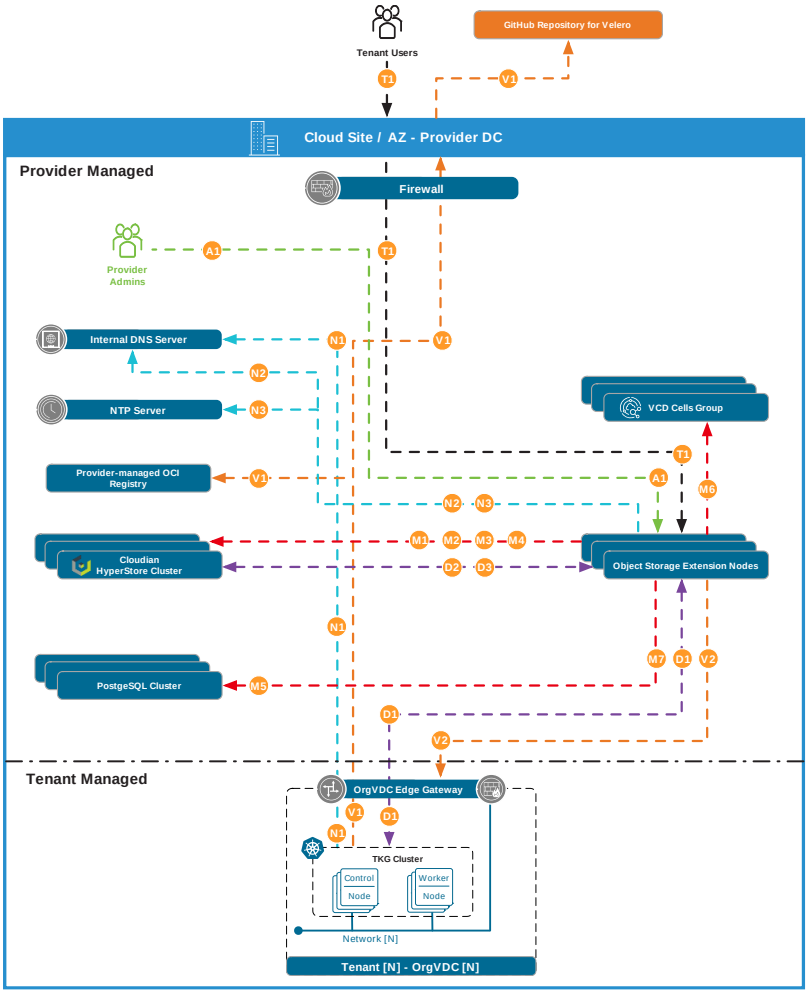

Figure 7: OSE Traffic Flows

Table 7 - OSE Traffic Flows				
Legend	Source	Destination	Protocol / Port	Description
T1	Tenant Users	OSE Service Public IP/FQDN	TCP 443	- Used for access OSE Tenant Portal - Used for S3 API Access

A1	Provider Admins	OSE Service Internal IP/FQDN	TCP 443	- Used for access OSE Provider Portal - Used for S3 API Access
V1	TKG Cluster	- GitHub Repository for Velero - Provider-managed OCI Registry	TCP 443	Used for download Velero / Kopia Package
V2	OSE Nodes	TKG Cluster API LB Virtual Service	TCP 6443	Used for issuing backup / restore operations to TKG Cluster API Server
M1	OSE Nodes	Cloudian HyperStore LB IP / FQDN	TCP 19443	Used for communication between OSE nodes and the Cloudian Admin Service
M2	OSE Nodes	Cloudian HyperStore LB IP / FQDN	TCP 8443	Used for communication between OSE nodes and the Cloudian Management Console (CMC)
M3	OSE Nodes	Cloudian HyperStore LB IP / FQDN	TCP 16443	(Recommended) Used for secured communication between OSE nodes and the Cloudian IAM Service
M4	OSE Nodes	Cloudian HyperStore LB IP / FQDN	TCP 16080	(Optional - Not recommended) Used for unsecured communication between OSE nodes and the Cloudian IAM Service
M5	OSE Nodes	PostgreSQL Cluster	TCP 5432	Used for OSE communication with the backend database
M6	OSE Nodes	VCD	TCP 443	Used for interactions with VCD and polling all TKG clusters information from VCD APIs
D1	TKG Cluster	OSE Nodes	TCP 443	Used for TKG Data backup traffic
D1	OSE Nodes	TKG Cluster	TCP 443	Used for TKG Data restore traffic
D2	OSE Nodes	Cloudian HyperStore LB IP / FQDN	TCP 443	(Recommended) Used for TKG Data backup traffic through secured S3 protocol
D2	Cloudian HyperStore Nodes	OSE Nodes	TCP 443	(Recommended) Used for TKG Data restore traffic through secured S3 protocol
D3	OSE Nodes	Cloudian HyperStore LB IP / FQDN	TCP 80	(Optional - Not recommended) Used for TKG Data backup traffic through unsecured S3 protocol
D3	Cloudian HyperStore Nodes	OSE Nodes	TCP 80	(Optional - Not recommended) Used for TKG Data restore traffic through unsecured S3 protocol
N1	TKG Cluster	DNS Server	TCP/UDP 53	Used for name resolution
N2	OSE Nodes	DNS Server	TCP/UDP 53	Used for name resolution
N3	OSE Nodes	NTP Server	UDP 123	Used for time synchronization

Table 8 – Required NAT Rules for OSE

Legend	Type	Internal IP	Internal Port	External IP	External port
T1	DNAT	OSE Service Internal IP	TCP 443	Public IP	TCP 443
V1 / D1 / N1	SNAT	TKG Cluster	Any	Public IP	Any

Table 9 – Required URLs for OSE		
Legend	URL	Description
V1	https://github.com/vmware-tanzu/helm-charts/releases/download/velero-3.0.0/velero-3.0.0.tgz	Used for pulling Velero / Kopia package from GitHub

Lifecycle Management

OSE builds upon the following components:

- VCD
- CSE
- Cloudian HyperStore

Any dependencies must be upgraded, if needed, before upgrading OSE. Refer to VMware Interoperability Matrix and Cloudian HyperStore documentation before commencing with the upgrade.

