# Best Practices for VPCs and Networking in Amazon WorkSpaces Deployments

*July 2020*

# Notices

Customers are responsible for making their own independent assessment of the information in this document. This document: (a) is for informational purposes only, (b) represents current AWS product offerings and practices, which are subject to change without notice, and (c) does not create any commitments or assurances from AWS and its affiliates, suppliers or licensors. AWS products or services are provided “as is” without warranties, representations, or conditions of any kind, whether express or implied. The responsibilities and liabilities of AWS to its customers are controlled by AWS agreements, and this document is not part of, nor does it modify, any agreement between AWS and its customers.

© 2020 Amazon Web Services, Inc. or its affiliates. All rights reserved.

# Contents

# Abstract

Today, many customers want to expand or migrate their desktop infrastructure environment onto AWS. This paper outlines best practices for implementing a virtual desktop environment using Amazon WorkSpaces. It offers guidance around factors affecting the AWS networking components that must be considered when deploying WorkSpaces.

# Introduction

Amazon WorkSpaces is a managed, secure cloud desktop service. Proper network configuration is essential to the successful implementation and ongoing operations of a WorkSpaces environment. As an End User Computing (EUC) service, any variation to the end user's experience when interacting with a WorkSpaces instance is very visible and can result in a loss of workforce productivity—especially if network connectivity has not been designed in accordance with best practices.

Networking for cloud-delivered desktops, such as Amazon WorkSpaces, should consider many aspects to help ensure that every user enjoys a consistent end user experience that enables them to be productive regardless of how and where they connect from. Considering all the factors that can have a negative influence on a user's WorkSpaces instance is essential when designing a WorkSpaces environment. While WorkSpaces reduces the number of design decisions that need to be considered compared to a traditional Virtual Desktop Infrastructure (VDI) environment, it is still a customer's responsibility to ensure that the required networking and access to supporting services and applications is available to support ongoing successful operation of their WorkSpaces environment.

Each WorkSpaces instance, while providing access to a single user's application portfolio, relies on a number of infrastructure and application services. The inaccessibility of any of these services could impact the end user in a number of ways including:

- Initial connectivity through the unavailability of authentication infrastructure or network connectivity
- The inability to patch and update the operating system and associated applications
- The failure to connect to application servers, ability to connect to internet-hosted applications, etc.

This document describes the fundamental capabilities of the AWS networking portfolio that can be used to deploy an Amazon WorkSpaces environment and explains how these can be used to tailor the environment to different use cases.

If you are familiar with AWS networking components, then the table in

Appendix A can be used to quickly find recommendations and considerations for each networking component that can be deployed within a WorkSpaces deployment.

# Networking for Amazon WorkSpaces

When designing the AWS networking that underpins a WorkSpaces environment, it is difficult to predict the future use of the environment, as the current users in scope for the solution are likely to have changing requirements over time. In addition, future users of the environment with different requirements also need to be considered. Therefore, it is a best practice to size and future proof the WorkSpaces environment as much as possible while balancing short term needs with future use cases but minimizing any potential for waste as well.

The first considerations for networking are the Region and its associated Availability Zones. The Physical Network Architecture section of this document covers this topic. After these fundamental decisions to identify the physical location of the WorkSpaces environment have been made, the next set of decisions are related to the logical network infrastructure that will support the WorkSpaces. The Logical Network Architecture section of this document covers these aspects including VPC (Virtual Private Cloud) design aspects.

Before we consider both the physical and logical aspects, it is necessary to understand the networking requirements and network traffic flows of the WorkSpaces service.

## WorkSpaces instance networking

Each WorkSpaces instance is associated with a specific VPC and the AWS Directory Service construct you used to create it. The WorkSpaces service requires a minimum of two subnets to operate, each in a different Availability Zone (AZ). This approach ensures distribution of your WorkSpaces across AZs such that in the event that a single AZ becomes unreachable, WorkSpaces in the unaffected AZs are still reachable.

Figure 1 shows how a single WorkSpaces instance is connected to the networks required for its successful operation.

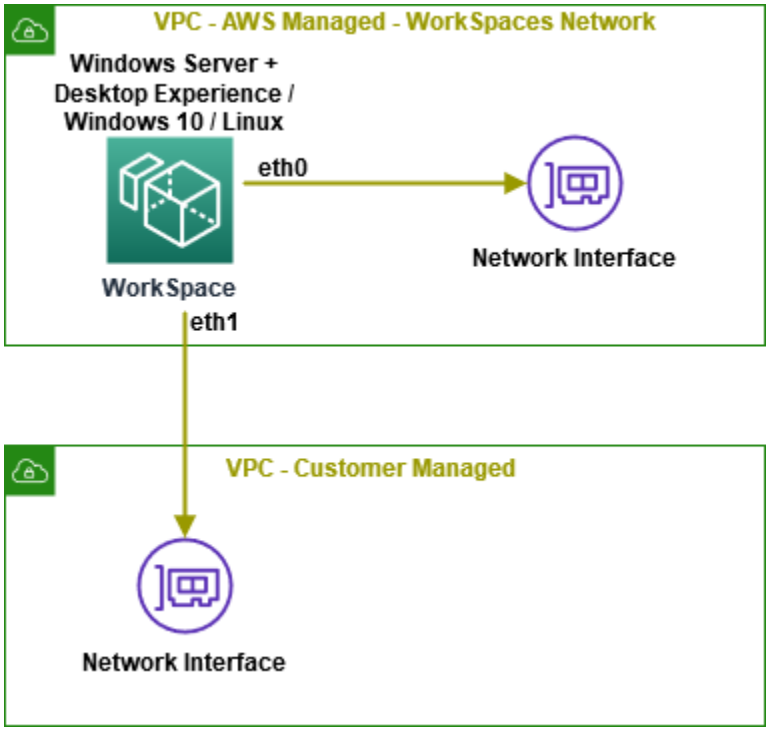


*Figure 1 – Networking for a single WorkSpace instance*

Each WorkSpace has two elastic network interfaces, a management network interface (eth0), and a primary network interface (eth1). The management network interface is where the network connection from your client endpoint access device terminates for streaming traffic—PCoIP or WorkSpaces Streaming Protocol (WSP)—and is also used by AWS for management of AWS provided software within the WorkSpace. For network routing to function correctly, you cannot use this private address space on any network that can communicate with your WorkSpaces network.

For a list of the private IP ranges that AWS uses on a per Region basis, see Management Interface IP Ranges.

> **Note**: Amazon WorkSpaces and their associated management network interfaces do not reside in your VPC, and you cannot view the management network interface or the Amazon Elastic Compute Cloud (Amazon EC2) instance ID in the AWS Management Console. However, you can view and modify the security group settings of the primary network interface (eth1) associated with each WorkSpace in the AWS Management Console. Also, the primary network interface of each WorkSpaces **does** count toward your elastic network interface Amazon EC2 resource quotas. For large deployments of WorkSpaces, you will need to open a support ticket via the AWS Management Console to increase your elastic network interface quotas.

The primary network interface of each WorkSpace (eth1) provides connectivity to resources inside your VPC, such as access to AWS Directory Service, the internet, and your corporate network. It is possible to attach security groups to this primary network interface (as you would do to any elastic network interface). AWS recommends the application of security groups to this elastic network interface based on the deployment (that is, the security context of the WorkSpace).

### Management network interface

You cannot control the management network interface via security groups. However, you can use a host-based firewall to block ports or control access to your WorkSpace. AWS does not recommend applying restrictions on the management network interface. If you decide to add host-based firewall rules to manage this interface, you must keep a few ports open so that the WorkSpaces service can manage the health and accessibility to the WorkSpace, as defined in the Amazon WorkSpaces Administration Guide in the WorkSpaces port requirements section.

It is also possible to use IP access control groups to restrict the client IP addresses that can connect to the Management Network Interface for streaming traffic as documented later in this document in the IP access control groups section here.

## Traffic flow

The traffic flow for Amazon WorkSpaces can be broken into two main components:

- The traffic between the client device and the Amazon WorkSpaces instance
- The traffic between the Amazon WorkSpaces instance and customer network

These components are considered in the next two sections.

## Client device to WorkSpace

The end-user device either running the Amazon WorkSpaces client or using Amazon WorkSpaces web access, regardless of its location (on-premises or remote), uses the same two ports for connectivity to the service. The client uses HTTPS/TCP over port 443 and port 4172/TCP+UDP (PCoIP/WSP) for communications and network health checks. Traffic on both ports is encrypted. Port 443 traffic is used for authentication and session information and uses TLS for encrypting the traffic. Pixel streaming traffic, keystrokes, and pointer movements are encrypted using up to AES-256-bit encryption for communication between the client and eth0 of the WorkSpace, via the streaming gateway.

We publish per-region IP ranges of our PCoIP/WSP streaming gateways and network health check endpoints. You can limit outbound traffic on port 4172 from your corporate network to the AWS streaming gateway and network health check endpoints by allowing only outbound traffic on port 4172 to the specific AWS Regions in which you're using WorkSpaces. By doing so, access to WorkSpaces can be restricted to a single Region. For the IP ranges and network health check endpoints, see Amazon WorkSpaces PCoIP Gateway IP Ranges.

## Amazon WorkSpaces to VPC

After a connection is authenticated from a client device to a WorkSpace and streaming traffic (PCoIP/WSP) is initiated, your WorkSpaces client will display a Windows or Linux desktop (your WorkSpace) that is connected to your VPC. Each WorkSpace's primary elastic network interface, identified as eth1, will have an IP address assigned to it by the Dynamic Host Configuration Protocol (DHCP) service that is provided by your VPC. The elastic network interface that is in your VPC has access to allowed resources in the VPC and to allowed networks that you have connected to your VPC (for example, via VPC peering, AWS Direct Connect connection, Site-to-Site VPN connection, etc.).

Elastic network interface access to your network resources is determined by the default security group (see more on security groups here) that your AWS Directory Service configures for each WorkSpace and any additional security groups that you assign to the elastic network interface. You can add security groups to the elastic network interface within your VPC at any time by leveraging the AWS Management Console or AWS CLI. Network access control lists or network ACLs (see more on network ACLs here) can also be used on a subnet to restrict network traffic. In addition to security groups, you can use your preferred host-based firewall on a given WorkSpace to limit network access to resources within the VPC.

# VPC design options

Now that we understand the network connectivity and traffic flow for each WorkSpace, let's consider some of the design options for designing the VPC into which you deploy the ENIs associated with your WorkSpaces instances.

There are a number of factors that influence the design of your VPC for Amazon WorkSpaces.

