vmware

Deploying Horizon 7 on VMware Cloud on AWS

VMware End User Computing Group

VMware, Inc.
3401 Hillview Ave
Palo Alto, CA 94304
www.vmware.com

Introduction

VMware Horizon® 7 for VMware Cloud™ on AWS delivers a seamlessly integrated hybrid cloud for virtual desktops and applications. It combines the enterprise capabilities of the VMware Software-Defined Data Center (SDDC), delivered as a service on AWS, with the market-leading capabilities of VMware Horizon 7 for a simple, secure, and scalable solution. You can easily address use cases such as on-demand capacity, disaster recovery, and cloud co-location without buying additional data center resources.

For customers who are already familiar with Horizon 7 or have Horizon 7 deployed on-premises, deploying Horizon 7 on VMware Cloud on AWS lets you leverage a unified architecture and familiar tools. This means that you use the same expertise you know from VMware vSphere® and Horizon 7 for operational consistency and leverage the same rich feature set and flexibility you expect. By outsourcing the management of the vSphere platform to VMware, you can simplify management of Horizon 7 deployments. For more information about VMware Horizon 7 for VMware Cloud on AWS, visit the Horizon 7 on VMware Cloud on AWS product page.

The purpose of this guide is to provide IT administrators with a set of steps and best practices on how to deploy Horizon 7 on VMware Cloud on AWS. This guide is designed to be used in conjunction with Horizon 7 documentation, VMware Workspace ONE and VMware Horizon 7 Enterprise Edition On-premises Reference Architecture guide, and VMware Cloud on AWS documentation.

Overview of Horizon 7 on VMware Cloud on AWS

You can deploy Horizon 7 on VMware Cloud on AWS to scale Horizon 7 desktops and applications on an elastic cloud platform.

VMware Cloud on AWS allows you to create vSphere Software-Defined Data Centers (SDDCs) on Amazon Web Services. These SDDCs include VMware vCenter Server® for VM management, VMware vSAN™ for storage, and VMware NSX® for networking. You can connect an on-premises SDDC to your cloud SDDC, and manage both from a single VMware vSphere® Web Client interface. Using your connected AWS account, you can access AWS services such as EC2 and S3 from virtual machines in your SDDC. For more information, see the VMware Cloud on AWS documentation.

Once you have deployed an SDDC on VMware Cloud on AWS, you can deploy Horizon 7 in that cloud environment just like you would in an on-premises vSphere environment. This enables Horizon 7 customers to outsource the management of the SDDC infrastructure to VMware. There is no requirement to purchase new hardware, and you can use the pay-as-you-go option for hourly billing on VMware Cloud on AWS.

Cloud Pod Architecture (CPA) is a Horizon 7 feature that allows you to scale your Horizon 7 deployment across multiple pods and sites for federated management. You can deploy Horizon 7 in a hybrid cloud environment when you use CPA to interconnect on-premises data centers and VMware Cloud on AWS data centers.

Important: A single pod and the Connection Servers in it must be located within a single data center and cannot span locations. Multiple locations must have their own separate pods. These pods can be managed individually or interconnected using Cloud Pod Architecture (CPA)

Since the Horizon 7 architecture is the same on-premises and in VMware Cloud on AWS, the deployment and management experience remain the same across on-premises sites and in the cloud. When using

multiple data centers, you must use a storage replication mechanism, such as DFS-R in a hub-spoke topology, for replicating user data.

You can also stretch CPA across two or more VMware Cloud on AWS data centers. Of course, use of CPA is optional. You can choose to deploy Horizon 7 exclusively in a single VMware Cloud on AWS data center without linking it to any other data center.

For details on feature parity between Horizon 7 on-premises and Horizon 7 on VMware Cloud on AWS, as well as interoperability of Horizon 7 and VMware Cloud on AWS versions, see the VMware Knowledge Base article Horizon 7 on VMware Cloud on AWS Support (58539).

Horizon 7 Deployment Scenarios on VMware Cloud on AWS

You can deploy Horizon 7 on VMware Cloud on AWS for the following scenarios.

Data Center Expansion	Use this scenario if you have an existing on-premises Horizon 7 infrastructure and need to expand capacity but don't want to procure additional hardware. Extend the Horizon 7 deployment to VMware Cloud on AWS by using Cloud Pod Architecture to connect on-premises pods with a pod in VMware Cloud. With this strategy, you can use cloud capacity and still manage on-premises and private cloud deployments in a single federated space. You can also utilize the cloud platform to provide temporary capacity for contractors and seasonal workers. The on-premises deployment is optional. Based on your needs, you can decide to consolidate and move the on-premises deployment completely to VMware Cloud on AWS.
Application Locality	Use this scenario when you want to move published applications that are latency-sensitive to VMware Cloud on AWS and need virtual desktops and RDS (Remote Desktop Session) hosts to be co-located with your published applications. You can also have other published applications that are still on-premises. When you extend your Horizon 7 deployment to VMware Cloud on AWS, you can allow end users to connect to the nearest virtual desktop or RDS host to launch the application regardless of whether the application is on-premises or on VMware Cloud on AWS.
Business Continuity (BC) and Disaster Recovery (DR)	The cost of building an on-premises BCDR infrastructure can be high. When you use VMware Cloud on AWS, you pay for the use of BCDR infrastructure during those times when the primary infrastructure is down or when you require a small pilot during normal operations for a quick Recovery Time Objective (RTO) during a disaster event. Having a unified Horizon 7 architecture across the primary site on-premises and the BCDR site on VMware Cloud on AWS makes the failover process simple. You can also deploy Cloud Pod Architecture across multiple VMware Cloud on AWS data centers for BCDR.

Deployment Architecture for Horizon 7 on VMware Cloud on AWS

Understanding Key Components of Horizon 7 on VMware Cloud on AWS

To set up a successful VMware Cloud on AWS deployment, you must configure the logical network that can support a Horizon 7 deployment on VMware Cloud on AWS.

Include the following components in the logical network configuration. Note that this document describes NSX-T components.

Management Component	The management component for the network includes vCenter Server.
Compute Component	The compute component for the network includes the following components: Unified Access Gateway appliances Load balancer Horizon Connection Servers Virtual machines
NSX-T Components	VMware NSX Data Center is the network virtualization platform for the Software-Defined Data Center (SDDC), delivering networking and security entirely in software, abstracted from the underlying physical infrastructure. The maximum number of ports per logical network is 1000. And since multiple VLANs are not supported on NSX with Horizon, the maximum size of the Horizon 7 pool is limited to 1000. Of course, you can create multiple pools using different logical networks. • Tier-0 router. Handles Internet, route or policy based IPSEC VPN, AWS Direct Connect and also serves as an edge firewall for the Tier-1 Compute Gateway (CGW). • Tier-1 Compute Gateway (CGW). Serves as a distributed firewall for all customer internal networks. • The Tier-1 Management Gateway (MGW). Serves as a firewall for the VMware maintained components like vCenter and NSX.

Horizon 7 Pod and Building Block On-premises

A typical Horizon 7 architecture design uses a pod strategy. A pod is a unit of organization determined by Horizon 7 scalability limits. Each pod has a separate management UI and therefore the typical design is to minimize the number of pods.

Customers usually include multiple building blocks in a Horizon 7 pod on-premises. A building block is a logical construct and should not be sized for more than the maximum number of desktops tested. See the VMware Knowledge Base article VMware Horizon 7 sizing limits and recommendations (2150348).

A building block consists of:

- Physical servers
- 1 vCenter Server and vSphere infrastructure
- Horizon 7 servers
- Shared storage
- Virtual desktops and/or RDS hosts for end users

Horizon 7 Pod on VMware Cloud on AWS

Horizon 7 pod architecture on VMware Cloud on AWS is slightly different from the on-premises architecture. On VMware Cloud on AWS:

- Each Horizon 7 pod consists of a single SDDC.
- Each SDDC contains a single vCenter Server, which means each pod has only a single vCenter Server.
- This architecture significantly reduces complexity because a Horizon 7 pod essentially consists of a single building block.