Multi-site / Multi-Availability Zones Considerations

In Multi-Site or Multi-AZ configuration, each cloud site will have a dedicated OSE instance installed which connects to the local VCD instance and the local Cloudian HyperStore cluster.

The Cloudian HyperStore clusters in all cloud sites or AZs must be configured in a Multi-Data Center setup. Refer to Cloudian HyperStore Documentation for more information about how to setup the Cloudian clusters. Once configured, the provider must configure new storage policies with the required replication method and number of replicas in each location before utilizing the platform to ensure that any user data is replicated as needed. This can be done from Cloudian Management Console (CMC) or from OSE directly. Finally, the provider can assign these storage policies to different tenants as needed using OSE APIs or portal.

For seamless operations, the VCD and OSE instances in all cloud sites must be able to resolve the FQDNs of the public endpoint FQDN of all VCD instances in all sites, OSE virtual service public FQDNs in all sites and the Cloudian clusters services FQDNs in all sites.

Each tenant's VCD organizations in each AZ must be associated together in a multi-site association so that tenants can access and use OSE in all AZs. In addition, each tenant's VCD organizations in each AZ must be configured with the same identity provider. In case of an AZ failure, tenants can access their backups on the object storage platform from other AZs through OSE portal. For more information, refer to Object Storage-as-a-Service Multi-site Architecture with VMware Cloud Director 10.3, Object Storage Extension 2.1 and Cloudian Hyperstore.

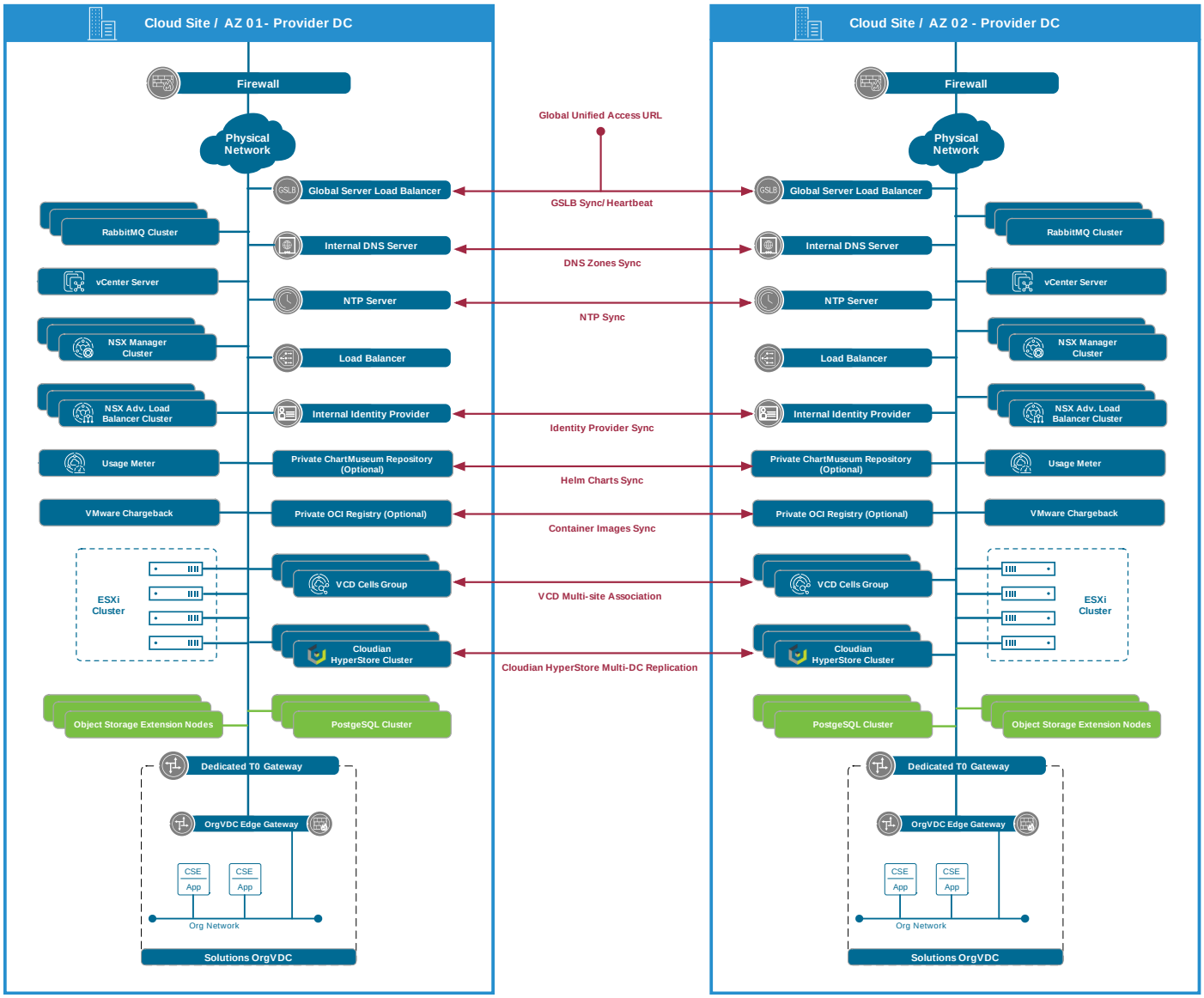

Figure 8: OSE Deployment in Multi-Site / Multi-AZ Configuration

VMware Cloud Director App Launchpad

To add a marketplace experience for provider-managed containerized applications, the provider should install and configure an ALP instance in each cloud site. The following sections detail the design and best practices for ALP architecture.

Deployment Specifications

ALP is distributed as an RPM package that can be installed on the following Linux distributions:

- CentOS Linux 7
- CentOS Linux 8
- Red Hat Enterprise Linux 8
- Photon OS 3 or later,
- Ubuntu 18 or later
- Debian 10 or later

For high availability, the provider will install ALP on two or more Linux VMs in each cloud site or AZ and configure ALP to connect to the local VCD and RabbitMQ instances. Highly-available ALP deployments do not currently support the Message Queue Telemetry Transport (MQTT) message bus.

The next step is to configure ALP to use a Provider Virtual Data Center (PVDC) to create an ALP service organization, Pay-As-You-Go (PAYG) OrgVDC and a service role which is assigned to the service account used during initial configuration. The OrgVDC will be used to host the catalogs which ALP will use to save the applications templates.