One of these includes the types of user (for example, consultants vs. permanent employees) that will be accessing the WorkSpaces instances and whether they need to be segregated for security requirements. This requirement can lead to designs influenced by risk.

Another factor is the location of a customer's existing infrastructure, data, applications, and users. The physical location of these leads to a design influenced by the placement of infrastructure, data, and users.

Lastly, where larger environments and customers are concerned, there are designs that are influenced by both risk and the placement of infrastructure, data, or users.

Each of these influences and some VPC designs that meet the requirements are considered in the following sections.

## VPC designs influenced by risk

The high-level design for a VPC being used to host a WorkSpaces environment leverages a VPC in an AWS Region with at least two Availability Zones (see here). Within each AZ at least one subnet is created to provide the eth1 network connectivity to each WorkSpace. This high-level design can satisfy the requirement to provide a simple WorkSpaces environment where the security and the associated risk profile of each user is the same.

There are a number of cases where there is a need to evolve this model to cater for multiple types of user that have different security requirements and risk profiles. In these instances, it is a best practice to group together users with a similar risk profile, but segregate these groups from each other. This might be due to concerns around data leakage, compliance requirements, potential malicious activities, a need to reduce risk for certain types of users, or to minimize the impact of a potential cyber-attack against a subset of WorkSpaces. An example of applying segregation to different types of users is shown in Figure 2.

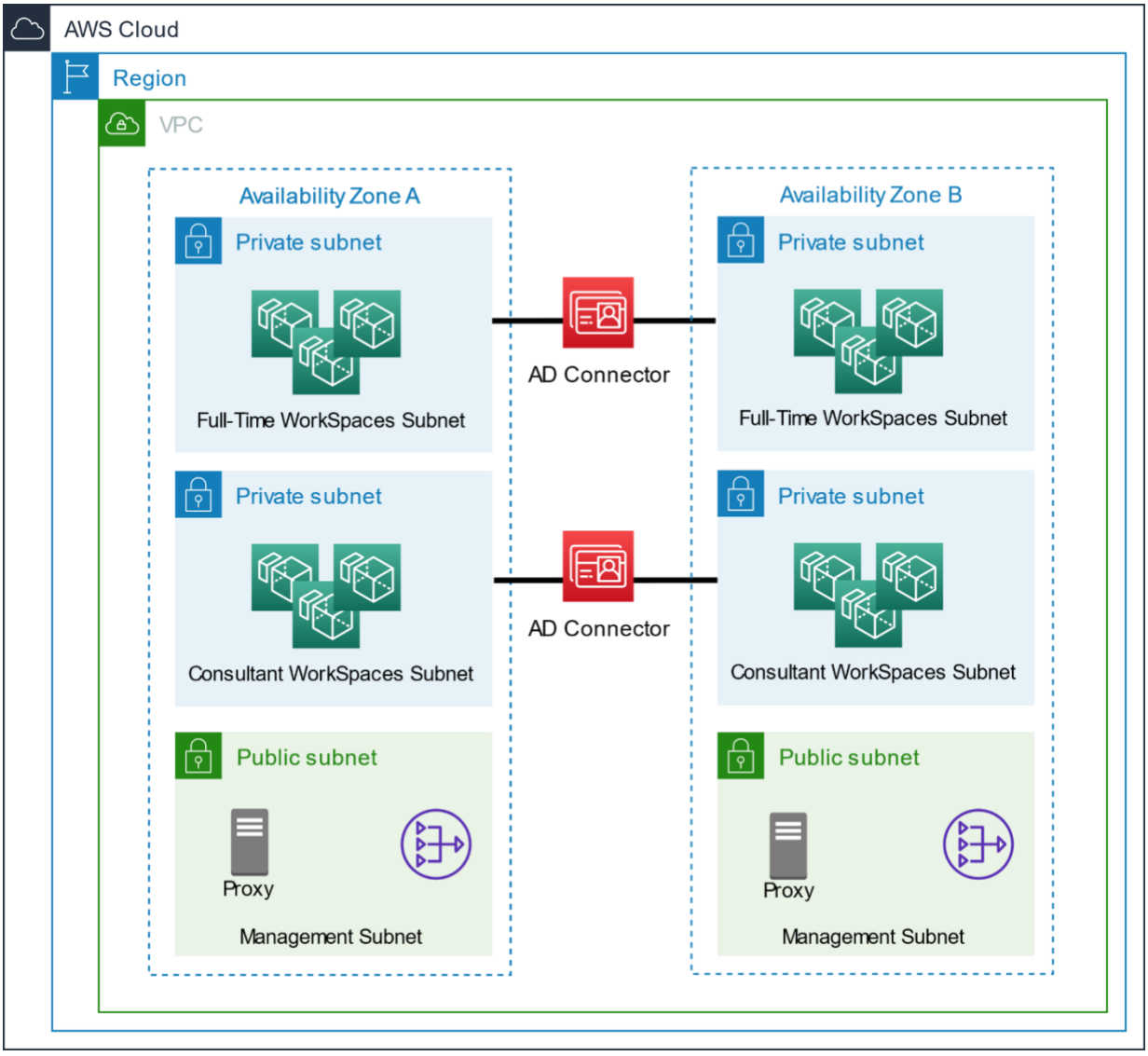


*Figure 2 – Full-Time vs Consultant WorkSpaces scenario*

In this example, two different types of users have been identified, full-time employees and consultants. To satisfy the requirement to segregate these two types of users from each other, two separate private subnets have been created in each AZ. One set of subnets host WorkSpaces for full-time staff, which are deemed to have a lower risk profile than the temporarily employed consultants, which are hosted on a separate set of subnets.

While in this example, segregation has been employed at a coarse-grained (high) level (that is, full-time vs. consultants), this same approach could be applied at a more

granular level to other types of user such as: BPO (Business Process Outsourcing) users, developers, support staff, or temporary staff. In this way, the different risk profiles for each set of users can be satisfied at a granular level within each subnet. Each subnet can be restricted in terms of network connectivity if desired, by implementing a distinct routing table that restricts the users to the areas of the network where they are permitted.

In addition, network access control lists (which are explored later in this document) can be used at the subnet level to restrict the network traffic at the port and IP address level. However, it is recommended that network access control lists (network ACLs) be used cautiously as they can have a significant impact on the operation of a WorkSpaces environment. It is important to note that network ACLs have no effect inside a subnet and only affect communication between subnets.

This approach can be expanded upon further by using security groups to impose a network security constraint at a more granular (low) level. This approach allows the network connection of each WorkSpace within a subnet to have a different security posture based on the risk profile associated with the user assigned to the individual WorkSpace. A general approach to consider is that security groups should be used for security controls and network ACLs should be used to manage any exceptions.

A security group can be applied on a per WorkSpace basis and therefore individual WorkSpaces within the same subnet can have different network security postures applied to them. The following example in Figure 3, shows different security groups applied to each of the two sets of subnets hosting WorkSpaces, in this case trusted and the untrusted users.

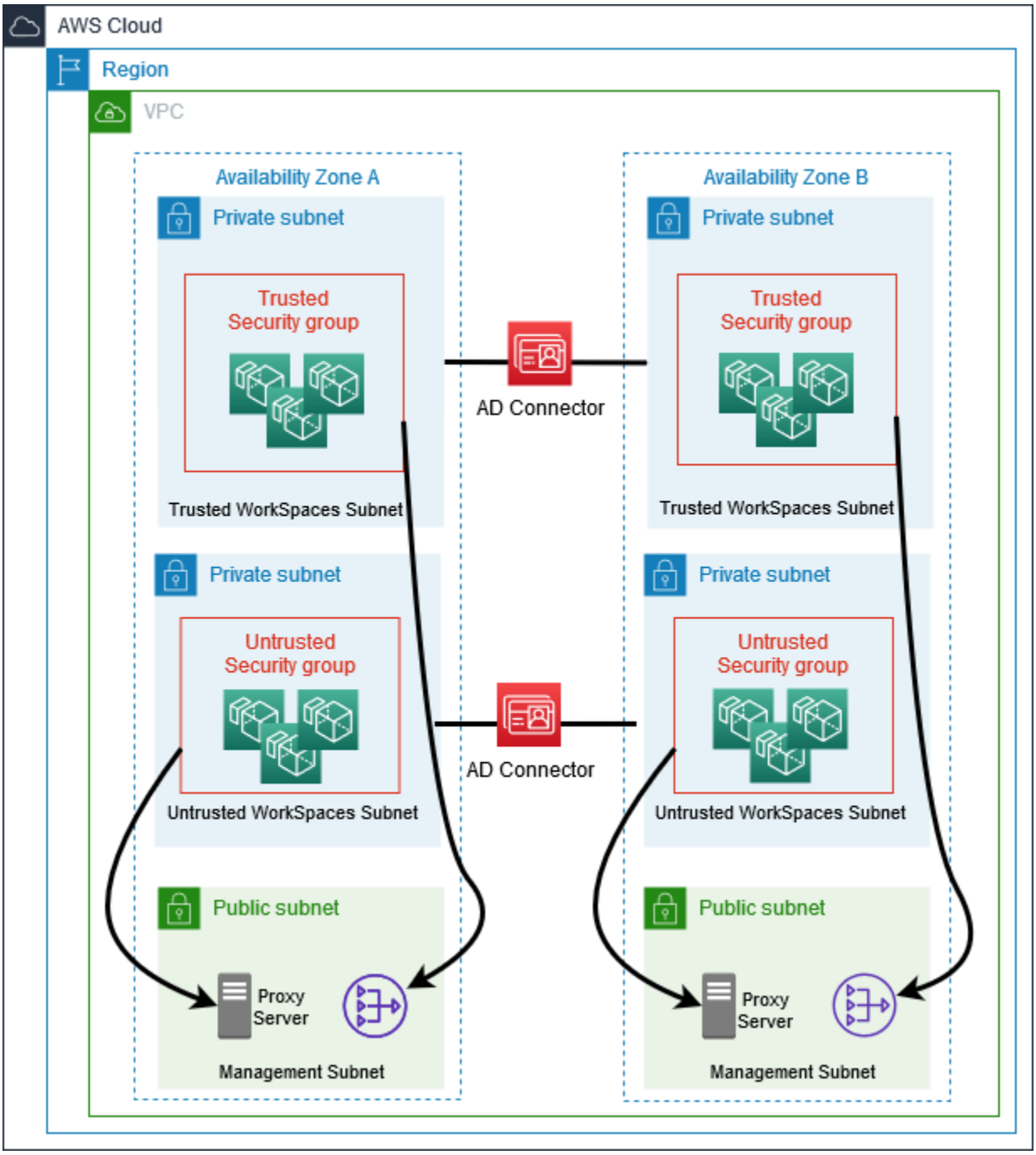


*Figure 3 – Trusted vs Untrusted users – Differing risk profiles satisfied by security groups*

In Figure 3, two different security groups have been created to apply to the trusted and untrusted users, respectively. While in this case, the application of security groups aligns to the subnets for each type of user, it is possible to have multiple and overlapping security groups within the same subnet to apply granular controls. It is important to remember that a default security group is associated with the directory being used with Amazon WorkSpaces and this is applied to a WorkSpace when provisioning through the AWS Management Console.