At this time, each SDDC only has 1 compute gateway, so that the number of connections is limited to ~2,000 VMs or sessions (the actual number depends on your VM traffic). This effectively limits the number of VMs or RDSH sessions per Horizon 7 pod to ~2,000 as well. Once the number of compute gateways per SDDC is increased, Horizon 7 on VMware Cloud on AWS will have comparable scalability with Horizon 7 on-premises.

Resource Pools

A resource pool is a logical abstraction for flexible management of resources. Resource pools can be grouped into hierarchies and used to hierarchically partition available CPU and memory resources.

Within a Horizon 7 pod on VMware Cloud on AWS, you can use vSphere resource pools to separate management components from virtual desktops or published applications workloads to make sure resources are allocated correctly.

After an SDDC instance on VMware Cloud on AWS is created, two resource pools exist:

- A Management Resource Pool with reservations that contains vCenter Server plus NSX, which is managed by VMware
- A Compute Resource Pool within which everything is managed by the customer

We recommend creating two sub–resource pools within the Compute Resource Pool for your Horizon 7 deployments:

- A Horizon Management Resource Pool for your Horizon 7 management components, such as connection servers
- A Horizon User Resource Pool for your desktop pools and published apps

See Figure 1 for schematics of the recommended architecture. Because the management components of Horizon 7 are shared among all virtual machines, you can avoid having any single virtual machine affect overall performance by deploying the management components in a separate resource pool with reservations. Alternatively, you can use different clusters to separate these components.

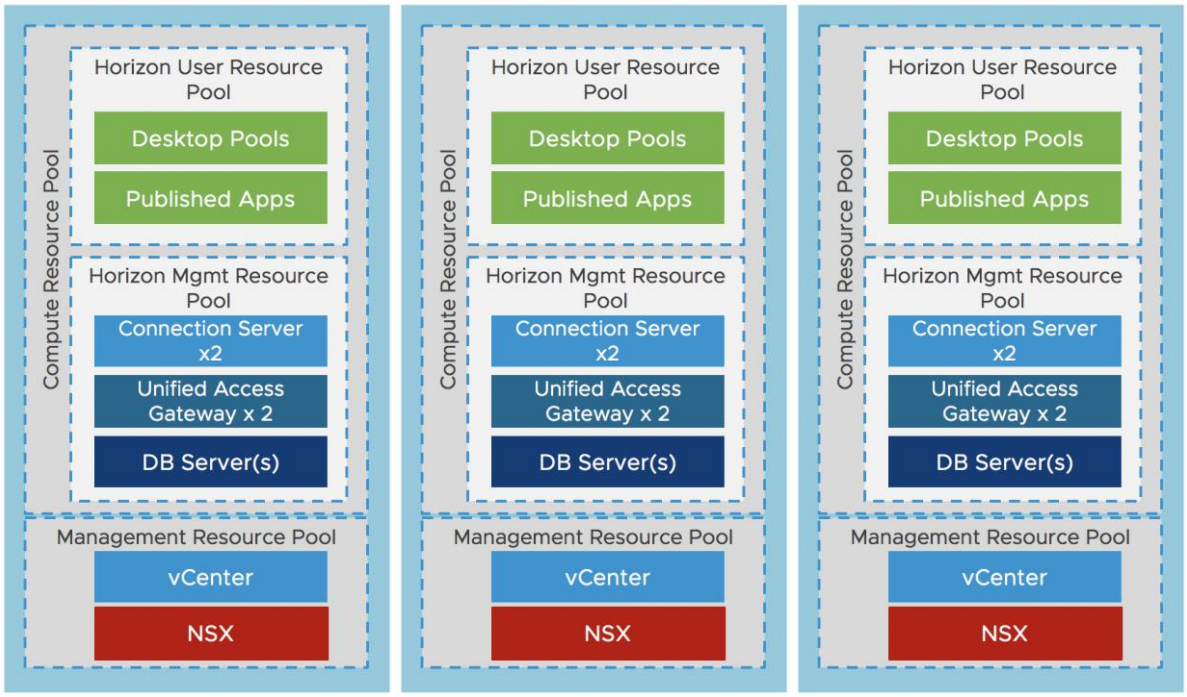


Figure 1: Horizon 7 Pod Architecture on VMware Cloud on AWS

Memory Reservations

Because physical memory cannot be shared between virtual machines, and because swapping or ballooning should be avoided at all costs, be sure to reserve all memory for all Horizon virtual machines, including management components, virtual desktops, and RDS hosts.

CPU Reservations

CPU reservations are shared when not used, and a reservation specifies the guaranteed minimum allocation for a virtual machine. For the management components, the reservations should equal the number of vCPUs times the CPU frequency (currently 2300 with VMware Cloud on AWS). Any amount of CPU reservations not actively used by the management components will still be available for virtual desktops and RDS hosts when they are not deployed to a separate cluster.

Virtual Machine–Level Reservations

As well as setting a reservation on the resource pool, be sure to set a reservation at the virtual machine level. This ensures that any VMs that might later get added to the resource pool will not consume resources that are reserved and required for HA failover. These VM-level reservations do not remove the requirement for reservations on the resource pool. Because VM-level reservations are taken into account only when a VM is powered on, the reservation could be taken by other VMs when one VM is powered off temporarily.

Leveraging CPU Shares for Different Workloads

Because RDS hosts can facilitate more users per vCPU than virtual desktops can, a higher share should be given to them. When desktop VMs and RDS host VMs are run on the same cluster, the share allocation should be adjusted to ensure relative prioritization.

As an example, if an RDS host with 8 vCPUs facilitates 28 users and a virtual desktop with 2 vCPUs facilitates a single user, the RDS host is facilitating 7 times the number of users per vCPU. In that scenario, the desktop VMs should have a default share of 1000, and the RDS host VMs should have a vCPU share of 7000 when not deployed to a separate cluster. This number should also be adjusted to the required amount of resources, which could be different for a VDI virtual desktop session versus a shared RDSH-published desktop session.

Table 1: **Reservations and Shares Overview**

	Resource Pool		**VM**		
	Reservation		**Reservation**		**Shares**
	Memory	**CPU**	**Memory**	**CPU**	**CPU**
Management	Full	Full (vCPU*freq)	Full	Full (vCPU*freq)	No
VDI	Full	No	Full	No	Default
RDSH	Full	No	Full	No	By ratio

Sizing Horizon 7 on VMware Cloud on AWS

Similar to deploying Horizon 7 on-premises, you will need to size your requirements for deploying Horizon 7 on VMware Cloud on AWS to determine the number of hosts you will need to deploy. Hosts are needed for the following purposes:

- Your virtual desktop or RDS workloads
- Your Horizon 7 infrastructure components such as connection servers, Unified Access Gateways, App Volumes managers, etc
- SDDC infrastructure components on VMware Cloud on AWS. These components are deployed and managed automatically for you by VMware, but you will need capacity in your SDDC for running them.

The methodology for sizing Horizon 7 on VMware Cloud on AWS is exactly the same as for on-premises. What is the different (and simpler) is the fixed hardware configurations on VMware Cloud on AWS. Work with your VMware sales team to determine the correct sizing.

Minimum SDDC Size

At the time of this update, the minimum number of hosts required per SDDC on VMware Cloud on AWS for production use is 3 nodes (hosts). For testing purpose, a 1-node SDDC is also available. However, since a single node does not support HA, we do not recommend it for production use. Horizon 7 can be deployed on a single node SDDC or a multi-node SDDC. If you are deploying on a single node SDDC, be sure to change the FTT policy setting on vSAN from 1 (default) to 0.