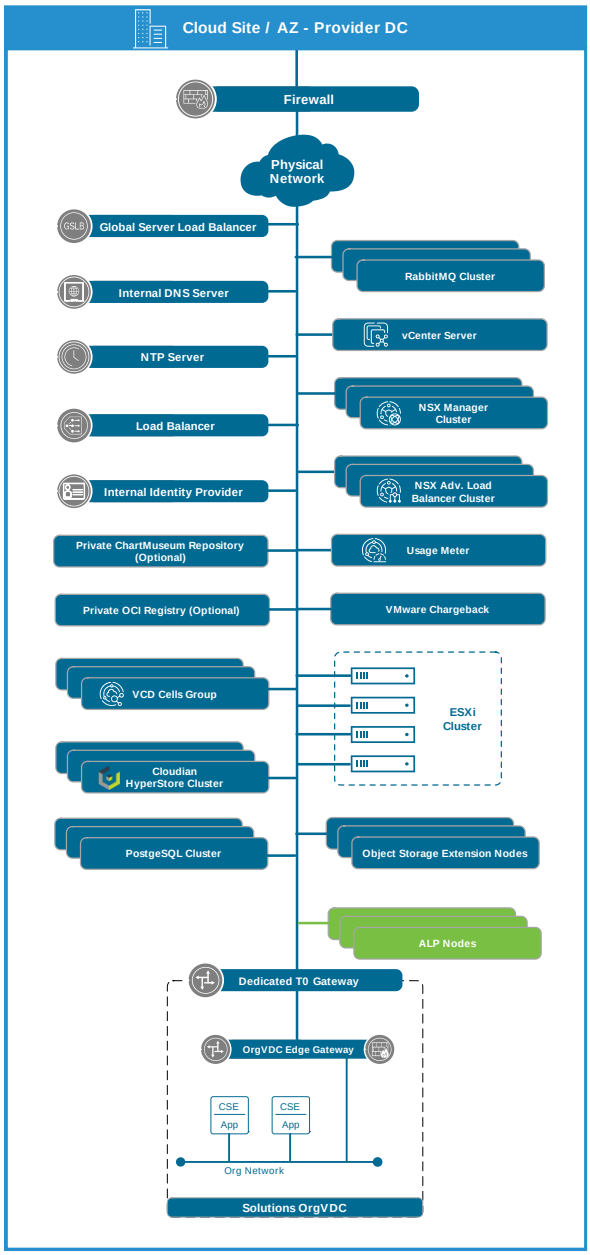

Figure 9: ALP Deployment in each AZ (new components marked in green)

Network Design

Basic Networking

In each cloud site or AZ, all ALP nodes must be able to communicate to the local VCD and RabbitMQ instances. ALP nodes must be able to resolve the VCD public address endpoint FQDN and the FQDN of the RabbitMQ instance to their corresponding internal IP addresses. Refer to the Traffic Flow and Communication Matrix for the complete ALP communication matrix.

External Access Requirements

ALP nodes must have internet access to connect to the following URLs:

- VMware Marketplace: https://gtw.marketplace.cloud.vmware.com/*

- VMware Cloud Services: https://console.cloud.vmware.com/*
- Bitnami Repository for Helm Charts: https://s3.*.amazonaws.com/*, if the provider does not have an internal provider managed ChartMuseum repository.

Refer to the Traffic Flow and Communication Matrix for the complete ALP communication matrix.

Applications Deployment on Kubernetes Clusters

ALP nodes must be able to access tenants' TKG clusters API endpoints to deploy containerized applications. Tenants' TKG clusters must be exposed to an external network which ALP nodes can communicate with. ALP can't deploy containerized applications on isolated TKG clusters.

ALP must be able to download the Helm chart packages from a Helm chart repository. ALP supports only either public Bitnami Repository or a provider managed ChartMuseum repository. In addition, TKG clusters must be able to download the applications images from the public Bitnami registry or any other location referenced in the chart packages.

ALP will follow this workflow to deploy a containerized application on a TKG cluster:

- ALP will pull the Helm chart packages from the repository.
- ALP will push the chart packages to the TKG cluster.
- The TKG cluster will download the application image from the location referenced in the chart package.

Refer to the Traffic Flow and Communication Matrix for the complete ALP communication matrix.

Traffic Flow and Communication Matrix

Figure 10 below shows the complete traffic flows for ALP. Table 10 shows the firewall ports that need to be opened on Tenant and Provider firewalls. Table 11 shows the required NAT rules, and Table 12 shows the URLs that the ALP nodes need to be able to reach to allow consuming containerized applications.