The application of the default security group can be overridden on a per WorkSpace basis. This approach would allow a mix of both trusted and untrusted users to coexist within the same subnet. This might be desirable in scenarios where there is an IP addressing constraint to reduce the number of subnets that need to be allocated to WorkSpaces. Conversely, the use of coarse-grained subnet-level separation affords an opportunity to group WorkSpaces with the same security profile into the same subnet. Through this approach, it is possible to audit the entire subnet to ensure only WorkSpaces with the same security profile reside on the same subnet.

For simplicity purposes, the diagram does not show this approach and the separation of trusted and untrusted users by subnet would be the preferred approach in a number of scenarios to allow IP address space traceability between the two types of users, which is useful in the event of a security breach investigation having to be undertaken.

In addition to the employment of distinct security groups for the differing user types, the VPC design also shows that trusted users have direct access to the internet through a NAT Gateway whereas the untrusted users have their internet traffic directed through a proxy server to constrain where the users can access the internet.

## VPC design influenced by placement of infrastructure, data, or user

While the risk profile associated with different types of users can influence the design of a VPC, so can the placement of infrastructure (new or existing on-premises), the location of data, and the location of users connecting to the WorkSpaces service.

Ultimately, any WorkSpaces environment hosts applications and these applications often need to consume or access data. To provide the optimal end-user experience, WorkSpaces should be located where the data they need can be accessed with relatively low network latency. Using this approach, applications consuming and interacting with the data are more responsive and end-users have a good user

experience. Therefore, it is critical to consider the physical placement of WorkSpaces alongside data wherever possible.

The location of the users connecting to WorkSpaces must also be considered to avoid incurring high latency connections between the client endpoint device being used by the user and the WorkSpace being connected to by the user. It is not always possible to identify a single Region to provide a WorkSpaces environment for all users and in these instances a multi-Region deployment such as that outlined in the next section must be considered. However, where users are generally located within the same geographic Region, a single regional implementation of WorkSpaces can be considered. Ultimately it is helpful to understand the network latency that your users may incur when connecting to the WorkSpace service in the chosen Region or Regions.

Lastly, the location of the resources required by each WorkSpace can also impact the end-user experience of users. The latency of proxy servers, Active Directory Domain Controllers, infrastructure for deploying operating system and application patches, anti-virus/quarantine servers, monitoring servers, file servers and application servers can all impact an end-user's experience. Therefore, if an existing infrastructure exists either on premises or already in AWS, then this must be factored into your choice of Region to host WorkSpaces.

Figure 4 shows an existing VPC (shown on the right) that has already been implemented in AWS that hosts Active Directory Domain Controllers. This VPC and its associated infrastructure can be re-used to benefit a new WorkSpaces environment through the use of VPC peering. In the architecture, a new VPC (shown on the left) has been created in the same region to host WorkSpaces and through the VPC peering connection these are able to authenticate with the existing Active Directory located in the existing VPC. If other infrastructure or application servers were also present in the existing VPC, then the low latency connection between the two VPCs would help to ensure a good user experience for all users using the WorkSpaces service.

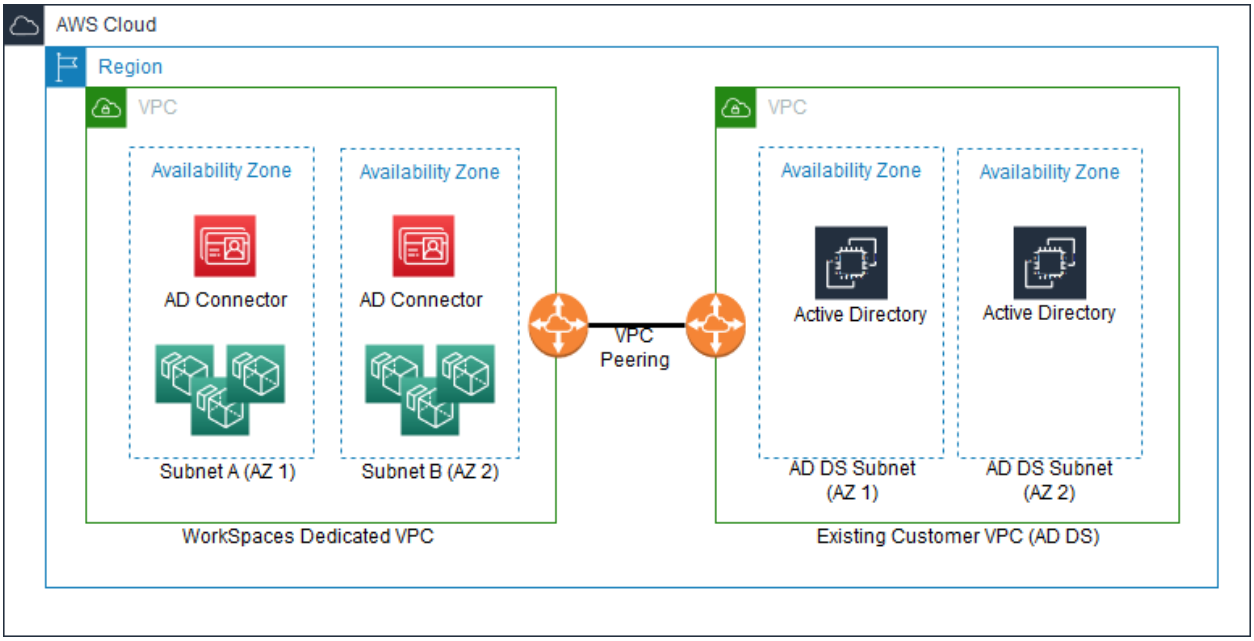


*Figure 4 – Managed Active Directory in an existing peered VPC*

Figure 5 shows a WorkSpaces architecture with a Managed Active Directory implemented across two private subnets and WorkSpaces implemented across an additional two private subnets. This design can be considered an anti-pattern as end-user workloads (that is, WorkSpaces) should be separated from shared and infrastructure services through the implementation of a separate VPC as outlined in Figure 4. This is because there is improved control over routing and placement of critical infrastructure. However, an exception may be a case where a small simple isolated environment is required for end-users.

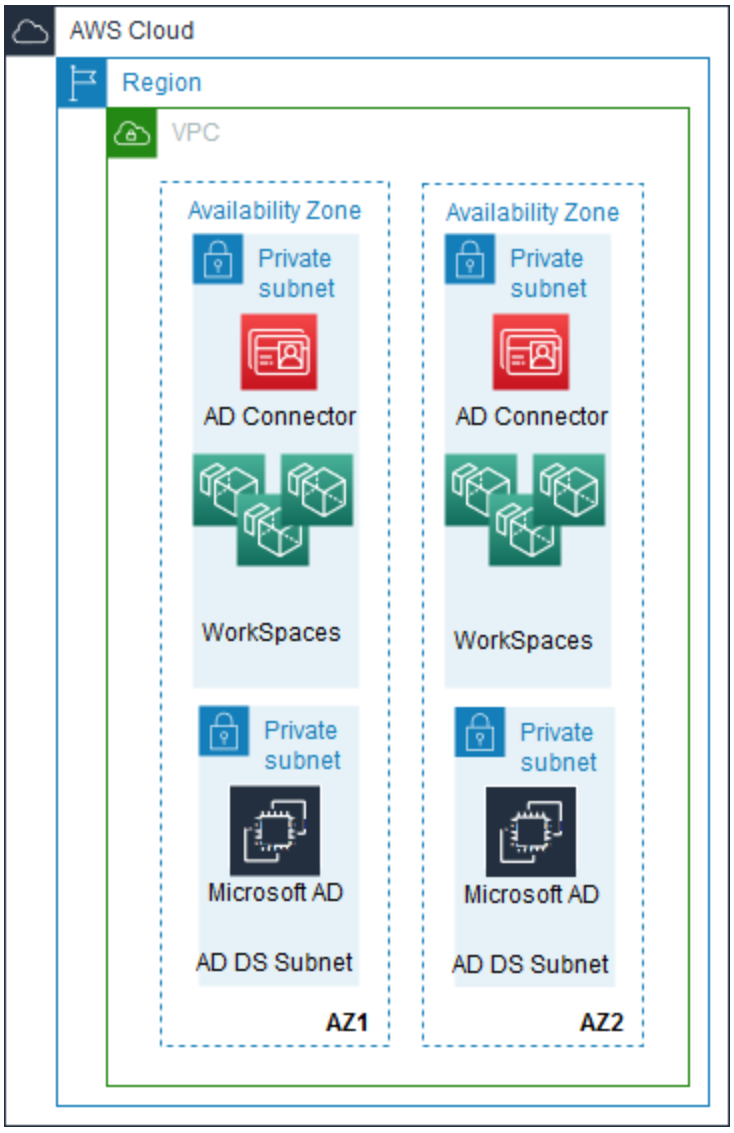


*Figure 5 – WorkSpaces and existing Active Directory in a common VPC*

The two previous VPC designs have been based on existing infrastructure already present in AWS. While Figure 4 captured a common deployment pattern, we have customers that have hybrid cloud environments where infrastructure and application servers still reside on-premises. In these instances, an approach that allows a WorkSpaces environment to interact with the on-premises infrastructure is required.

Figure 6 shows an existing on-premises network with Active Directory Domain Controllers and shared services connecting to a new WorkSpaces environment hosted in a VPC. Connectivity between the two exists through the deployment of AWS Direct Connect (see here) between the VPC and on-premises network and through the use of a Site-to-Site VPN (see here) to provide a redundant connection should the primary

AWS Direct Connect connection fail. Further resiliency can be provided for this architecture through the deployment of Active Directory Domain Controllers on EC2 instances. This would allow WorkSpaces authentication to take place even if both the Direct Connect connection and VPN should fail.

If internet connectivity is required, the WorkSpaces VPC can provide this through the use of an internet gateway (see here). The VPC also ensures that AWS services that have VPC endpoints (see here) can be reached from the private WorkSpaces without the traffic having to traverse the internet.