Network Configuration for Horizon 7 Deployment on VMware Cloud on AWS

The following section describes network configuration using NSX-T.

After you deploy an SDDC instance on VMware Cloud on AWS, two isolated networks exist, a management network and a compute network. Each has its own NSX Edge Gateway and an NSX Distributed Logical Router for extra networks in the compute section.

The recommended network architecture consists of a double DMZ and a separation between Horizon management components and the RDSH and VDI virtual machines.

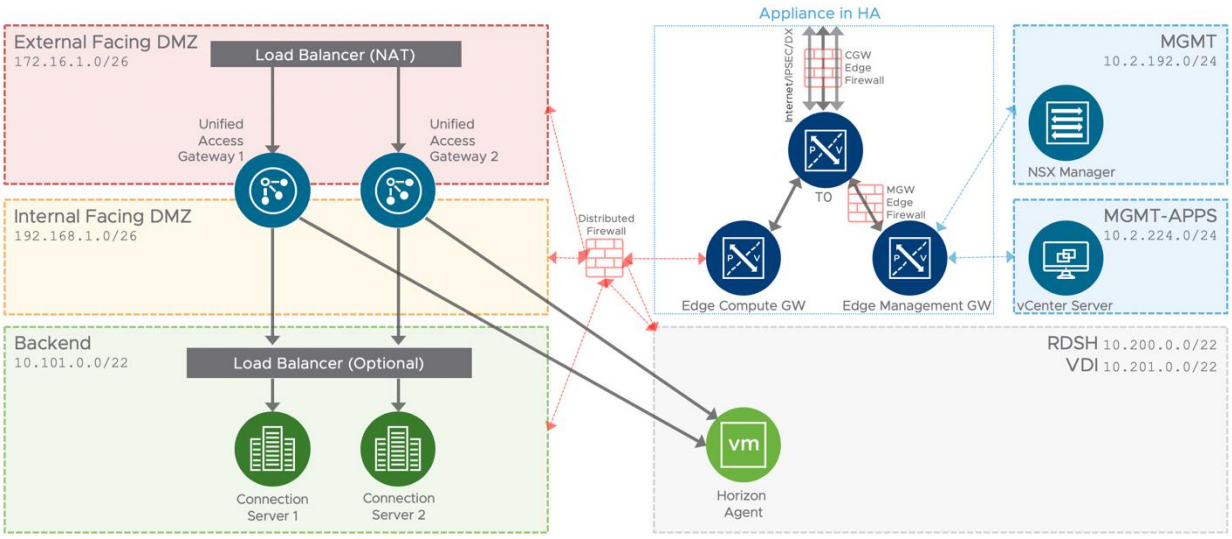


Figure 2: Network Diagram (Subnets are for Illustrative Purposes Only)

Because the Horizon Connection Server must communicate with the vCenter Server, traffic must be allowed on the MGW Edge Firewall.

A third-party load balancer such as F5 LTM or AWS Elastic Load Balancer (ELB) must be deployed to allow multiple Unified Access Gateway appliances and Connection Servers to be implemented in a highly available configuration.

When direct external access is required, configure a public IP address with Network Address Translation towards the Unified Access Gateway virtual IP of the load balancer.

For external management or access to external resources, create a VPN or direct connection to the tier 0 router (illustrated as a light-grey line in the diagram). You can configure a route-based IPsec VPN or a policy-based IPsec VPN.

- Route-based VPN uses the routed tunnel interface as the endpoint of the SDDC network to allow access to multiple subnets within the network. Local and remote networks are discovered using BGP advertisements.
- Policy-based VPN allows access to a subnet of the SDDC network.

AWS Direct Connect (DX) is a cloud service solution that makes it easy to establish a dedicated network connection from your premises to AWS. Using AWS Direct Connect, you can establish private connectivity between AWS and your datacenter, office, or colocation environment, which in many cases can reduce your network costs, increase bandwidth throughput, and provide a more consistent network experience than Internet-based connections.

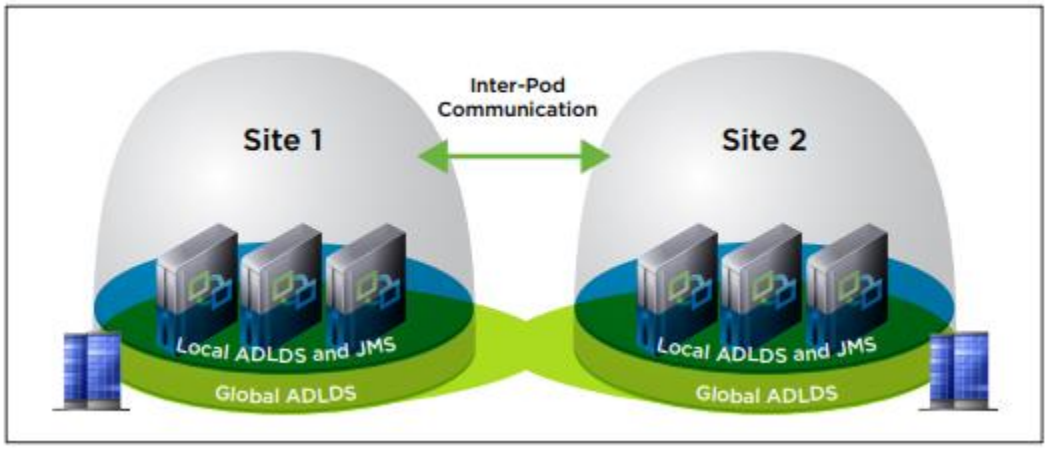
AWS Direct Connect lets you establish a dedicated network connection between your network and one of the AWS Direct Connect locations. Using industry standard 802.1q VLANs, this dedicated connection can be partitioned into multiple virtual interfaces. This allows you to use the same connection to access public resources such as objects stored in Amazon S3 using public IP address space, and private resources such as Amazon EC2 instances running within an Amazon Virtual Private Cloud (VPC) using private IP space, while maintaining network separation between the public and private environments. Virtual interfaces can be reconfigured at any time to meet your changing needs.

Architecting Horizon 7 Cloud Pod Architecture (CPA) for VMware Cloud on AWS

Cloud Pod Architecture (CPA) is a standard Horizon 7 feature that allows you to connect your Horizon 7 deployment across multiple pods and sites for federated management. It can be used to scale up your deployment, to build hybrid cloud, and to provide redundancy for Business Continuity and Disaster Recovery. CPA introduces the concept of a global entitlement (GE) that spans the federation of multiple Horizon pods and sites. Any users or user groups belonging to the global entitlement are entitled to access virtual desktops and RDS published apps on multiple Horizon 7 pods that are part of the CPA.

Important: CPA is not a stretched deployment; each Horizon 7 pod is distinct and all Connection Servers belonging to each of the individual pods are required to be located in a single location and run on the same broadcast domain from a network perspective.

Here is a logical overview of a basic two site/ two pod CPA implementation



For the full documentation on how to set up and configure CPA, refer to Administering View Cloud Pod Architecture in the Horizon 7 documentation and VMware Workspace ONE and VMware Horizon 7 Enterprise Edition On-premises Reference Architecture

Below, we will discuss how CPA can be used for Horizon 7 on VMware Cloud on AWS

Using CPA to build Hybrid Cloud and Scale for Horizon 7

You can deploy Horizon 7 in a hybrid cloud environment when you use CPA to interconnect Horizon 7 on-premises and Horizon 7 pods on VMware Cloud on AWS. You can easily entitle your users to virtual desktop and RDS published apps on-premises and/or on VMware Cloud on AWS. You can configure it such that they can connect to whichever site is closest to them geographically as they roam.

You can also stretch CPA across Horizon 7 pods in two or more VMware Cloud on AWS data centers with the same flexibility to entitle your users to one or multiple pods as desired

Of course, use of CPA is optional. You can choose to deploy Horizon 7 exclusively in a single VMware Cloud on AWS data center without linking it to any other data center.