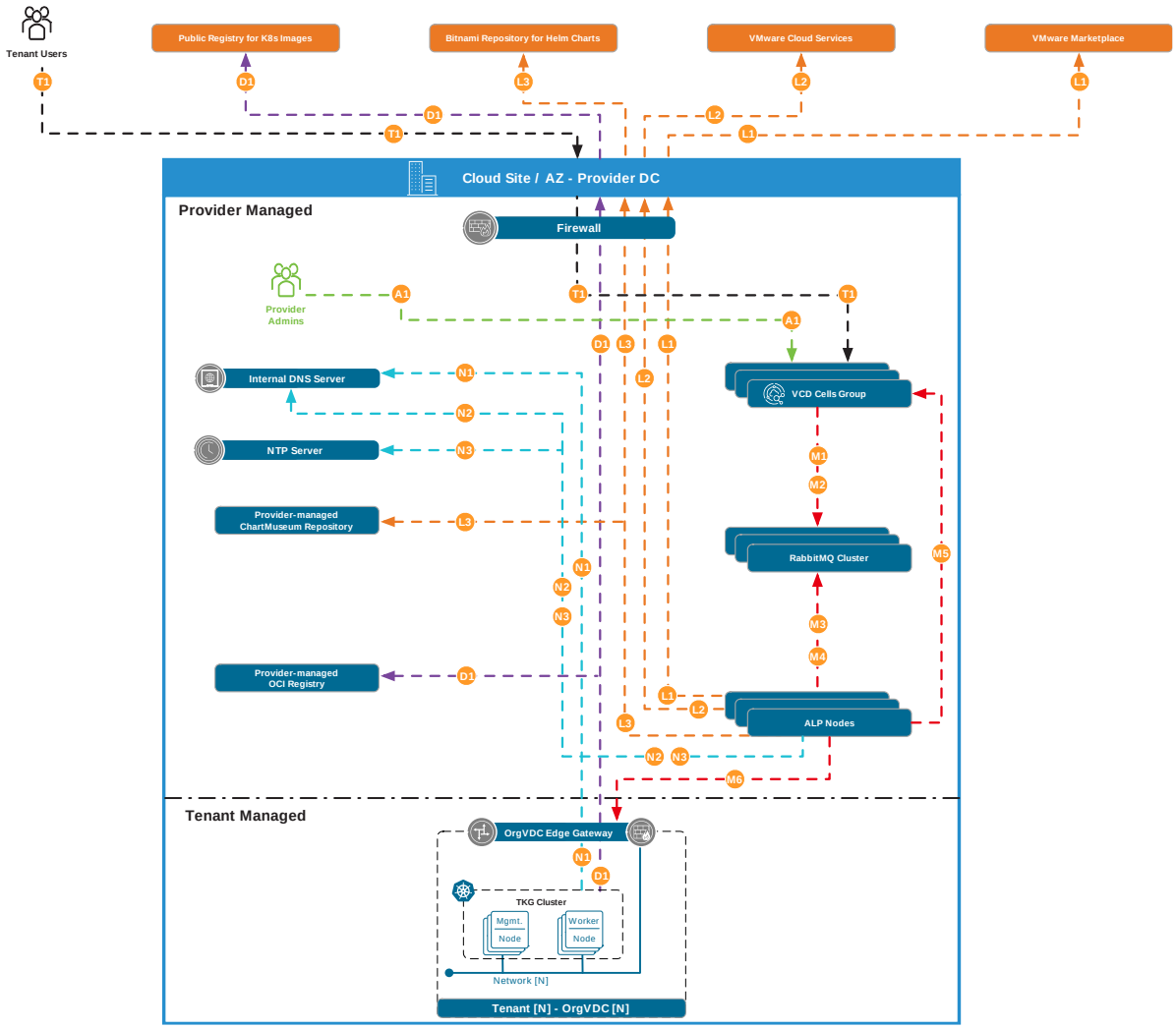

Figure 10: ALP Traffic Flows

Table 10 - ALP Traffic Flows				
Legend	Source	Destination	Protocol / Port	Description
T1	Tenant Users	VCD	TCP 443	Used for access ALP Tenant Portal
A1	Provider Admins	VCD	TCP 443	Used for access ALP Provider Portal
L1	ALP Nodes	VMware Marketplace Service	TCP 443	Used for access to VMware Marketplace
L2	ALP Nodes	VMware Cloud Services	TCP 443	Used for access to VMware Cloud Services
L3	ALP Nodes	Bitnami Object Storage for Helm Charts	TCP 443	Used for downloading K8s applications helm charts
M1	VCD	RabbitMQ	TCP 5671	(Recommended) Used for secured AMQP communication between VCD and RabbitMQ
M2	VCD	RabbitMQ	TCP 5672	(Optional - Not recommended) Used for unsecured AMQP communication between VCD and RabbitMQ

M3	ALP Nodes	RabbitMQ	TCP 5671	(Recommended) Used for secured AMQP communication between ALP nodes and RabbitMQ
M4	ALP Nodes	RabbitMQ	TCP 5672	(Optional - Not recommended) Used for unsecured AMQP communication between ALP nodes and RabbitMQ
M5	ALP Nodes	VCD	TCP 443	Used for communication and API calls from ALP nodes to VCD
M6	ALP Nodes	TKG Cluster API LB Virtual Service	TCP 6443	Used for applications deployment API calls to TKG Cluster API server
D1	TKG Cluster	- Public Image Registry - Provider-managed OCI Registry	TCP 443	Used for downloading K8s applications images
N1	TKG Cluster	DNS Server	TCP/UDP 53	Used for name resolution
N2	ALP Nodes	DNS Server	TCP/UDP 53	Used for name resolution
N3	ALP Nodes	NTP Server	UDP 123	Used for time synchronization

Table 11 – Required NAT Rules for ALP					
Legend	Type	Internal IP	Internal Port	External IP	External port
L1 / L2 / L3	SNAT	ALP nodes	Any	Public IP	Any
D1 / N1	SNAT	TKG Cluster	Any	Public IP	Any

Table 12 – Required URLs for ALP		
Legend	URL	Description
L1	https://gtw.marketplace.cloud.vmware.com/*	Used for access to VMware Marketplace
L2	https://console.cloud.vmware.com/*	Used for access to VMware Cloud Services
L3	https://s3.*.amazonawscom/*	Used for downloading K8s applications helm charts

Lifecycle Management

ALP builds upon the following components:

- VCD
- CSE

All dependencies must be upgraded, if needed, before upgrading ALP. Refer to VMware Interoperability Matrix before commencing with the upgrade.

Multi-site / Multi-Availability Zones Considerations

In a Multi-Site or Multi-AZ configuration, each cloud site will have a dedicated ALP instance installed which connects to the local VCD and RabbitMQ instances. Tenants will use ALP to deploy containerized applications to their TKG clusters in the same cloud site or AZ.