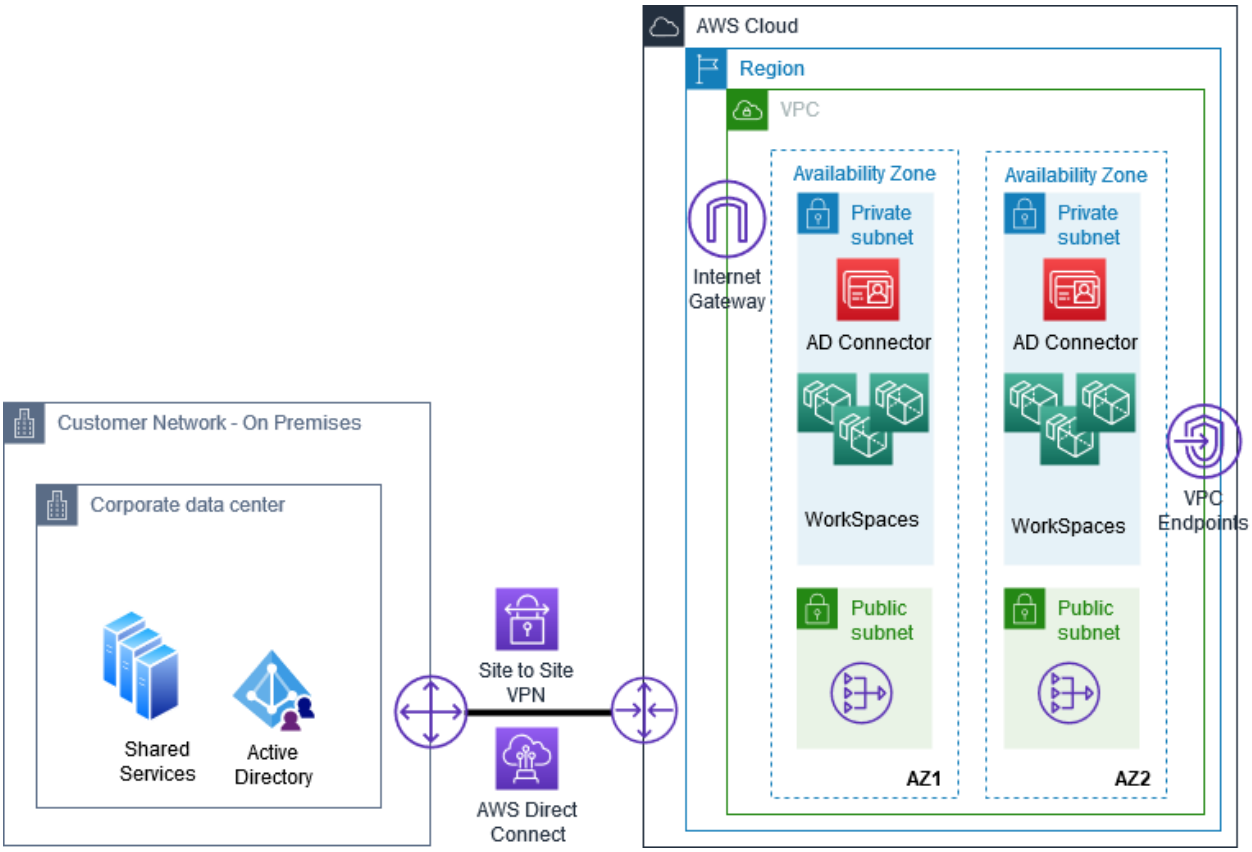


*Figure 6 – Existing on-premises Active Directory connected to WorkSpaces using Direct Connect/Site-to-Site VPN*

# VPC design for multi-regional WorkSpaces deployments with multiple directories

The previous VPC designs have been influenced by risk or by the physical locations of resources (data, infrastructure, or users). It is possible that smaller implementations of WorkSpaces can align with either of these approaches since the overall design could be dominated by one overriding set of requirements. However, for larger implementations a combination of factors must be considered and therefore a hybrid of the previous designs should be considered.

Figure 7 shows a multi-regional WorkSpaces environment implemented across three AWS Regions, six Availability Zones, and an existing on-premises network. This design satisfies a geographically dispersed user base since a user within a user base that spans multiple geographies can connect to their nearest WorkSpace Region. This in turn affords the opportunity for a user to connect using a lower latency network connection than they would do if WorkSpaces were implemented within a single Region globally.

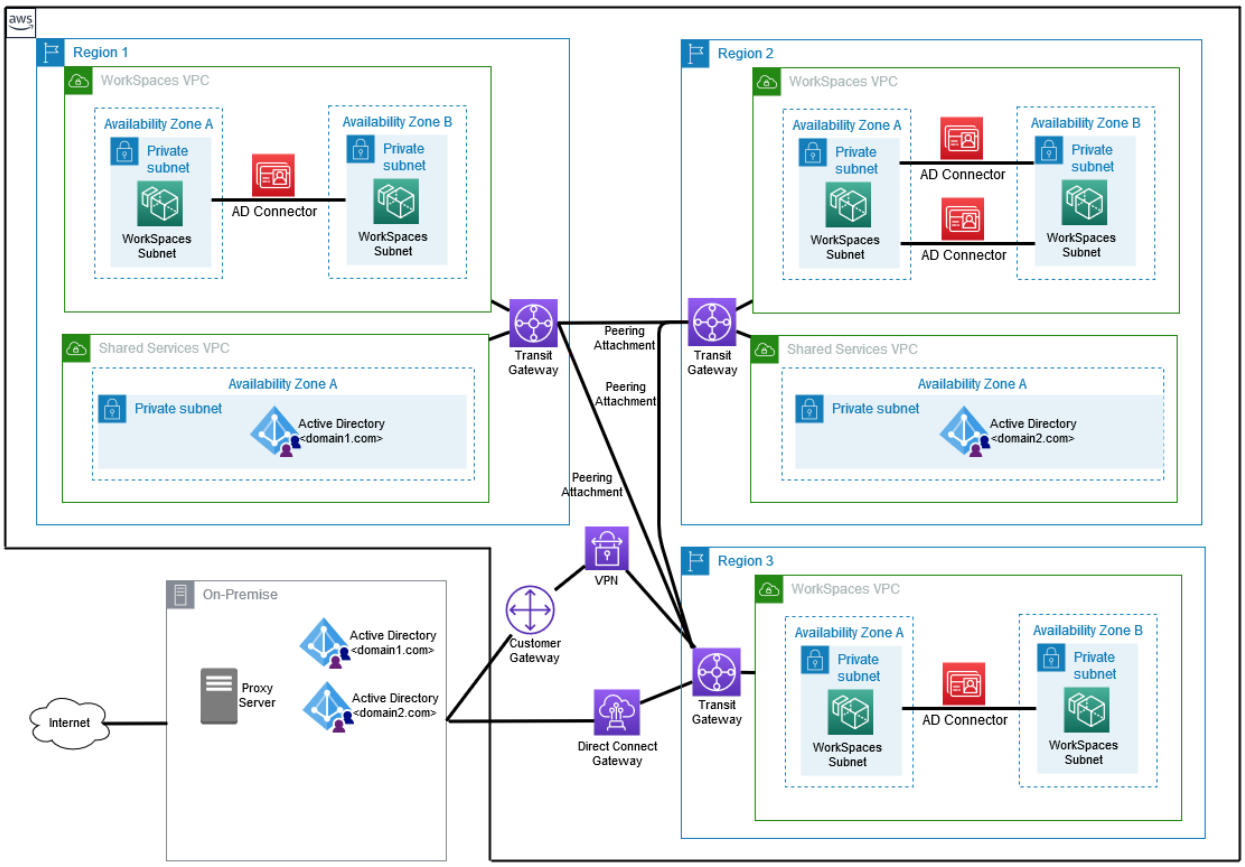


*Figure 7 – Multi-Regional WorkSpaces deployments with multiple directories*

In addition to multiple Regions, there are also multiple Active Directory domains that the WorkSpaces environments participate within. In this example, there are two separate Active Directory forests containing domains called domain1.com and domain2.com. These are completely separate and untrusted, yet the WorkSpaces service can serve users in both forests since there are different Active Directory Connectors associated with a domain in each forest that permits the dual domains to be connected independently to the WorkSpaces service.

In addition to multiple Active Directory forests, Figure 7 also shows the use of the AWS Transit Gateway service (see here) to link all the geographically dispersed VPCs together into a single network. This approach also allows the shared use of a common AWS Direct Connect connection from AWS back to the on-premises network across all Regions and the implementation of a Site-to-Site VPN as a redundant connection in the event of loss of connectivity when using AWS Direct Connect.

Internet connectivity is provided to all WorkSpaces globally through the existing on-premises proxy server and in addition, other shared on-premises application servers can also be accessed. A variation on this design where local internet connectivity is provided is possible through the use of internet gateways in each Region along with additional proxy servers.

An important design aspect to note is the use of multiple Active Directory Connectors (ADC) in Region 2. This approach allows WorkSpaces to be deployed on a single set of subnets in Region 2 that are joined to either domain1.com or domain2.com forests. The first set of ADCs is used for domain1.com and the second set is used for domain2.com. While this approach provides the ability to have mixed domain membership of WorkSpaces in a single Region on a single pair of subnets, it should also be noted that the design only provides domain2.com Domain Controllers in the same Region. Therefore, authentication traffic for domain1.com has to traverse the broader network using AWS Transit Gateway to authenticate with Region 1 or on-premises Domain Controllers.

It must be noted that this design does not provide an optimal end-user experience since authentication and accessing Group Policy Objects for domain1.com has to take place remotely and can be impacted if network connectivity to the on-premises network is lost. Therefore, it is recommended that Domain Controllers are highly available and accessible across more than one network connection to ensure that authentication can take place in the eventuality where the primary network connection fails.

Lastly, Region 3 has no local Domain Controllers and so suffers from the same performance and availability challenge where not only does the authentication traffic have to traverse the broader network but so do Group Policy Objects, login scripts etc. that reside on the Domain Controllers. While this design will suit some deployments its lack of local Domain Controllers and dependency on on-premises infrastructure will be considered too high a risk for large enterprises.

## Streaming and authentication traffic flow for Amazon WorkSpaces

Figure 8 shows the streaming and authentication traffic flow for Amazon WorkSpaces. This diagram is useful to understand a full WorkSpaces environment that is being

accessed by end-users and that also has connectivity to an existing on-premises network.