Using CPA to Provide Business Continuity (BC) and Disaster Recover (DR) for Horizon 7

Unlike traditional BCDR solution for apps where replication of all data from primary site to secondary site is needed, we recommend a different approach for Horizon 7, using CPA. Since majority of VDI and RDS deployments use non-persistent and stateless virtual machines that can be created and recreated very quickly, it is senseless to replicate them across sites. CPA can be used across on-premise Horizon 7 pods (primary site) and Horizon 7 pods on VMware Cloud on AWS (secondary site) for the purpose of BCDR. By using VMware Cloud on AWS as a secondary site for BCDR, you can take advantage of the hourly billing option and the pay-as-you-go benefit.

During normal operations, keep a small host footprint on VMware Cloud on AWS where you will deploy your Horizon 7 instance, store your updated golden images and create a small pool of VMs. Note that there is a minimum number of hosts requirement per SDDC. When the primary site goes down, you can simply create the new virtual desktops as well as new hosts on the secondary site from the exact same golden image. Use Global Entitlements to ensure that your end-users can access desktops on the secondary site.

You will need to keep persistent data such as user profiles, user data, and golden images synced between the two sites by using a storage replication mechanism, such as DFS-R in a hub-spoke topology or another 3rd party file share technology. If you also use App Volumes and User Environment Manager, appstacks and file share data will also need to be replicated from the primary site to the secondary site.

An important consideration in leveraging VMware Cloud on AWS as a secondary site for BCDR involves host availability at the AWS data center when you need your BCDR capacity. While there are usually spare hosts available that can be used to expand your secondary site, depending on your RTO (Recovery Point Objective) and growth requirement, you may not be able to reach your target number right away. The only way to guarantee the number of hosts you need right away is to reserve them ahead of time, but the tradeoff is the high cost. There are things you can do to optimize your availability and cost:

- Segment end-user population into tiers in terms of RTO. Some user segments may require a secondary desktop right away. You should have desktops created and on standby for them. Other user segments may be able to tolerate longer RTO and may require a secondary desktop within hours. In this case, you can wait for new hosts and desktops created. Each new host takes about 10 min to create, assuming the data center has available physical server.

- New hosts in the same cluster are created serially whereas hosts in different clusters are created in parallel. For faster host availability, it is better to have more clusters. Note: the current cluster limit recommended by VMware Cloud on AWS is 16 hosts per cluster.

Work with your VMware sales representative to ensure that you will have adequate BCDR capacity when you need it.

Below is an example of how you can set up and configure a BCDR site on VMware Cloud on AWS to protect your primary site. This works similarly regardless of whether the primary site is on-premises or on VMware Cloud on AWS.

In this example, our customer has a 1500 user VDI pod / site on-premise and you want to set up a secondary pod/site on VMware Cloud on AWS for the purpose of BCDR. They have determined that they will need 16 hosts on VMware Cloud on AWS for when the entire 1500 users are all using the secondary site.

They first create a secondary pod on VMware Cloud on AWS with 3 hosts (current minimum number of hosts per SDDC) and pay the reserve instance price. On the 3 hosts, they deploy 2 connection servers, 2 Unified Access Gateways, AD Domain Controller, and an event database. They also store a copy of their golden images in the secondary pod.

Initialize CPA between their primary pod and secondary pod. Put primary pod in site 1, and secondary pod in site 2.

They have also segmented their users into 3 tiers by their RTO:

- Tier 1 users: these users are essential personnel and need a secondary desktop right away when the primary pod/site goes down. There are 100 of them

- Tier 2 users: these users will require a secondary desktop within about 2 hours after the primary pod/site goes down. There are 600 of them.

- Tier 3 users: these users can wait up to 1 day before getting their secondary desktops. There are 800 of them.

During normal operations, they have 3 pools on the primary pod/site, one for each tier of users, with names of primary_pool_tier1, primary_pool_tier2, primary_pool_tier3. On the secondary pod/site they also create 3 pools, one for each tier of users, with names of secondary_pool_tier1, secondary_pool_tier2, and secondary_pool_tier3. Secondary_pool_tier1 is created with 100 VMs. Secondary_pool_tier2 and secondary_pool_tier3 are created with 1 VM each.

Then create 3 global entitlements:

- GE1 consists of primary_pool_tier1 and secondary_pool_tier1, and all of the Tier 1 users

- GE2 consists of primary_pool_tier2 and secondary_pool_tier2, and all of the Tier 2 users

- GE2 consists of primary_pool_tier3 and secondary_pool_tier3, and all of the Tier 3 users

When they experience a site-wide outage of the primary site, all users will be automatically logged off. As they try to log back in, the administrator has configured a pre-authentication message to inform them when each tier of users should expect to be able to get a desktop. This prevents users from repeatedly try to log into the secondary site that does not yet have a desktop ready for them. This message must be configured at the pod-level (rather than globally) at this time.

As instructed, the Tier 1 users will try to log back into their desktops right away. Since there's already a pool of 100 VMs ready for these users, they will be transparently connected to their secondary desktops.

In order to accommodate the 600 Tier 2 users, the administrator will first have to create 7 hosts, which will take roughly 70 minutes since hosts in the same cluster are created serially. Once the hosts have been created, the administrator can then expand secondary_pool_tier2 from 1 VM to 700 VMs. And since these are instant clones, the pool expansion would take 10-15 min or so. The creation process is manual for now, as the automated creation of new hosts based on increased demand is not yet available. At the pre-specified time, Tier 2 users will start logging into their desktops and be transparently connected to a desktop on the secondary site.

Once the Tier 2 user desktops have been provisioned, the administrator can repeat the same process and move onto Tier 3 users.

Once the primary site is back online again, users will be automatically connected to their primary desktop the next time they log in. The administrator can simply delete the secondary desktops on the secondary site, and then delete the unused hosts on the secondary site.

Note that the workflow above currently only works with global entitlements involving 2 sites, a primary site and a secondary site. If you have a scenario where you want to use the same DR site for two different primary sites, you still need to create two separate set global entitlements, one set for primary site 1 and secondary site, and another for primary site 2 and secondary site.

You can optionally configure a Global Load Balancer (GSLB) between the two sites and your end-users (such as F5, AWS Route 53, or others). The global load balancer provides a single-namespace capability that allows the use of a common global namespace when referring to CPA. Using CPA with a global load balancer provides your end users with a single connection method and desktop icon in their Horizon Client or Workspace ONE console. Without the global load balancer and the ability to have a single namespace for multiple environments, end-users will be presented with two different icons (corresponding to the number of pods on which desktops have been provisioned for them), which may potentially get confusing.

Business Continuity (BC) and Disaster Recover (DR) for Horizon 7 Full Clone Desktops

The BCDR workflow recommended in the previous section works well for non-persistent instant clones. There are some considerations for protection persistent full clone desktops.

First, do your users require the mirror image desktops after a primary site failure? If the answer is yes, then you'll need to replicate your primary full clone desktops periodically to the secondary site. This is the most costly type of protection – for every primary full clone desktop, you'll need an equivalent secondary full clone desktop on VMware Cloud on VMC, running at all times. You'll also need to script the import of secondary desktops into the connection servers on the secondary site as a manual full clone pool.

Most customers find that, given the cost of providing a fully mirrored desktop, it is acceptable to give their persistent full clone desktop users a secondary desktop that is a pristine copy of same golden image. Any user customization or data not saved in a file share and replicated to the secondary site will be lost, so you'll need to ensure that all important user data reside in a file share. You can then use the sample

workflow above to provision either an instant clone desktop or a full clone desktop on the secondary site for BCDR purpose.

Configuring VMware Cloud on AWS for Horizon 7 Deployment

The recommendation for a production environment is to use a minimum of three hosts in a cluster. Using a single host is recommended only for testing because with a single host, there is no HA. By default, a single-node SDDC gets deleted after 30 days.