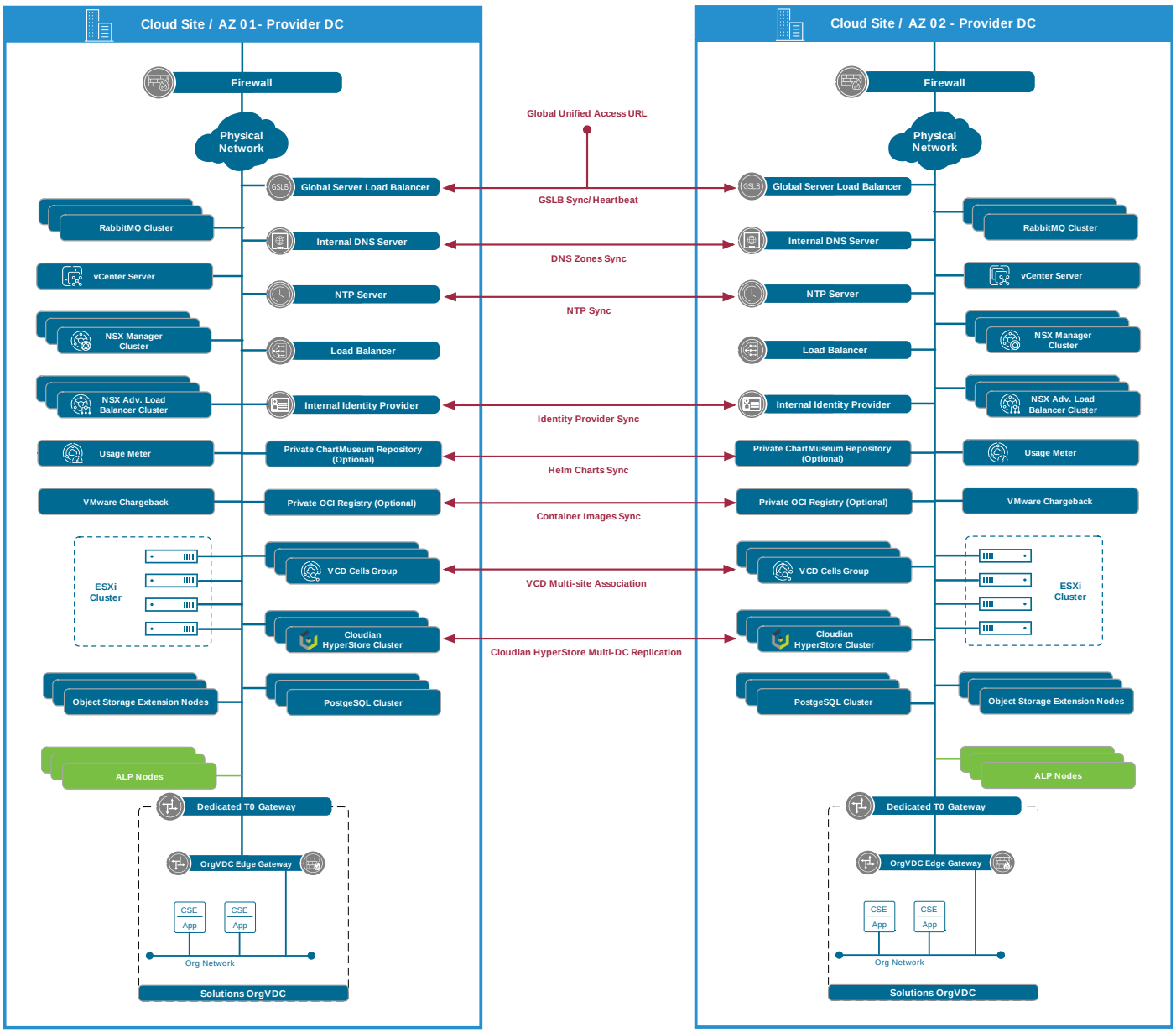

Figure 11: ALP Deployment in Multi-Site / Multi-AZ Configuration

VMware Cloud Director Extension for Data Solutions

The provider should install and onboard DSE to the VCD instances in all cloud sites to deliver a portfolio of on-demand data services based on VMware Tanzu Data Solutions. The following sections detail the design and best practices for DSE.

Deployment Specifications

DSE is an extension for VCD which is based on Solution Add-On framework. DSE is distributed as an ISO package and the provider administrator can use this package to install DSE on the VCD instance. When the provider runs DSE installer on VCD, DSE UI plugin and Run-time Defined Entity (RDE) artifacts are installed.

The next step is to configure the DSE instance with the data solutions packages registry and publish the different data solutions to tenants. Once published, tenants must prepare their TKG clusters by installing the DSE Operator on each TKG cluster. For more information, refer to the VMware Official Documentation.

Note: The VCD cells group has a DSE instance installed in the next figure.

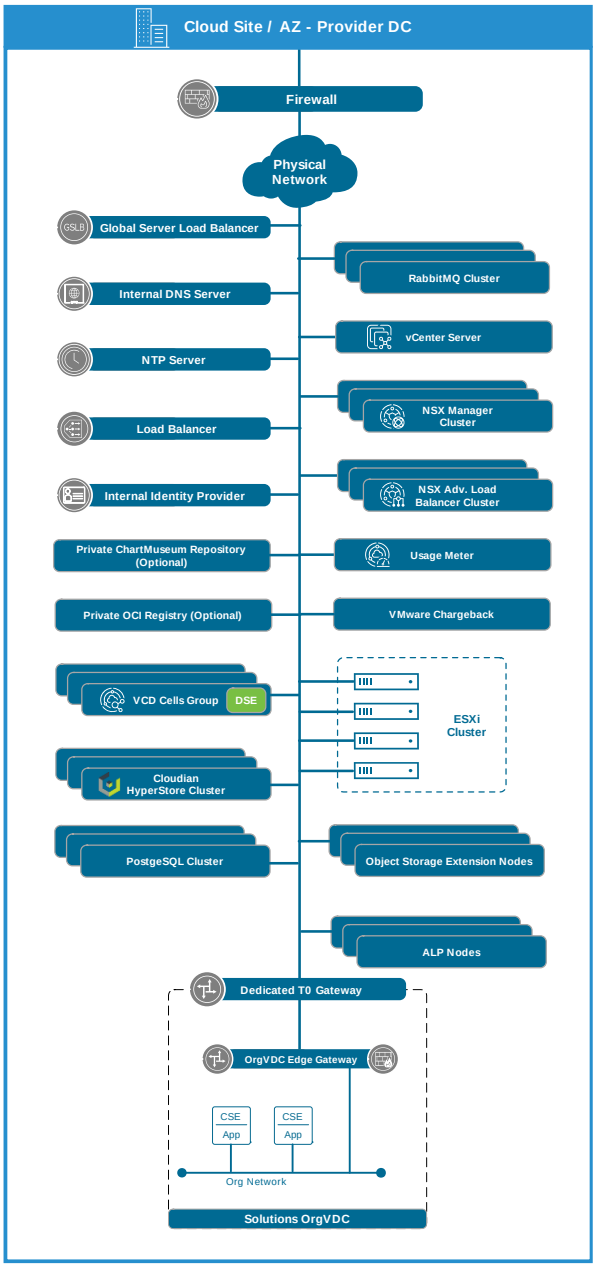

Figure 12: DSE Deployment in each AZ (new components marked in green)

Network Design

The provider must have an account for the VMware Harbor Container Registry for Tanzu. The provider can download DSE packages for data solutions products and push them to a private Open Container Initiative (OCI) registry. If tenants' TKG clusters have public network access, the provider can configure DSE to use VMware Harbor Container Registry for Tanzu.