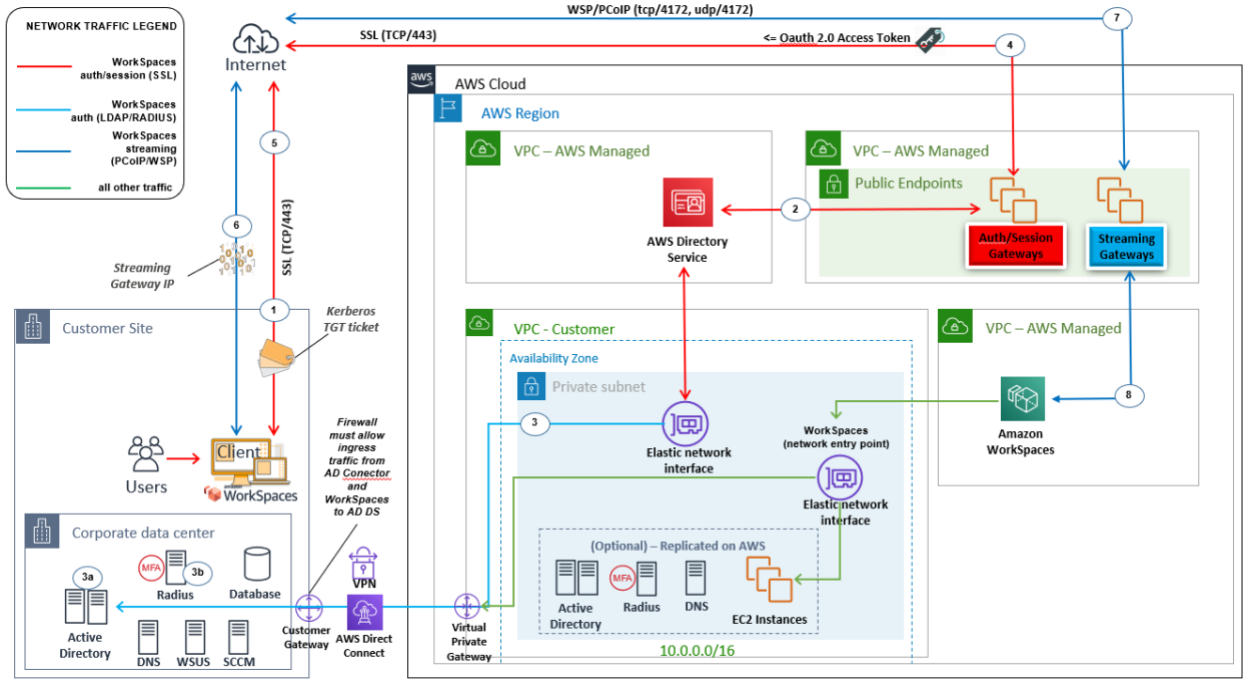


*Figure 8 – Authentication, authorization, and streaming request flow for Amazon WorkSpaces*

Let's walk-through the authentication, authorization, and streaming request flow for WorkSpaces:

1. In the preceding scenario, users are connecting to the WorkSpace service from their corporate office. The user first authenticates (step 1) with a public endpoint (similar to a VPN authentication) using Active Directory credentials and a one-time MFA[1] token. During authentication, the user's credentials are encrypted and sent from the client to the authentication endpoint over a TLS 1.2 tunnel.
2. The credentials are used to authenticate the identity stored in the customer Active Directory (steps 2, 3).
3. If an ADC (Active Directory Connector) is being used, the ADC performs LDAP authentication to the Active Directory (3a, 3b).

[1] The use of MFA with WorkSpaces for authentication is optional but recommended.

4. If the user successfully completes authentication, a one-time OAuth 2.0 token is provided to the client (step 4). If the user fails authentication, no further actions are possible.
5. If the client receives the OAuth 2.0 token post-authentication, the client provides the token to the WorkSpaces service over a TLS 1.2 tunnel (step 5).
6. The token is received by the WorkSpaces service and is used to retrieve information about the user's WorkSpace. The information about the WorkSpace can only be retrieved using a valid OAuth 2.0 token after a user has completed authentication. Once information about the user's WorkSpace is retrieved, the IP address of a PCoIP/WSP Streaming gateway is provided to the user/client over a TLS 1.2 tunnel.
7. The client receives information about the PCoIP/WSP Streaming gateway, and requests a session with the PCoIP/WSP gateway to initiate a session using the OAuth 2.0 token provided (step 6).
8. The OAuth 2.0 token is used by the PCoIP/WSP gateway to request information about the user's WorkSpace from the WorkSpaces service (step 7). This request is again completed over a TLS 1.2 tunnel.
9. Once the gateway receives information about the user's WorkSpace, a streaming session is initiated from the WorkSpace to the user/client via the PCoIP/WSP gateway (step 8). The streaming traffic, all pointer/keyboard interactions and USB traffic are encrypted using AES-256.

From the preceding steps, it can be seen that there is a critical dependency on the availability of network connectivity between the Amazon WorkSpaces infrastructure and your Active Directory. Therefore, the availability of WorkSpaces to your users is dependent on the reachability of your domain controllers. If domain controllers only reside on premises, then the links between the customer VPC and your on-premises network must be highly available to ensure that your users can access their WorkSpaces in the event of the loss of a primary network connection. Options for providing AWS to on-premises network connectivity are considered in the later sections of this document.

Another important point to note is that the private IP addresses of the WorkSpaces instances that are associated with eth1 are never exposed to the public internet due to the use of the streaming gateway.

# Physical network architecture

## AWS Regions

The considerations for the networking components that an Amazon WorkSpaces deployment uses start with choosing one or more AWS regions in which the WorkSpaces instances will physically reside. A Region is a separate geographic area, which provides multiple, physically separated and isolated Availability Zones, which are connected with low latency, high throughput, and highly redundant networking.

Amazon WorkSpaces Design Considerations:

- **Availability of Amazon WorkSpaces service in Region.** The Amazon WorkSpaces services are not currently available in all AWS Regions; therefore, the list of available Regions will need to be reviewed to inform the regional decision. The list of Regions where the Amazon WorkSpaces service is currently available can be obtained from the Amazon WorkSpaces pricing page here: https://aws.amazon.com/about-aws/global-infrastructure/regional-product-services/
- **Location of users.** In which geographies are the users located in? Single, multiple, global? If multiple or global then a multi-Region WorkSpaces implementation will need to be considered. It is recommended that network latency is less than 100 ms when accessing WorkSpaces in the chosen Region. If this is exceeded and user experience is impacted, then additional Regions should be considered.
- **Location of data.** The Windows applications running on the WorkSpaces instances need to interact with data whether it is stored on a database server, file server, or other location. Therefore, the network latency when accessing this data should be factored into the decision process for selecting the Region in order to ensure that the effects of network latency and as a consequence the impact to the user's experience is reduced.
- **Location of application servers.** Much like the physical location of the data that users interact with has a direct impact on the end-user experience of using Amazon WorkSpaces, so does the location of application servers. If application servers are located physically a long distance away (that is, high network latency) from user's WorkSpaces, then their end-user experience with the applications that have a dependency on the application server can be severely degraded.

- **Business Processes**. The business processes that users participate in and the location of where the data and applications reside for these processes must be factored into choosing the most suitable Region. Therefore, it is important to be aware of the business processes and not just the applications, data, and users. For example, if a business process that users in a Region participate in requires data that originates in a different remote Region then the way in which the users interact with this data is critical. If the users need to manipulate large datasets of the data directly using Windows applications, then it is likely that it will be more efficient for the user to use an application hosted locally to the data (for example, a second WorkSpaces instance or locate the user's WorkSpace to the remote Region if possible) rather than transfer the data between Regions to manipulate it within their local WorkSpaces instance.

# Availability Zones

Following the choice of Region, and depending on the chosen AWS Region, a decision might need to be made on which Availability Zones (AZs) to use. The Amazon WorkSpaces service is not currently available in all AZs in every AWS Region and therefore the choice of which AZs to use needs to factor this limitation.

Amazon WorkSpaces Design Considerations:

- **Availability of the Amazon WorkSpaces service in the chosen AZs.** As previously outlined, the WorkSpaces service is not available in all AZs in all Regions, therefore the Regions must be selected based on availability. To determine which AZs in your account support WorkSpaces, see Availability Zones for Amazon WorkSpaces in the *Amazon WorkSpaces Administration Guide.*
- **Availability of any supporting AWS Services in the chosen AZs.** While the availability of the WorkSpaces in AZs can influence the selection of AZs, so can the availability of any supporting AWS services that are required. If a supporting service is not available in the chosen AZ, then traffic will need to traverse between AZs and therefore subnets, resulting in a slightly more complicated design than otherwise.

- **Desired availability of WorkSpaces instances.** It is recommended that all AZs where the Amazon WorkSpaces service is available in a Region are used for deployments to ensure that any possible impact due to the inaccessibility of a single AZ is reduced. This can be achieved by using multiple directories within the WorkSpaces service in a single Region, since a single directory only supports two AZs.

# Logical network architecture

The logical network architecture considerations for an Amazon WorkSpaces implementation cover three types of connections. Firstly, external connectivity from an existing customer's on-premises network to the WorkSpaces environment for reaching the private (eth1) interface of the WorkSpaces instances, secondly connectivity within the customer's AWS environment, and lastly internet connectivity from the WorkSpaces environment. Each of these are explored in the following sections.

## External connectivity: On-premises network to AWS

External connectivity from an on-premises network to AWS might not be relevant for all customers. For example, if you do not have an on-premises network or have no need to connect your WorkSpaces environment to your on-premises environment, because it is a standalone environment.

Where external network connectivity from an existing customer's on-premises network to AWS is required, it is available through two different approaches. The first is the use of a dedicated private network connection with AWS Direct Connect, the second is through the establishment of a Site-to-Site Virtual Private Network (VPN) connection between the customer's on-premises network and the customer's Virtual Private Cloud (VPC).

To achieve a consistent bandwidth and latency between your on-premises network and your WorkSpaces environment, it is better to consider the use of dual Direct Connect connections to ensure that users frequently have a consistent end user experience. The use of combined Direct Connect connection and Site-to-Site VPN will lead to different bandwidth and latencies across the two links and therefore a more variable end user experience for your WorkSpaces users.

## AWS Direct Connect

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

There are two key components that are used for AWS Direct Connect.

- **Connections**. Create a connection in an AWS Direct Connect location to establish a network connection from your on-premises network to an AWS Region.
- **Virtual Interfaces**. Create a virtual interface to enable access to AWS services. A public virtual interface (VIF) enables access to public services, such as Amazon S3 and also the WorkSpaces service itself for streaming traffic. A private VIF enables access to your VPC. When using a private VIF, routes must be published to ensure that traffic can be routed between your VPC and your on-premises network. A transit VIF is a special kind of private VIF for connection to an AWS Transit Gateway and only one can exist per physical connection. A mixture of up to 50 public and private VIFs can be supported per physical connection if using a dedicated Direct Connect connection. If using a hosted connection, only a single VIF is supported.