To deploy Horizon 7 on VMware Cloud on AWS:

1. Create an SDDC instance on VMware Cloud on AWS. See the VMware Cloud on AWS documentation.
2. Deploy Horizon 7.5 or later on VMware Cloud on AWS. See the Horizon 7 documentation.
3. Set up the Horizon 7 environment on VMware Cloud on AWS.

Horizon 7 Environment on VMware Cloud on AWS

When you set up the Horizon 7 environment on VMware Cloud on AWS, you must install and configure the following components:

- Install Active Directory, DNS, DHCP, and KMS servers.
- Optionally, install RDS license servers.
- Install Horizon Connection Server and replica server version 7.5 or later.
- Use cloudadmin@vmc.local for the vCenter Server credentials and select VMware Cloud on AWS when adding the vCenter Server to Horizon.

For a single-node cluster, modify the vSAN VM storage policy to "No data redundancy."

Deploy a Unified Access Gateway appliance and connect it to the Connection Server if your deployment supports remote users.

- Use Unified Access Gateway version 3.3.
- Only deploy a single NIC with the OVF Deploy wizard. For multiple NICs, use the PowerShell script to include the password and encode special characters in the .INI configuration file. For more information, see the Unified Access Gateway documentation.
- Deploy the NICs to the Compute-ResourcePool, WorkloadDatastore, and Workloads folder.
- Specify netmask0-2 for the NICs.
- Deploy a load balancer if you are using two or more Connection Servers.

- Optionally, install a Horizon event database on Microsoft SQL Server 2016.

Install Horizon Agent on the master images for RDS hosts and VDI virtual desktop VMs. This agent communicates with the Connection Servers.

Deploy Horizon 7 over Hybrid Cloud

You might already have Horizon 7 environments on-premises. The Horizon 7 pod on-premises and your Horizon 7 pod on VMware Cloud on AWS can be managed separately. Alternatively, you can extend your on-premises Horizon 7 environment to the cloud by linking it with your Horizon 7 on VMware Cloud on AWS environment using Cloud Pod Architecture (CPA). Deploying your Horizon 7 over hybrid cloud enables you to manage your on-premises deployment and your cloud deployment in a single federated space.

For hybrid cloud deployment, follow these steps.

1. Configure VPN and firewall rules to enable the Connection Server instance on VMware Cloud on AWS to communicate with the Connection Server instance on-premises.

2. Prepare Microsoft Active Directory (AD) and choose to set up a one-way trust or a two-way trust.

3. Ensure that your on-premises Horizon 7 version is 7.0 or later.
   Note: The Horizon 7 version deployed on-premises does not have to match the Horizon 7 version deployed on VMware Cloud on AWS. However, you cannot mix a Horizon 6 pod (or lower) with a Horizon 7 pod within the same CPA.

4. Use Cloud Pod Architecture to connect the Horizon 7 pod on-premises with the Horizon 7 pod on VMware Cloud on AWS.

5. For easy sharing of images and ISO images, you can use the vCenter Content Library on each vCenter Server.

Connection and Firewall Configuration for Deploying Horizon 7 on VMware Cloud on AWS

To set up a successful hybrid cloud deployment, you must follow these connection and firewall rules.

Connection Rules

Use the VMC Console in VMware Cloud on AWS to create a VPN in the SDDC management network to connect to your on-premises management network. Next, configure the management gateway with firewall rules. Finally, specify DNS server addresses for the management network. Your networking team can configure the on-premises VPN using information you download from the SDDC.

You must configure the following VPN connections between components in the logical network:

- From the management component to the on-premises component
- From the compute component to the on-premises component
- From the compute component to the management component

You can also use AWS Direct Connect to set up a connection between Horizon 7 and VMware Cloud on AWS. For more information on configuring VPNs or using AWS Direct Connect, see the VMware Cloud on AWS Getting Started document.

Firewall Rules

You can run the Firewall Rule Accelerator in VMware Cloud on AWS for all VPNs to create all the required firewall rules.

The following table describes firewall rules for the Management Gateway on VMware Cloud on AWS.

Table 2. Management Gateway Firewall Rules

Rule Name	Service Name	Ports	Action	Source	Destination
Any SSO	SSO (TCP 7444)	7444	Allow	Any	vCenter
vCenter (ANY) to Management-On-Prem	Any (All Traffic)	Any	Allow	vCenter	Compute/On-prem subnet
ESXi (ANY) to Management-On-Prem	Any (All Traffic)	Any	Allow	ESXi	Compute/On-prem subnet
Management-On-Prem to vCenter (HTTPS)	HTTPS (TCP 443)	443	Allow	Compute/On-prem subnet	vCenter
Management-On-Prem to vCenter (ICMP)	ICMP (All ICMP)	Any	Allow	Compute/On-prem subnet	vCenter
Management-On-Prem to ESXi (Provisioning)	Provisioning (TCP 902)	902	Allow	Compute/On-prem subnet	ESXi
Management-On-Prem to ESXi (Remote Console)	Remote Console (TCP 903)	903	Allow	Compute/On-prem subnet	ESXi
Management-On-Prem to ESXi (ICMP)	ICMP (All ICMP)	Any	Allow	Compute/On-prem subnet	ESXi
Default Deny All	Any (All Traffic)	Any	Deny	Any	Any

The following table describes firewall rules for the Compute Gateway needed to install on VMware Cloud on AWS.

Table 3. Compute Gateway Firewall Rules

Rule Name	Service Name	Ports	Action	Source	Destination
Compute (ANY) to Internet and VPN	Any (All Traffic)	Any	Allow	Any	All Internet and VPN
Management-On-Prem (ANY) to BackEnd	Any (All Traffic)	Any	Allow	On-Premises Management subnet	Management Subnet

For stricter control and external access, see the *Ports and Services* chapter in the VMware Horizon 7 Security guide.

Preparing Active Directory for Hybrid Cloud Deployment

If you are deploying Horizon 7 in a hybrid cloud environment by linking the on-premises pod with the VMware Cloud on AWS pod, you must prepare the on-premises Microsoft Active Directory (AD) to access the AD on VMware Cloud on AWS.

If you are deploying the Horizon 7 pod on VMware Cloud on AWS as a standalone (that is, not part of a hybrid cloud deployment), you can skip the preparation of the on-premises AD.

Use the following guidelines to prepare AD for your hybrid cloud deployment if you want the on-premises AD domain controllers to service the Horizon 7 pod on VMware Cloud on AWS.

Note: The access time might be slow due to the latency between the on-premises pod and VMware Cloud on AWS.

On VMware Cloud on AWS, deploy a read-only AD domain controller.

- Configure a trust from the domain running on VMware Cloud on AWS to your existing domain. When you allow the domain running on VMware Cloud on AWS to access the on-premises domain, the domain running on VMware Cloud on AWS can serve as a resource domain. Configuring a trust enables your users to sign in with single sign-on (SSO), using their existing corporate credentials, to services running within VMware Cloud on AWS.

Link Horizon 7 Pods on VMware Cloud on AWS

You can use the Cloud Pod Architecture feature to connect Horizon 7 pods regardless of whether the pods are on-premises or on VMware Cloud on AWS. When you deploy two or more Horizon 7 pods on VMware Cloud on AWS, you can manage them independently or manage them together by linking them with Cloud Pod Architecture.

- On one Connection Server, initialize Cloud Pod Architecture and join the Connection Server to a pod federation.
- Once initialized, you can create a global entitlement across your Horizon 7 pods on-premises and on VMware Cloud on AWS.

- Optionally, when you use Cloud Pod Architecture, you can deploy a global load balancer (such as F5, AWS Route 53, or others) between the pods. The global load balancer provides a single-namespace capability that allows the use of a common global namespace when referring to Horizon CPA. Using CPA with a global load balancer provides your end users with a single connection method and desktop icon in their Horizon Client or Workspace ONE console.