All tenants' TKG clusters must be able to connect to the local VCD instance so that DSE can deploy and manage data solutions packages on the TKG clusters. The tenants' TKG clusters download the packages from either the public VMware Harbor Container Registry for Tanzu or the provider managed OCI registry.

Traffic Flow and Communication Matrix

Figure 13 below shows the complete traffic flows for DSE. Table 13 shows the firewall ports that need to be opened on Tenant and Provider firewalls. Table 14 shows the required NAT rules, and Table 15 shows the URLs that are required to download DSE operator and Tanzu Data Solutions packages.

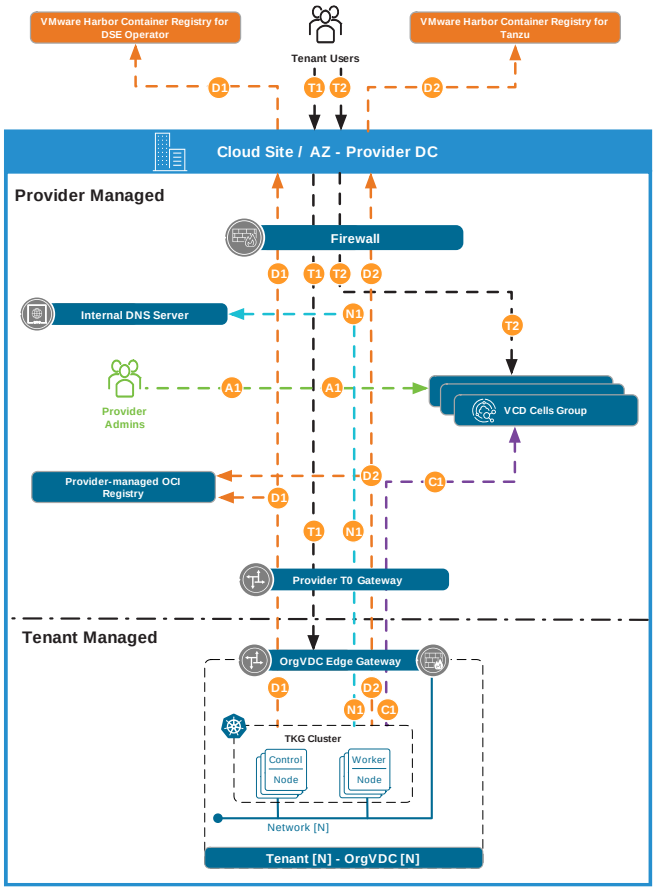

Figure 13: DSE Traffic Flows

Table 13 - DSE Traffic Flows				
Legend	Source	Destination	Protocol / Port	Description
T1	Tenant Users	- TKG Cluster API LB Virtual Service - Tanzu Data Solutions LB Virtual Services	- TCP 6443 - User Selected	- Used for access to TKG K8S Cluster API Service - Used for access to hosted Tanzu Data Solutions on TKG cluster
T2	Tenant Users	VCD	TCP 443	Used for access VCD Tenant Portal / CSE Tenant Portal / DSE Tenant Portal

A1	Provider Admins	VCD	TCP 443	Used for access VCD Provider Portal / CSE Provider Portal / DSE Provider Portal
D1	TKG Cluster	- VMware Harbor Container Registry for DSE Operator - Provider-managed OCI Registry	TCP 443	Used for downloading DSE Operator package
D2	TKG Cluster	- VMware Harbor Container Registry for Tanzu - Provider-managed OCI Registry	TCP 443	Used for downloading Tanzu Data Solutions packages
C1	TKG Cluster	VCD	TCP 443	- Used by TKG cluster to poll VCD to get resources change requests (deploy operators, deploy instances, update, delete, etc.) - Used by TKG cluster to reconcile cluster resources to VCD
N1	TKG Cluster	DNS Server	TCP/UDP 53	Used for name resolution

Table 14 – Required NAT Rules for DSE					
Legend	Type	Internal IP	Internal Port	External IP	External port
T1	DNAT	TKG Cluster	User Selected	Public IP	- TCP 6443 - User Selected
D1 / C1 / N1	SNAT	TKG Cluster	Any	Public IP	Any

Table 15 – Required URLs for DSE		
Legend	URL	Description
D1	https://projects.registry.vmware.com/vcdds/*	Used for downloading DSE Operator package
D2	https://registry.tanzu.vmware.com/*	Used for downloading Tanzu Data Solutions packages

Lifecycle Management

DSE builds upon the following components:

- VCD
- CSE
- Data Solutions products packages

To upgrade DSE, make sure that all dependencies are upgraded, if needed, before upgrading DSE. Refer to VMware Interoperability Matrix before commencing with the upgrade.

Multi-Site / Multi-Availability Zones Considerations

In Multi-Site or Multi-AZ configuration, each VCD instance in each cloud site will have a dedicated DSE instance installed. Each DSE instance will deploy and manage Tanzu Data Solutions packages to the local tenants' TKG clusters. In case of using a private OCI registry for data solutions packages, each cloud site must have its own registry to prevent any dependencies between different cloud sites or AZs.