  The use of a public VIF to transport WorkSpaces streaming traffic is an option that can provide a more reliable network transport for users accessing WorkSpaces from an on-premises network. This leads to a more predictable user experience since the conditions of the network connection such as latency, packet loss etc. are less variable compared to an internet connection.

There are two different models for AWS Direct Connect: A dedicated connection model and a hosted connection model. When using a dedicated connection model, capacities of 1 Gbps or 10 Gbps per connection can be provisioned between your router and an AWS Direct Connect location. With a hosted connection, a partner provides the connection to connect from your router to an AWS Direct Connect location. Various bandwidth connections are available from 50 Mbps up to 10 Gbps. It is important to consider that if an AWS Transit Gateway (see here) is planned to be used that the transit VIF can only use either a dedicated connection or a hosted connection of 1Gbps or greater. Further information regarding hosted and dedicated connections can be read here: https://docs.aws.amazon.com/directconnect/latest/UserGuide/WorkingWithConnections.html

For connections where the use of encryption is mandated by internal security requirements, it is possible to employ a VPN over your AWS Direct Connect connection to ensure that traffic is encrypted and secure. This combination is called AWS Direct Connect Plus VPN and more can be read here: https://docs.aws.amazon.com/whitepapers/latest/aws-vpc-connectivity-options/aws-direct-connect-plus-vpn-network-to-amazon.html

Amazon WorkSpaces Design Considerations:

- **Bandwidth.** Sufficient bandwidth to cater for the resultant traffic that the number of concurrent sessions will generate in terms of authentication and application traffic to on-premises servers needs to be available. In addition, if a public virtual interface (VIF) is being employed for the WorkSpaces streaming traffic then this will also impact on the amount of bandwidth required for the Direct Connect connection. High-Level bandwidth requirements for the streaming traffic for the Amazon WorkSpaces service are outlined in the WorkSpaces FAQ here: https://aws.amazon.com/workspaces/faqs/
- **Latency.** The latency on the Direct Connect link must also be considered since this can impact the end user experience when using WorkSpaces. Latencies above 5 ms can start to impact the performance of Windows desktops when home drives are hosted on a Windows file server with a latency higher than this. Latency to on-premises databases or application servers from the WorkSpaces instances also needs to be considered and monitored to ensure that users receive a good user experience.
- **Jitter.** Jitter is defined as the variance in the time delay in milliseconds between data packets on a network. A large variance in these time delays can impact a user's experience of WorkSpaces. Therefore, jitter should be considered within the design and addressed if there is a large variance.
- **User concurrency.** The number of concurrent users within the WorkSpaces environment should factor in the amount of bandwidth that needs to be provisioned for a Direct Connect connection. The amount of bandwidth required to your on-premises environment will need to be calculated based on the expected usage resulting from application and authentication traffic. In addition, if a public Virtual Interface (VIF) is being used with the Direct Connect connection to transport the WorkSpaces streaming traffic then this bandwidth must also be included in the overall calculation.
- **Number of deployed WorkSpaces.** Like the number of users, the number of deployed WorkSpaces also needs to be factored into the sizing of the AWS Direct Connect connection in terms of bandwidth. If Operating System and application patches, antivirus updates etc. are being distributed from the on-premises infrastructure then all the updates will be distributed to each WorkSpace and sufficient bandwidth must be in place to satisfy any requirements to patch or deploy updates within a defined amount of time.

- **Physical location of supporting services.** The location of services (for example, Active Directory, Software Distribution, antivirus, Monitoring servers etc.) that support the ongoing operation of the WorkSpaces instances will impact the sizing of the AWS Direct Connect connection and the latency of this link can also impact the end user experience for users using the WorkSpaces environment. The link should be sized to factor in this traffic.
- **Desired availability.** Your WorkSpaces environment will have an expected level of availability from your business users. If an AWS Direct Connect connection is required for the day to day operation of the WorkSpaces and the successful use of applications installed on the WorkSpaces instances, then it is important to consider the availability of this connection. If you were to lose the Direct Connect connection what would the impact be to your business? The resultant risk and your organization's risk appetite should be used to determine if a secondary Direct Connect connection is required, whether a fallback VPN connection is required, a redundant Direct Connect connection or indeed whether no redundancy is required.
- **Encryption.** As outlined previously, where security requirements mandate the use of encryption, it is possible to encrypt AWS Direct Connect connections to ensure that all traffic is secure while in transit.

Further information regarding the use of AWS Direct Connect can be found here:

https://docs.aws.amazon.com/directconnect/latest/UserGuide/Welcome.html

## AWS Site-to-Site VPN

By default, WorkSpaces that you launch into an Amazon VPC cannot communicate with your own (remote) network. You can enable access to your remote network from your VPC by attaching a Virtual Private Gateway to the VPC, creating a custom route table, updating your security group rules, creating an AWS Site-to-Site VPN connection, and configuring routing to pass traffic through the connection. This is an alternative approach for providing remote connectivity to the use of Direct Connect outlined previously.

A Virtual Private Gateway or AWS Transit Gateway is the VPN concentrator on the AWS side of the Site-to-Site VPN connection. You can create a virtual private gateway and attach it to the VPC from which you want to create the Site-to-Site VPN connection.

Amazon WorkSpaces Design Considerations:

The design considerations when using a Site-to-Site VPN are very similar to the Direct Connect connection outlined previously. However, in addition to these, the following should be considered.

- **Variable latency.** Since the AWS Site-to-Site VPN service traverses the public internet, the latency on these types of connection can vary more than an AWS Direct Connect connection. As a consequence, the resultant end user experience can vary accordingly. The choice of using an AWS Site-to-Site VPN should factor in the potential for variable latency.

Further information regarding the use of VPNs with AWS can be found here:

https://docs.aws.amazon.com/vpn/latest/s2svpn/VPC_VPN.html

## Direct Connect gateway

Use AWS Direct Connect gateway to connect your VPCs. You associate an AWS Direct Connect Gateway with either an AWS Transit Gateway (when you have multiple VPCs in the same Region) or a Virtual Private Gateway (VGW) for a single VPC.

A Direct Connect gateway is a globally available resource. You can create the AWS Direct Connect gateway in any public Region and access it from all other public Regions.

When a Direct Connect gateway is used with a Direct Connect connection, it provides the ability to connect to multiple Transit Gateways that in turn connect to multiple Regions rather than being constrained to the Region where the Direct Connect connection terminates.

Amazon WorkSpaces Design Considerations:

- **Number of routes.** There is a maximum number of public routes per Border Gateway Protocol (BGP) session on a public virtual interface for Direct Connect. In addition, there is also a maximum number of private routes on a BGP session on a private virtual interface. These limits are documented here: https://docs.aws.amazon.com/directconnect/latest/UserGuide/limits.html
- **Number of Virtual Private Gateways per AWS Direct Connect gateway.** Limits are documented here: https://docs.aws.amazon.com/directconnect/latest/UserGuide/limits.html

- **Number of Direct Connect dedicated connections per Region per account.** Limits are documented here: https://docs.aws.amazon.com/directconnect/latest/UserGuide/limits.html
- **Direct Connect gateway cannot be used to connect a VPC in the China Regions.**

Further information regarding the use of Direct Connect Gateway can be found here:

https://docs.aws.amazon.com/directconnect/latest/UserGuide/direct-connect-gateways.html

# Internal connectivity: Within AWS

Customer connectivity within the AWS Cloud leverages a number of capabilities within the Amazon Virtual Private Cloud (Amazon VPC) service. These are explored in the following sections.

## VPC (Virtual Private Cloud)

The fundamental building block for networking in an AWS Region is a VPC (Virtual Private Cloud). A VPC must be created in order for Amazon WorkSpaces instances to be provisioned. Each VPC is allocated a range of IP addresses by each customer and the creation of a VPC for WorkSpaces must consider a number of factors.

It is important to note that a VPC imposes some constraints that may impact applications planned for hosting on Amazon WorkSpaces. Firstly, by default, it is not possible to broadcast IP traffic in an Amazon VPC (see VPC FAQ here: https://aws.amazon.com/vpc/faqs). Secondly, the number of VPCs that can be created by default in a Region is 5, but it should be noted that this limit can be increased by raising a support request with AWS through the AWS Management Console.

Amazon WorkSpaces Design Considerations:

- **Unique IP Address Range.** A Unique IP address range needs to be allocated per VPC. This must be a CIDR (Classless Internet Domain Routing) block of IP addresses and takes the form of: 10.0.0.0/20 or 10.0.0.0/22 for example. Most ranges used within VPCs fall within the private (non-publicly routable) IP address ranges specified in RFC1918; however, you can use publicly routable CIDR blocks for your VPC.

- **Non-overlapping IP address range.** Non-overlapping IP address ranges must be used between VPCs and your on-premises network to avoid any address conflicts and therefore IP routing challenges if you intend for those networks to communicate.
  In addition to avoiding overlapping IP address ranges with your on-premises network, it is also important to avoid overlapping with the IP address ranges reserved for the WorkSpaces Management Interface. These ranges are listed here:
  https://docs.aws.amazon.com/workspaces/latest/adminguide/workspaces-port-requirements.html#network-interfaces
- **Suitably sized CIDR block.** A suitably sized CIDR block must be allocated to the VPC. The maximum size is a /16 block, however, the block size should be determined based on the number of network addresses that will be consumed by the Amazon WorkSpaces service. Each Amazon WorkSpace instance will require the consumption of a single IP address from this block. However, additional addresses are required for supplementary services such as a directory service or any supporting infrastructure that needs to reside in the same VPC as the WorkSpaces instances.

> **Note**: It is possible to add IP addresses to an existing VPC that already has a CIDR block associated with it. This approach and its limitations are documented here:
> https://docs.aws.amazon.com/vpc/latest/userguide/VPC_Subnets.html#vpc-resize

## Subnets

Following the creation of a VPC and allocation of a CIDR block to it, the next step is to create the subnets where the WorkSpaces instances will be provisioned and connected to. Subnets are created and associated with a single AZ and therefore align with the physical architecture of the WorkSpaces environment.

Amazon WorkSpaces Design Considerations:

- **IP Address Reservations.** Five IP addresses are consumed per subnet by AWS by default in every subnet (see here). You will therefore need to size your subnets with this in mind.