Without the global load balancer and the ability to have a single namespace for multiple environments, end users will be presented with a possibly confusing array of desktop icons (corresponding to the number of pods on which desktops have been provisioned for them). For more information on how to set up Cloud Pod Architecture, see the *Administering Cloud Pod Architecture in Horizon 7* document.

Use Cloud Pod Architecture to link any number of Horizon 7 pods on VMware Cloud on AWS. The maximum number of pods must conform to the limits set for pods in Cloud Pod Architecture. See, the VMware Knowledge Base article VMware Horizon 7 Sizing Limits and Recommendations (2150348).

When you connect multiple Horizon 7 pods together with Cloud Pod Architecture, the Horizon 7 versions for each of the pods can be different from one another. The only limitation is that they all be Horizon 7 v7.0 or higher (i.e. no mixing of Horizon 6 pods).

Shared Content Library

Content Libraries are container objects for VM, vApp, and OVF templates and other types of files, such as templates, ISO images, text files, and so on. vSphere administrators can use the templates in the library to deploy virtual machines and vApps in the vSphere inventory. Sharing golden images across multiple vCenter Server instances, between multiple VMware Cloud on AWS and/or on-premises SDDCs guarantees consistency, compliance, efficiency, and automation in deploying workloads at scale.

For more information, see Using Content Libraries in the *vSphere Virtual Machine Administration* guide in the VMware vSphere documentation.

Licensing

Enabling Horizon 7 to run on VMware Cloud on AWS requires two separate licenses: capacity license for VMware Cloud on AWS and Horizon Subscription License.

For POC or pilot deployment of Horizon 7 on VMware Cloud on AWS, you may use a temporary eval license or your existing perpetual license. However, to enable Horizon 7 for production deployment on VMware Cloud on AWS, you must purchase the new Horizon Subscription License. To obtain the new Horizon Subscription License or for more information on how to upgrade your existing perpetual license to subscription license and associated discounts, please contact your VMware representative.

Different Types of Horizon Subscription Licenses

Horizon Subscription Licenses come in five major flavors:

- Horizon Apps Subscription –deploying RDSH apps only. A single license can be used for deploying both on-premises or on VMware Cloud on AWS. An on-premise vSphere license is included.

- Horizon Apps Subscription Add-on – deploying RDSH apps only. A single license can be used for deploying on VMware Cloud on AWS only. No on-premise vSphere license is included and therefore this license is lower cost.

- Horizon Subscription – deploying RDSH apps and VDI. A single license can be used for deploying both on-premises or on VMware Cloud on AWS. An on-premise vSphere license is included.

- Horizon Subscription Add-on – deploying RDSH apps and VDI. A single license can be used for deploying on VMware Cloud on AWS only. No on-premise vSphere license is included and therefore this license is lower cost.

- Workspace One Subscription – deploying Horizon RDSH apps and VDI, as well as Workspace One mobility solution.

Except for Workspace One, all other licenses above have Concurrent User and Named User options.

You can use different licenses (including perpetual licenses) on different Horizon pods whether the pods are connected by CPA or not. You cannot mix different licenses within a pod since each pod only takes 1 type of license. For example, you cannot use both perpetual license and subscription license for a single pod. You also cannot use both the Horizon Apps Subscription license and the Horizon Subscription license for a single pod. Suppose you have two pods deployed, pod A on-premises and pod B on VMware Cloud on AWS and the two pods are connected by CPA, you can use a different license type on each pod. For example, you can use the Horizon Enterprise perpetual license for pod A, and the new Horizon Subscription license for pod B.

The best subscription license you need for your Horizon 7 on VMware Cloud on AWS deployment will depend on your use case. Here are some examples:

- You are setting up a new H7 deployment for 2,000 VDI users on VMware Cloud on AWS. There are no on-premises components. Purchase 2,000-user Horizon Subscription license in this case.

- You have an existing Horizon 7 pod on-premises for 2,000 users, and you want to deploy a pod on VMware Cloud on AWS for an addition 1,000 users for full time VDI use. The best license type is the Horizon Subscription Add-on for your Horizon 7 pod on VMware Cloud on AWS. You would keep your perpetual license for on-premise pod until renewal and then decided whether to move to Horizon Subscription license for your on-prem pod.

- You have an existing Horizon 7 pod on-premises for 2,000 users, and you want to deploy a pod on VMware Cloud on AWS as BCDR capacity for the 2,000 users on-premise. The best license type is to upgrade your existing 2,000-user perpetual license to 2,000-user Horizon Subscription license. This new license would allow these 2,000 users to connect to virtual desktops either on-premises or on VMware Cloud on AWS. If you still have significant time left on your perpetual license, another option is to keep the perpetual license and purchase 2,000-user Horizon Subscription Add-on license – your perpetual license includes a vSphere on-premise license. Then upgrade to the 2,000-user Horizon Subscription license when your perpetual license is up for renewal.

License Enablement

Regardless of whether you are deploying Horizon 7 on-premises or on VMware Cloud on AWS, if you are using any of the subscription licenses, you must install the Horizon Cloud Connector to enable subscription license management for Horizon 7. The Horizon 7 Cloud Connector is a virtual appliance that connects a Horizon 7 pod with Horizon Cloud Service features.

A MyVMware account from https://my.vmware.com is required for Horiozn 7 subscription license. Once you purchase the subscription license, a record will be created in the Horizon Cloud Service using your MyVMware email address, and your subscription license information will be visible to the Horizon Administrator console.

As part of the subscription license fulfillment process, you will receive email with the link to download the Horizon 7 Cloud Connector as an OVA file and follow instructions to deploy the Cloud Connector, from vSphere web client, alongside your new or existing Horizon 7 pods. Once the Cloud Connector is deployed and paired with the Connection Server in the Horizon 7 pod with the Horizon Cloud Service, which manages the Horizon 7 subscription license between connected Horizon 7 pod(s). Unlike the Horizon 7 perpetual license, with a subscription license, you do not need to retrieve or manually enter a license key for Horizon 7 product activation. However, supporting component license keys, e.g., license key for vSphere, license key for App Volumes and others, will be delivered separately and must be manually keyed in to activate the product by the administrator.

Review the Horizon 7 documentation for more details on how to deploy the Horizon 7 Cloud Connector Virtual Appliance. You will need a separate Cloud Connector for each pod.

Currently, Horizon Cloud Service is only used to enforce subscription licenses for Horizon 7. Over time, additional features will be available on Horizon Cloud Service for Horizon 7 deployments.

Deploying Desktops on VMware Cloud on AWS with Instant Clone, App Volumes, and User Environment Manager

Instant Clone

In addition to using Full Clones, you can also leverage Instant Clone Technology (starting with Horizon 7.7) coupled with App Volumes (starting with App Volumes 2.15) to accelerate the delivery of user-customized and fully personalized desktops. Dramatically reduce infrastructure requirements while enhancing security by delivering a brand-new personalized desktop and application services to end users every time they log in:

- Reap the economic benefits of stateless, nonpersistent virtual desktops served up to date upon each login.
- Deliver a pristine, high-performance personalized desktop every time a user logs in.
- Improve security by destroying desktops every time a user logs out.

When you install and configure Horizon 7 for instant clone for deployment on VMware Cloud on AWS, do the following:

When adding VMware Cloud on AWS vCenter Server to the Horizon configuration, be sure to select the **VMware Cloud on AWS** check box.

- CBRC is not supported or needed on VMware Cloud on AWS. CBRC has been disabled by default.
- On the master image, add the domain's DNS to avoid customization failures.
- When creating Horizon instant-clone pools on VMware Cloud on AWS, use the following settings in the provisioning wizard:
    - **Compute-ResourcePool** resource pool
    - **Workloads** folder
    - **WorkloadDatastore** datastore

Multi-VLAN is not yet supported when creating Horizon instant-clone pools on VMware Cloud.