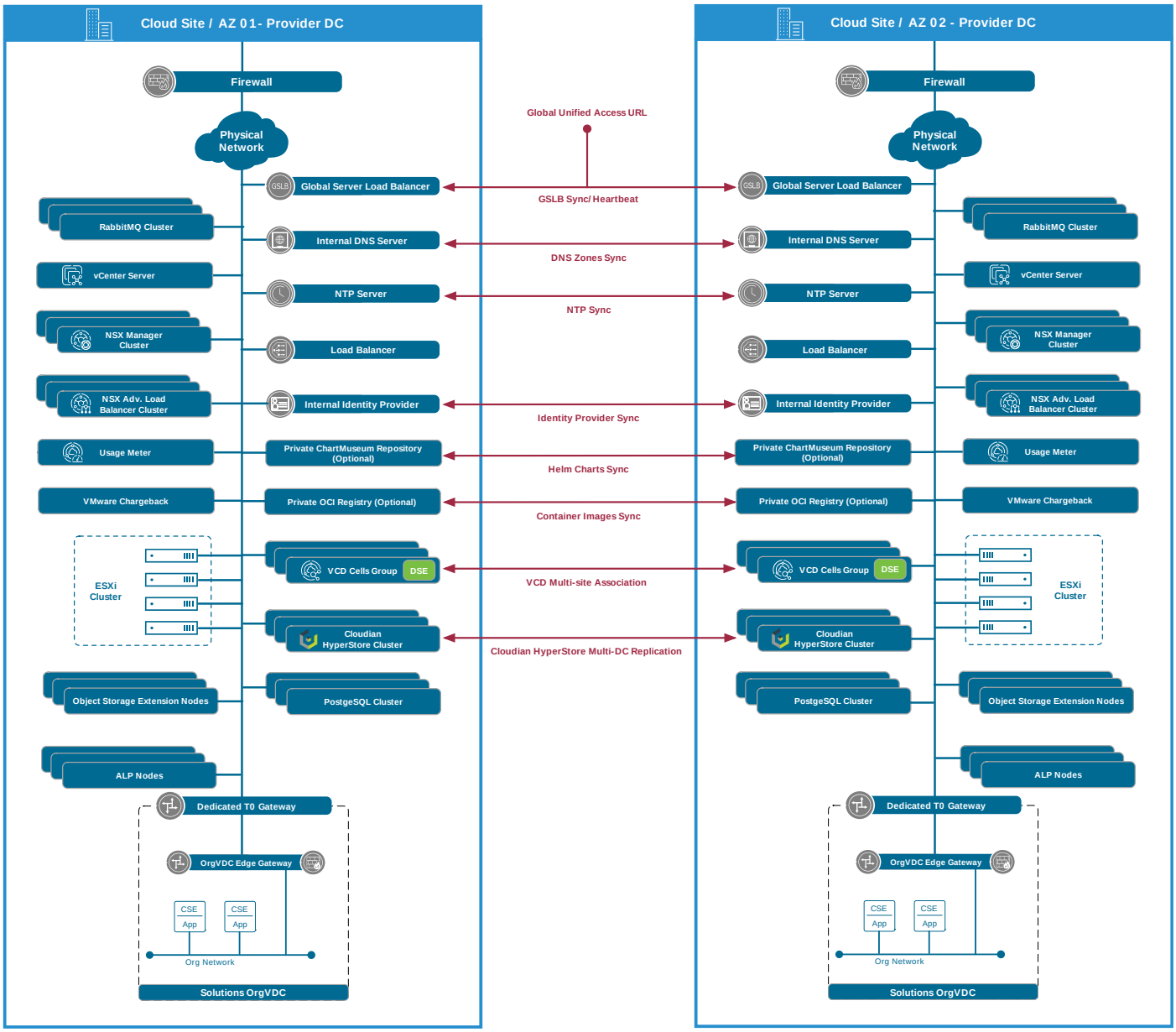

Figure 14: DSE Deployment in Multi-Site / Multi-AZ Configuration

Usage Reporting

Provider Usage Reporting

VMware Cloud Director Containers Service Extension

The provider must report and meter TKG usage data in all cloud sites. There are two options:

- The provider can manually report TKG clusters usage per tenant to Commerce portal for CSP-Cloud Builders. To improve manual reporting experience, the provider can configure a custom placement policy for each customer. This allows providers to have a specific resource pool for TKG clusters. The provider can then label this resource pool for TKG usage and easily report usage specific to the created label.
- Alternatively, the provider can use VMware Aria Operations and create a custom view to track CSE provisioned TKG clusters and report the usage manually through the commerce portal. This will require VMware Aria Operations Advanced or Enterprise license and the latest Aria Operations Management Pack for Cloud Director. The Cloud Director Adapter must be enabled to collect CSE-managed TKG Clusters metrics. For more information about how to create custom views in Aria Operations, refer to the VMware Official Documentation.

It is important to note that TKG cluster VMs still need to be metered and reported separately as normal VMs for usage of other products within or outside Flex-Core bundle.

VMware Cloud Director Object Storage Extension

The provider must report and meter OSE usage data with the Cloudian Hyperstore in all cloud sites. To generate the report, the provider must leverage Cloudian's Smart Support feature that provides system usage and status information.

Smart Support employs proactive analysis and alerts to help the Cloudian Support team maximize system uptime and performance. The provider will have to enable Smart Support on Cloudian's system to be able to report usage. If enabled, system logs are generated once per day and sent to Cloudian Support. Communication is one-way only. Data is sent only via HTTPS through port 443. No user object data is either transmitted or accessible via Smart Support.

Cloudian will generate a month-end report based on the data from Smart Support system. This monthly usage report will then be used for reporting in the commerce portal. For more information, refer to the latest Product Usage Guide which is available through Partner Connect Portal.

VMware Cloud Director App Launchpad

ALP comes with no additional cost. It is included within the Flex-Core bundle.

VMware Cloud Director Extension for Data Solutions

DSE comes with no additional cost. The plugin is included within the Flex-Core bundle. However, the provider must report and meter the consumption of VMware Data Solutions (RMQ/SQL) on 'a per core' basis.

The usage for VMware SQL with MySQL and VMware SQL with PostgreSQL is calculated based on the used CPU cores for each container where the service runs. To calculate the usage of those services, identify all containers where they run. Determine the configured vCPUs for each container and add the count of configured vCPUs and convert to Cores. For RMQ, all configured Kubernetes vCPU Resource Limits for Containers with RMQ Nodes must be identified for reporting. Containers with Kubernetes Operators do not need to be identified for reporting. For more information, refer to the latest Product Usage Guide which is available through Partner Connect Portal.

Tenant Showback / Chargeback

Tenant showback / chargeback is an important aspect in any new service offering. The following sections examine the different methods that the provider can do showback / chargeback for the tenant usage of the different products with KaaS offering.