- **Subnet sizing.** Subnets should be sized based on the forecasted number of WorkSpaces that will reside in them as they are permanent and cannot be changed once created. Creating a subnet that is too small will result in wastage due to the five IP address reservation listed previously and also incur management and operational overheads (maintaining additional NACLs, route tables, additional subnets in Active Directory Sites and Services etc.). Creating a subnet that is too large will also result in wastage due to a large IP address allocation not being consumed by WorkSpaces instances. Therefore, a balance must be struck in terms of not sizing a subnet too small or too large.
- **User segregation.** As described earlier (see here), it may be desirable to separate different types or categories of users from each other on a per subnet level. An example use case would be for users performing payment operations where PCI compliance needs to be considered. The separation of non-PCI users from PCI users reduces the scope of security controls that need to be applied across the entire user base since only PCI users need to have the strongest security controls applied to them to maintain PCI compliance.
- **Contactable Dependent Services.** WorkSpaces instances are members of a Windows Active Directory and therefore an Active Directory must be contactable from the subnets that are chosen for deployment. Other dependent services for your Windows applications must also be considered.
- **Private vs. public subnets.** Refer to the next two sections.

### Private subnets

A private subnet is one that has been allocated a private address range as specified in RFC1918. The use of private subnets is recommended with WorkSpaces to ensure that the WorkSpaces are not directly connected to the internet. This approach improves the security posture of WorkSpaces instances over and above WorkSpaces connected to a public subnet. Internet connectivity from individual WorkSpaces to the internet is still possible through the use of NAT (see here) and internet gateways (see here) as outlined later in this document.

### Public subnets

A public subnet is one that has a route table associated with it that has a route to an internet gateway. Instances in a public subnet with Elastic IPv4 addresses, which are public IPv4 addresses enable them to be reached from the internet. While the WorkSpaces service does provide an opportunity to connect WorkSpaces to a public

subnet this is not the recommended approach since instances are directly exposed to the internet and therefore potentially vulnerable to malware and network attacks.

## Route table

A route table contains a set of rules, called routes, that are used to determine where network traffic is directed. Each subnet in your VPC must be associated with a route table. The table controls the routing for the subnet. A subnet can only be associated with one route table at a time, but you can associate multiple subnets with the same route table.

Amazon WorkSpaces Design Considerations:

- **Essential Services Can Be Reached**. Ensure that WorkSpaces instances can route to your Active Directory for user authentication, software distribution, antivirus servers etc.
- **Application Servers Can Be Reached**. Ensure that all application servers (web servers, file servers, database servers etc.) are reachable with a route contained within the route table.
- **Internet Connectivity is Available (Optional)**. Ensure that a route is included to the NAT Gateway if the WorkSpaces instances will access the internet using a NAT Gateway. If a NAT Gateway is not desired because either the WorkSpaces will not have internet connectivity or an existing on-premises proxy server will be used, then this consideration can be discounted.

Further considerations on the use of and design of route tables can be found here:

https://docs.aws.amazon.com/vpc/latest/userguide/VPC_Route_Tables.html

## Security groups

A security group acts as a virtual stateful firewall for your WorkSpaces instances to control inbound and outbound traffic. Security groups can be applied to WorkSpaces instances at provisioning time or post-provisioning and act at the instance level, not the subnet level. Therefore, different WorkSpaces instances can have the same or different security groups applied to them to enforce a differing degree of network control depending on your requirements.

Security groups are useful to achieve a fine-grained level of control over WorkSpaces instances when user segregation is required for different types of user (for example, support staff, Business Process Outsourcing (BPO) staff, developers, knowledge workers, task workers, power users, etc.). Each of these groups can have a different

security group associated with them to enforce varying levels of network access control depending on the network resources that the users are entitled to access.

Amazon WorkSpaces Design Considerations:

- **Port Restrictions.** Enforcing a restriction on the range of TCP and UDP ports that a WorkSpaces instance can use is considered best practice. However, where a large application portfolio is in scope for deployment to WorkSpaces instances, the practicalities of creating and maintaining a set of whitelisted ports can be challenging from an administrative standpoint. This is due to the requirement to have intimate knowledge of the application servers being connected to by each application on an ongoing basis.
- **IP Address Restrictions.** Security groups can be configured for WorkSpaces instances to specifically only allow access to specific whitelisted addresses, for tighter security controls. However, where a large application portfolio is in scope for deployment to WorkSpaces instances the practicalities of creating and maintaining a set of whitelisted IP addresses can be challenging from an administrative standpoint. This is due to the requirement to have intimate knowledge of the application servers being connected to by each application on an ongoing basis. Any IP address changes due to infrastructure or application server moves, decommissioning, failover, or load balancing, will impact the conversation map and therefore the security group configuration.
- **Enable Specific Ports.** Within a WorkSpaces environment that is considered to have a high risk profile (see here), it is considered best practice to define a set of common ports that are whitelisted to permit traffic for maintenance or administrative purposes. For example, enabling TCP port 3389 for RDP administrative access to WorkSpaces can be extremely useful for providing remote support to individual WorkSpaces. If this TCP port is enabled, then the access should be limited to a range of source IP addresses that are identifiable as administrative IP ranges.

Further considerations for the use of security groups can be found here:

https://docs.aws.amazon.com/vpc/latest/userguide/VPC_SecurityGroups.html

## Network access control list

A network access control list (network ACL) is an optional layer of security for your VPC that acts as a stateless filter for controlling traffic in and out of one or more subnets. You might set up network ACLs with rules in order to add an additional layer of security to your VPC such that each instance in a specific subnet might be permitted to access a

specific range of IP address on-premises for example. Each network ACL rule set is associated with one or more subnets.

Amazon WorkSpaces Design Considerations:

- **Coarse-Grained Control**. Network ACLs provide control over the IP traffic associated with all WorkSpaces instances on a per-subnet level and are therefore considered to offer a coarse-grained network security control compared to security groups, which offer a per instance level of control. Network ACLs only work at the subnet boundary and do not affect traffic between peers within the subnet.
- **Open by Default**. By default, the rules in network ACLs are open and allow all inbound and outbound IPv4 and if applicable IPv6 traffic. If the default rule is deleted, then all traffic will be blocked.
- **Mandatory**. Each subnet must have a network ACL associated with it.
- **One to Many**. A network ACL can be applied to multiple subnets.
- **Stateless**. As Network ACLs are stateless, they will not accept traffic unless it is specifically authorized in a rule. This means that extra care must be taken when using Network ACLs to ensure you do not accidentally disrupt functionality. Stateful firewalls, such as security groups, are more straightforward to use as they are aware of traffic passing through them and will automatically accept returning established traffic without requiring a rule.

Further considerations for the use of network ACLs can be found here:

https://docs.aws.amazon.com/vpc/latest/userguide/vpc-network-acls.html

## DNS

When WorkSpaces instances are launched in a VPC, a hostname is automatically created and assigned to them. These names must be resolvable via DNS in order for Kerberos authentication to work with your Active Directory. DNS is therefore an essential infrastructure service within the operation of a WorkSpaces environment. There are a few options to consider for DNS.

By default, AWS provides a DNS server and WorkSpaces instances will be registered with this. An alternative approach is to use an existing on-premises DNS infrastructure that is reachable from the VPC where the WorkSpaces instances will reside. Lastly, it is possible to extend your existing DNS infrastructure into AWS and host your own DNS infrastructure on EC2 instances or alternatively use the Amazon Route 53 resolver.

When choosing to host your own DNS, it is important to update the DHCP Options Sets associated with the WorkSpaces VPC to ensure that WorkSpaces instances are configured with the correct IP addresses of your DNS servers. DHCP Options Sets are discussed in the next section.

Amazon WorkSpaces Design Considerations:

- **Use of existing on-premises DNS**. As previously outlined, it is important to update DHCP option sets to include the IP addresses of your DNS servers, not for WorkSpaces instances themselves (see the next section) but for other servers sharing the same subnets.
- **Extend existing on-premises DNS into AWS**. Amazon EC2 or Route 53 can be used to extend your on-premises DNS into AWS. Route 53 resolver endpoints can be used as a way to interact between your on-premises and AWS DNS resolution.
- **Use of the AWS provided DNS**. AWS provides a DNS service within your VPC by default. This can be used to resolve hostnames for WorkSpaces and EC2 instances deployed within your VPC.

Further information regarding the configuration and use of DNS within an Amazon VPC can be found here: https://docs.aws.amazon.com/vpc/latest/userguide/vpc-dns.html

## DHCP options sets

With an Amazon VPC, DHCP services are provided by default for your instances. When you create a VPC, we automatically create a set of DHCP options and associate them with the VPC. This set includes two options: `domain-name-servers=AmazonProvidedDNS`, and `domain-name=domain-name-for-your-Region`. `AmazonProvidedDNS` is an Amazon DNS server, and this option enables DNS for instances that need to communicate over the VPC's internet gateway. The string `AmazonProvidedDNS` maps to a DNS server running on a reserved IP address at the base of the VPC IPv4 network range, plus two. For example, the DNS server on a 10.0.0.0/16 network is located at 10.0.0.2 and is called the Route 53 resolver. For VPCs with multiple IPv4 CIDR blocks, the DNS server IP address is located in the primary CIDR block. The DNS server does not reside within a specific subnet or Availability Zone in a VPC.

DHCP options sets are used within an Amazon VPC to define scope options, such as the domain name or the name servers, that should be handed to your instances via DHCP. Correct functionality of Windows services and WorkSpaces within your VPC depends on this DHCP scope option and you need to set it correctly. In each of the

designs discussed earlier, you would create and assign your own scope that defines your domain name and name servers. This ensures that domain-joined Windows instances or WorkSpaces are configured to use the Active Directory DNS.

The following table is an example of a custom set of DHCP scope options that must be created for WorkSpaces and AWS Directory Services to function correctly.

| Parameter | Value |
|---|---|
| **Name tag** | Creates a tag with a key = name and value set to a specific string.<br>For example: example.com |
| **Domain name** | example.com |
| **Domain name servers** | DNS Server addresses, separated by commas.<br>For example: 10.0.0.10, 10.0.1.10 |
| **NTP servers** | If required, specify the IP addresses of your NTP servers. |
| **NetBIOS servers** | If being used, enter the comma-separated IP addresses of WINS servers.<br>Example: 10.0.0.11, 10.0.1.11 |
| **NetBIOS node type** | 2 |

In the **Domain name servers** parameter shown in the table above, the IP addresses defined in your scope options are specific to your environment. However, it should be noted that Windows WorkSpaces do not use this setting. DNS server IP addresses are provided by Amazon WorkSpaces Windows services running within your WorkSpaces instance operating systems and are set at service startup.