App Volumes

App Volumes provides real-time application delivery and management, now for on-premise and on VMC:

- Quickly provision applications at scale.
- Dynamically attach applications to users, groups, or devices, even when users are already logged in to their desktop.
- Provision, deliver, update, and retire applications in real time.
- Provide a user-writable volume, allowing users to install applications that follow them across desktops.
- Provide end users with quick access to a Windows workspace and applications, with a personalized and consistent experience across devices and locations.
- Simplify end-user profile management by providing organizations with a single and scalable solution that leverages the existing infrastructure.
- Speed up the login process by applying configuration and environment settings in an asynchronous process instead of all at login.
- Provide a dynamic environment configuration, such as drive or printer mappings, when a user launches an application.

For more information on how to configure, see "Configuring App Volumes Manager for VMware Cloud on AWS" in the App Volumes Administration guide for App Volumes.

Transfer App Volumes from vSphere to VMC

For migration or BCDR purpose, you can transfer your appstacks or user writable volumes from on-premise to the VMware Cloud on AWS environment using your vSphere client in a two-step process.

From the vSphere client:

1. Create a VM with thin provisioning and attach the volume that you want to transfer to the VM.
2. Select the VM and export it as an OVF template from File > Export to OVF Template.

From the VMware Cloud on AWS web client:

1. Click Actions > Deploy OVF Template.
2. Follow on-screen instructions and when you have to select the storage format, select Thin provision.

Once the VM is created, browse the datastore where the OVF was exported and move the VMDK file with its metadata to the cloudvolumes directory.

Ensure that you change the template location in the metadata file to point to the new datastore.

User Environment Manager

Use VMware User Environment Manager for application personalization and dynamic policy configuration across any virtual, physical, and cloud-based environment. Install and configure the User Environment Manager on VMware Cloud on AWS just like installing on-premises.

Estimating Data Egress Cost

Unlike on-premises, deploying Horizon 7 on VMware Cloud on AWS incurs data egress cost based on the amount of data egress traffic your environment will generate. It is important to understand and estimate the data egress traffic.

Understanding Different Types Data Egress Traffic

Depending on your deployment use case, you may be incurring cost for some or all of the following types of data egress traffic:

- End-user traffic via Internet - You have configured your environment where your end users will connect to their virtual desktops on VMware Cloud on AWS remotely via the Internet. Any data leaving the VMware Cloud on AWS data center will incur egress charge. Egress data consists of the following components: outbound data from Horizon 7 protocols and outbound data from remote experience features (for example, remote printing). While the former is typically predicable, the latter has more variance and depends on the exact activity of the user

- End-user traffic via on-premises - You have configured your environment where your end users will connect to their virtual desktops on VMware Cloud on AWS via your on-premise data center. In this case, you will have to link your data center with the VMware Cloud on AWS data center using VPN or Direct Connect. Any data traffic leaving the VMware Cloud on AWS data center back to your data center will incur egress charge. And if you have Cloud Pod Architecture (CPA) configured between the on-premise environment and the VMware Cloud on AWS environment, you will incur egress charge for any CPA traffic between the two pods (although CPA traffic is typically fairly light)

- External application traffic - You have configured your environment where your virtual desktops on VMware Cloud on AWS has to access applications hosted either in your on-premise environment or in another cloud. Any data traffic leaving the VMware Cloud on AWS data center to these other data centers will incur egress charge

- Note that data traffic within your VMware Cloud on AWS org or between the org and AWS services in that same region is exempt from egress charge. However, any traffic from the org to another availability zone or to another AWS region will be subject to egress charge

Data ingress (i.e. data flowing into VMware Cloud on AWS data center) is free of charge.

Estimating Data Egress Traffic with SysTrack from Lakeside

Since the data egress cost is priced per GB, the best way to estimate your data egress cost is to estimate your likely data egress traffic by using a monitoring tool in your existing on-premises environment (whether it's already virtualized or not). Make sure you estimate the different types of data egress traffic listed above separately as applicable. One such monitoring tool is SysTrack from Lakeside Software.

Lakeside's SysTrack workplace analytics solution contains an extensive set of tools to provide relevant planning information for desktop transformation. Best of all, this is available at no cost to VMware customers via the SysTrack Desktop Assessment (SDA). Through the SDA, customers can collect detailed environmental information, including recommendations for deployment options and resource requirements, with only the need to deploy the SysTrack agent to systems being considered for transformation. For advanced cases, the on-premises version of SysTrack can be used as well. This guide will make the assumption that such a deployment is already in place. For additional setup details or questions about the SDA, the Quick Start Guide is a good resource.

Once SysTrack is deployed in the environment, it will immediately begin collecting relevant information from devices on which it's installed. For this guide we'll focus on the most interesting facets of data collection for the network:

- Per Device Network Usage
- Per Session Protocol Bandwidth Usage
- Per Application (and Destination) Bandwidth Usage

The key is thinking about how best to combine these pieces of information into something that's useful for planning costs. Because the ingress data is free on VMware on AWS, we'll focus on the egress data as observed from the devices in the collection. That will take the form of "transmitted" data from that device to other destinations, and you'll see more details about this as we go into methodology we suggest.

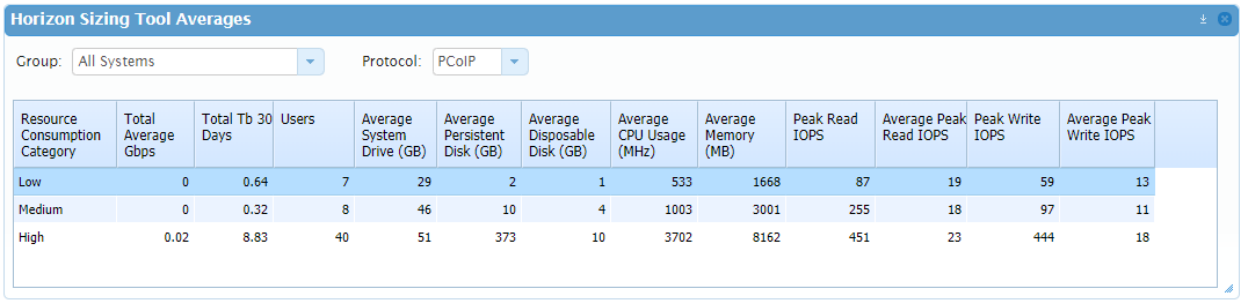
With SysTrack you can make use of two different styles of calculation depending on the level of detail you'd like to see. We've created several dashboards that customers can use to easily visualize this information in SysTrack.

Note – All figures in these dashboards are measured in bits, not bytes.

1. Basic Network Egress Bandwidth Calculation

   The Horizon Sizing Tool dashboard provides some basic numbers to help you plan your migration to the VMware cloud. The table below breaks users down into categories based on egress bandwidth consumption. Resource consumption metrics are supplied for each category as well as egress bandwidth consumption for applications and three remote display protocols: ICA, Blast and PCoIP.

   The resource consumption metrics can be fed into some of VMware's sizing tools (Horizon Sizing Tool & Horizon Sizing Estimator). These tools provide guidance on how to plan for the number and size of systems you will be deploying.