VMware Cloud Director Containers Service Extension

Using CSE plugin, the tenants can monitor their TKG clusters and view:

- TKG clusters basic information.
- TKG clusters Kubernetes resource information.
- TKG clusters vApp details.

The provider can use VMware Chargeback 8.10 for tenant's showback/chargeback. VMware Chargeback 8.10 natively integrates with VCD and reports tenant's CSE usage. The provider can configure pricing policies for CSE usage which automatically generates a monthly bill to the tenant about their TKG clusters cost. This will require VMware Aria Operations Advanced or Enterprise license and the latest Aria Operations Management Pack for Cloud Director. The Cloud Director Adapter must be enabled to collect CSE-managed TKG Clusters metrics.

Alternatively, the provider can use VCD / CSE APIs to collect tenant's TKG Clusters and their specifications. Then, the provider should feed these attributes to an external billing system to generate the required bills for each tenant. For more information about VCD/ CSE APIs, refer to this blog post.

VMware Cloud Director Object Storage Extension

Using OSE plugin, tenants can monitor their object storage consumption and view:

- Number of Buckets, objects, consumed storage and the number of users consuming object storage.
- TKG clusters backup status and schedules.
- TKG clusters restore operations status and history.

For tenant's showback/chargeback, the provider needs to consider the tenant's object storage usage, the storage policy assigned to the tenant, and if replication across sites is enabled.

The provider can use VMware Chargeback 8.10. VMware Chargeback 8.10 natively integrates with VCD and reports tenant's usage of object storage. The provider can configure pricing policies which automatically generates a monthly bill to the tenant about their object storage cost. This will require VMware Aria Operations Advanced or Enterprise license and the latest Aria Operations Management Pack for Cloud Director. The Cloud Director Adapter must be enabled to collect object storage usage metrics from OSE.

Alternatively, the provider can use Cloudian HyperStore APIs to generate the usage data of each tenant and the assigned storage policy, then feeds this usage data into an external billing system to generate the required bills for each tenant.

VMware Cloud Director App Launchpad

Tenants can monitor their usage of different containerized applications offered through ALP by using ALP tenant portal. Users can view:

- Number of entitled applications.
- Top applications deployed by tenants.
- Top users who deployed application instances.
- Total number of running application instances.

- Number of instances deployed from each application.

The same information can be collected using ALP APIs. For more information about ALP APIs, refer to the VMware Official Documentation.

Currently, there is no automated way for the provider to showback/chargeback tenant's usage of ALP. The provider can showback/chargeback tenant's usage of ALP and the offered containerized applications by extracting each tenant's usage of different applications by using ALP APIs. Then, the provider should feed this usage into an external billing system to generate the required bills for each tenant. For more information about ALP APIs, refer to the VMware Official Documentation.

VMware Cloud Director Extension for Data Solutions

Tenants can monitor their usage of Tanzu Data Solutions by using the DSE plugin. Users can view:

- Entitled data solutions.
- The number of installed instances of each data solution and on which TKG clusters.
- The properties of each installed instance.

Tenants can also monitor the installed instances by using Prometheus and Grafana. DSE uses Prometheus and Grafana for data monitoring, alerting and visualization. They must be installed on a Tanzu Kubernetes workload cluster. The detailed steps can be found here.

Currently, there is no automated way for the provider to showback/chargeback tenant's usage of Tanzu Data Solutions. The provider can showback/chargeback tenant's Tanzu Data Solutions usage by extracting the number of vCPUs consumed by the installed instances of each data solution using automated scripts, then feed it into an external billing system to generate the required bills for each tenant.

Glossary

ALP	VMware Cloud Director App Launchpad
AMQP	Advanced Message Queuing Protocol
AZ	Availability Zone
CEIP	Customer Experience Improvement Program
CMC	Cloudian Management Console
CSE	VMware Cloud Director Containers Service Extension
DSE	VMware Cloud Director Extension for Data Solutions
GLB	Global Load Balancer
IaaS	Infrastructure-as-a-Service
KaaS	Kubernetes-as-a-Service
MQTT	Message Queue Telemetry Transport
OCI	Open Container Initiative
OrgVDC	Organization Virtual Data Centre
OSE	VMware Cloud Director Object Storage Extension
OVA	Open Virtualization Appliance
PAYG	Pay-as-you-go
PVDC	Provider Virtual Data Centre
RDE	Run-time Defined Entity
RMQ	RabbitMQ
SANs	Subject Alternative Names
TKG	Tanzu Kubernetes Grid
UI	User Interface
VCD	VMware Cloud Director

About the Author

Shady Ali ElMalatawey is a Senior Staff Solutions Architect in the global Multi-cloud Architecture team. He has 13 years of experience and has been in different roles in VMware since joining in 2018. He is working with VMware Cloud Service Providers to provide architecture best practices and technical guidance on different VMware products under the VMware Partner Connect program. He holds three VMware Certified Design Expert (VCDX) certificates, and his number is 249. In addition, he is a VMware Education SME, an exam contributor and a VCDX Panelist. He has been member of vExpert Program since 2014, vExpert Pro Program since 2020, vExpert NSX Program since 2022 and vExpert Cloud Providers Program since 2020.

Acknowledgments and Reviewers

Special Acknowledgment to:

Milind Gadre - Vice President Engineering, VCPP Engineering

Reviewers in alphabetic order:

Avnish Kumar Tripathi - Staff Multi-Cloud Solutions Architect, APJ

Gerrit Lehr - Principal Multi-Cloud Strategist, EMEA

Joseph Polcar - Staff Multi-Cloud Solutions Architect, AMER

Justin Brandon - Multi-Cloud Solutions Architect, AMER

Matt Elliott - Staff Multi-Cloud Solutions Architect, AMER

Sangeeta Panigrahy – Senior Product Manager, VCPP Product Management

Sean Massey - Staff Multi-Cloud Solutions Architect, AMER

Stephen Evanchik - Principal Engineer, VCPP Engineering

Steven Zhang - Staff II Engineer, VCPP Engineering

Timo Sugliani - Senior Staff Multi-Cloud Solutions Architect, EMEA

Urbano Criola - Staff Multi-Cloud Solutions Architect, APJ

William Zhoe – Senior Engineering Manager, VCPP Engineering