Resource Consumption Category	Total Average Gbps	Total Tb 30 Days	Users	Average System Drive (GB)	Average Persistent Disk (GB)	Average Disposable Disk (GB)	Average CPU Usage (MHz)	Average Memory (MB)	Peak Read IOPS	Average Peak Read IOPS	Peak Write IOPS	Average Peak Write IOPS
Low	0	0.64	7	29	2	1	533	1668	87	19	59	13
Medium	0	0.32	8	46	10	4	1003	3001	255	18	97	11
High	0.02	8.83	40	51	373	10	3702	8162	451	23	444	18

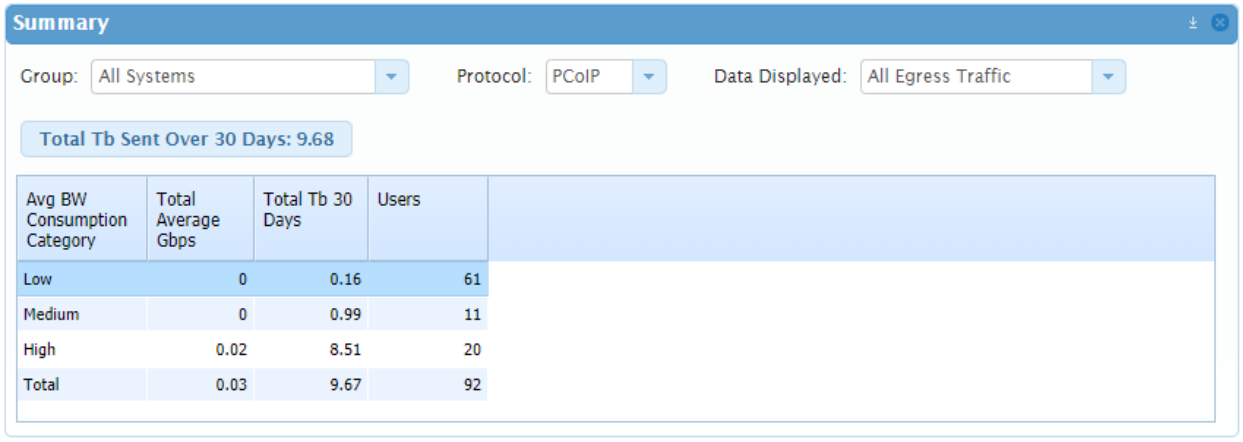
2. Advanced Network Egress Bandwidth Calculation

   The Advanced Horizon Sizing Tool offers advanced sizing calculations which can be used to estimate costs around migrating to the VMware cloud. VMware cloud pricing is based around the level of data transmitted from the cloud back to the client (i.e. egress bandwidth consumption).

Within the dashboard, each user is put into a category based on egress bandwidth consumption. This can be based on actual SysTrack data if you are migrating from an existing virtual environment or estimated if migrating from a physical environment.

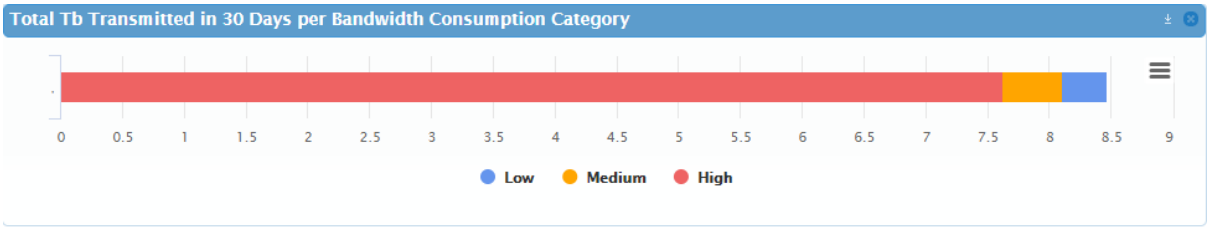
You can select the protocol you would like to use to estimate values for users on physical machines, as well as the type of egress data displayed – protocol, application or a combination of both.

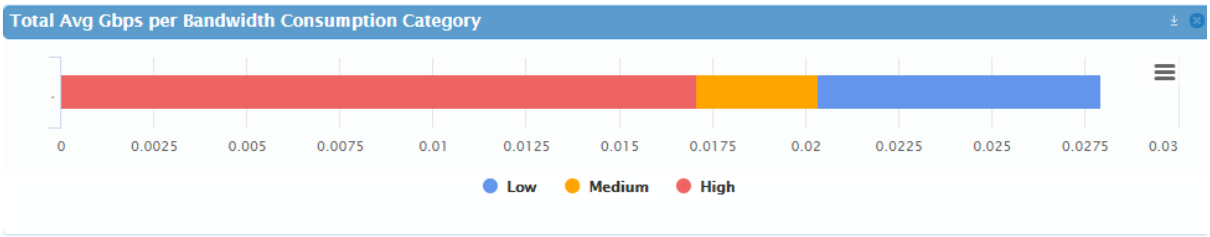
A summary table shows these categories and tells you the total and average for egress bandwidth consumption.



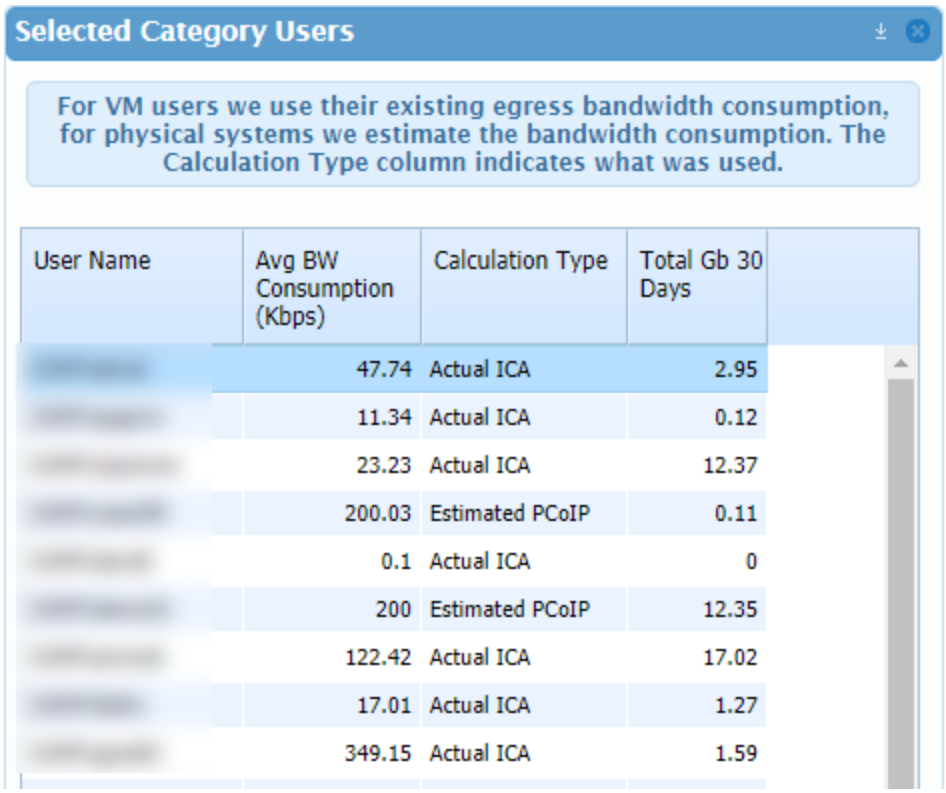
Avg BW Consumption Category	Total Average Gbps	Total Tb 30 Days	Users
Low	0	0.16	61
Medium	0	0.99	11
High	0.02	8.51	20
Total	0.03	9.67	92

This same data is also presented in graph form for a more visual presentation:





You can also see more detail on a per user basis depending on the category selected in the summary table. The table below shows total and average bandwidth consumption as well as calculation type, which indicates which data source was used to generate the consumption values.



Here is an example of how to determine monthly egress bandwidth per VM in gigabytes.

If you want to simply calculate the egress bandwidth for an average VM, you can take the “Total Tb 30 Days” and “Users” values in the advanced dashboard summary table and perform the following calculation:

Monthly GB egress bandwidth per VM = (([Total Tb 30 Days]/8) * 1024) / [Users]

From the data in the screenshot below, the calculation would become this:

Monthly GB egress bandwidth per VM = ((9.67/8) * 1024) / 92 = 13.5 GB