For WorkSpaces using either a Microsoft Managed Active Directory or Simple Active Directory, the setting of DNS servers is determined by the IP addresses of the first two domain controllers. Where an ADC is used, the setting is determined based on the configuration value for “Existing DNS settings” on the ADC. It is important to be aware that a change to this value does not automatically get propagated to existing WorkSpaces. Therefore, it is recommended to set the `HKLM:\SOFTWARE\Amazon\Skylight\DomainJoinDns` registry value, which is a `REG_SZ` value, to a comma-delimited list of two IP addresses (for example, 10.0.0.1,10.0.0.2) via a Group Policy Preference to ensure that WorkSpaces can be updated with new IP addresses when a change needs to be made.

The **Domain name servers** DHCP option setting will only be used by non-WorkSpaces EC2 instances residing on the same subnet as your WorkSpaces. If you have existing DNS servers you would like to use for your WorkSpaces environment, then these must be listed here. Conversely if you would like to use the Amazon provided DNS server or Route 53 service, then the IP addresses for the respective DNS servers must be listed instead.

Amazon WorkSpaces Design Considerations:

- **Correct Configuration of the DHCP Options Set.** As outlined in the previous table, it is important to consider all the parameters as these all can influence the operation of your WorkSpaces environment.
- **DNS Filtering.** It is not possible to filter traffic to or from a DNS server using network ACLs or security groups.

## VPC endpoints and endpoint services

Figure 6 earlier in this document shows a VPC Endpoint deployed into a customer VPC. A VPC endpoint enables you to privately connect your VPC to supported AWS services and VPC endpoint services powered by AWS PrivateLink without requiring an internet gateway, NAT device/gateway, VPN connection, or AWS Direct Connect connection. Instances in your VPC do not require public IP addresses to communicate with resources in the service. Traffic between your VPC and the VPC endpoints configured for AWS services do not leave the Amazon network.

Endpoints are virtual devices. They are horizontally scaled, redundant, and highly available VPC components that allow communication between your instances in your VPC and services without imposing availability risks or bandwidth constraints on your network traffic.Figure 7

Amazon WorkSpaces Design Considerations:

- **Use VPC endpoints where possible.** To prevent the need for public traffic to AWS services, use VPC endpoints where they are available for services. This ensures that traffic is retained within AWS and there are no availability or bandwidth constraints to consider.
- **Use the Amazon WorkSpaces AWS PrivateLink endpoint.** Amazon WorkSpaces Public APIs are accessible with AWS PrivateLink. AWS PrivateLink is a type of VPC Endpoint called an “Interface VPC Endpoint”. It is recommended that AWS PrivateLink is used to access the WorkSpaces public APIs wherever possible to restrict API traffic within your Amazon VPC and isolate API traffic from the internet. Further information is available here: https://docs.aws.amazon.com/vpc/latest/userguide/vpce-interface.html

## Elastic network interface

As outlined earlier in the document in Figure 1, each WorkSpaces instance is provisioned with two network interfaces (eth0 and eth1). These are elastic network interfaces. The use of elastic network interfaces is covered earlier in this document (see here).

As a consequence of having an elastic network interface within your VPC per WorkSpaces instance it is important to ensure that sufficient capacity is available to create an elastic network interface for every WorkSpaces instance created in your VPC. Any limit encountered in the number of ENIs that can be created will impact your ability to provision WorkSpaces instances.

Amazon WorkSpaces Design Considerations:

- **Ensure that sufficient elastic network interface capacity is available**. This is essential in order to create an elastic network interface for each WorkSpaces instance.
- **Do not modify the elastic network interface associated with a WorkSpaces instance.** The WorkSpaces service is a managed service and therefore fully manages the creation of ENIs for WorkSpaces instances. Any modification to an elastic network interface outside of the WorkSpaces service may impact the ability of the instance to connect to and communicate with your VPC.

Further information is available here:
https://docs.aws.amazon.com/vpc/latest/userguide/VPC_ElasticNetworkInterfaces.html

## VPC peering

A VPC peering connection is a network connection between two VPCs that enables you to route traffic between them using private IPv4 or IPv6 addresses. An example of a VPC peering connection is shown earlier in this document in Figure 4. Instances in either VPC can communicate with each other as if they are within the same network, if routes have been added to the respective subnets in the VPCs. You can create a VPC peering connection between your own VPCs, or with a VPC in another AWS account. The VPCs can be in different Regions (also known as inter-Region VPC peering connection).

When creating a VPC peering connection, a connection request must be created in association with one VPC and accepted by the other VPC. If the accepting VPC is in another AWS account, then the acceptance action will need to be undertaken in the separate AWS account.

Name resolution is another consideration that must be considered when creating a peering connection. The DNS settings for the peering connection may need to be modified to enable name resolution to take place seamlessly between the VPCs.

To send private IPv4 traffic from your WorkSpaces instances to instances in a peer VPC, you must add a route to the route table that is associated with your subnet in which your WorkSpaces instances reside. The owner of the peered VPC must also add a route to their subnet's route table to direct traffic back to the WorkSpaces instances in your VPC. To avoid the route having a `blackhole` state, this must be undertaken while a peering connection is in the `pending-acceptance` state. A route in a state of `blackhole` will have no effect until the VPC peering connection is in the `active` state.

The shared services design (Existing Customer VPC) in Figure 4 in this document outlines a high-level approach for leveraging VPC peering between a dedicated WorkSpaces VPC and a Shared Services VPC that contains core infrastructure services such as Active Directory Domain Controllers and other supporting services (for example, software distribution, antivirus, etc.).

Amazon WorkSpaces Design Considerations:

- **CIDR blocks cannot overlap**. It is not possible to create a VPC peering connection when CIDR blocks overlap between VPCs.
- **Partial overlapping CIDR blocks cannot exist**. If the VPCs have multiple IPv4 CIDR blocks and one or more of these conflicts between the VPCs, then it is not possible to create a peering connection.
- **Transitive routing is not possible**. If three VPCs (A, B, C) are connected such that A connects to B and B connects to C, it is not possible to route traffic between A and C through B.
- **Name Resolution**. How name resolution using DNS will take place needs to be considered when creating the peering connection to ensure that all hostnames can be resolved between the VPCs. Amazon Route 53 should be considered as one of the options.
- **Update Route Tables**. Route tables need to be updated to ensure that traffic can flow between the peered VPCs.

Further information is available here:
https://docs.aws.amazon.com/vpc/latest/peering/what-is-vpc-peering.html

### AWS Transit Gateway

An AWS Transit Gateway is a network transit hub that you can use to interconnect your VPCs and on-premises networks.

An AWS Transit Gateway acts as a regional virtual router for traffic flowing between your VPC and VPN/Direct Connect connections. An AWS Transit Gateway scales elastically based on the volume of network traffic.

Figure 7 earlier in this document shows how an AWS Transit Gateway can be employed in a multi-regional WorkSpaces design to provide network connectivity between Regions and also back to an on-premises network.

Amazon WorkSpaces Design Considerations:

- **Transitive routing through the AWS Transit Gateway is possible**. The use of an AWS Transit Gateway addresses the transitive routing limitation imposed when using VPC peering.

Further considerations for the use of AWS Transit Gateway can be found here:

https://docs.aws.amazon.com/vpc/latest/tgw/what-is-transit-gateway.html

## Internet connectivity

Internet connectivity can be provided to WorkSpaces in a variety of ways. The approach to provide connectivity often depends on the existing security policies of your organization. Existing on-premises internet connectivity can be used or alternatively AWS services such as the internet gateway and NAT Gateway can be used to provide internet access. The use of AWS services to provide internet connectivity to WorkSpaces is outlined in the next two sections.

### Internet gateway

An internet gateway is a horizontally scaled, redundant, and highly available VPC component that allows for communication between instances in your VPC and the internet. It therefore imposes no availability risks or bandwidth constraints on your network traffic.

An internet gateway serves two purposes: to provide a target in your VPC route tables for internet-routable traffic, and to perform network address translation (NAT) for instances that have been assigned public IPv4 addresses (see public subnets here).

When determining whether to provision an internet gateway for your WorkSpaces environment it should be determined whether there is an existing on-premises proxy that must be used, due to internal security requirements, that would prevent the use of the internet gateway. In some scenarios, it may be permissible for a new proxy service to be created in AWS to prevent all internet traffic flowing back to the on-premises network. In other scenarios a proxy server may not be required at all. The requirement to use a proxy server or not will be based on an organization's security posture and requirements.

Amazon WorkSpaces Design Considerations:

- **Will users need access to browse the internet from AWS and not via the on-premises network?** This might be desirable to reduce the amount of internet traffic traversing a VPN or Direct Connect connection that is generated by WorkSpaces instances.
- **Will operating system patches and application updates need to be obtained directly via the internet using AWS internet connectivity or will they be provided using existing on-premises infrastructure?**

If either or both of the answers to these questions is true, then an internet gateway will be required. Further considerations of the internet gateway can be found here:

https://docs.aws.amazon.com/vpc/latest/userguide/VPC_Internet_Gateway.html

## NAT Gateway

A Network Address Translation (NAT) Gateway can be used to enable instances in a private subnet to connect to the internet or other AWS service, but prevent the internet from initiating a connection with those instances.

> **Note**: If you want to use a NAT Gateway, an internet gateway is required to ensure that the NAT Gateway has internet connectivity.

When thinking about whether to provision a NAT Gateway for your WorkSpaces environment, it should be determined whether there is an existing on-premises proxy that must be used due to internal security requirements that would prevent the use of a NAT Gateway. In some scenarios, it may be permissible for a new proxy service to be created in AWS to prevent all internet traffic flowing back to the on-premises network.

Amazon WorkSpaces Design Considerations:

- **When permitted by policy, use a NAT Gateway and do not use public subnets for WorkSpaces.** While Amazon WorkSpaces instances can be deployed in public subnets, it is recommended not to expose them to the internet but rather deploy them in private subnets and optionally provide internet connectivity via a NAT Gateway.
- **A NAT Gateway can only be deployed in a public subnet.** This will require a route to your WorkSpaces VPC and subnets in order to support outbound internet connectivity from WorkSpaces instances.
- **Each NAT Gateway is deployed in a specific Availability Zone.** To provide resiliency, deploy a NAT Gateway per WorkSpaces Availability Zone.
- **A NAT Gateway can simplify the provisioning of internet connectivity.** When WorkSpaces are accessing the internet via a NAT Gateway, they all appear to be using the same public IP address. Where you need to explicitly allow (whitelist) access, this approach makes it easier since only a single IP address needs to be allowed. Conversely, if an internet gateway was being used instead, each WorkSpace would have its own IP address and therefore a range of IP addresses must be added to the allow list.
- **Do not "Enable Internet Access" on your WorkSpaces directory when using a NAT Gateway.** By not selecting the **Enable Internet Access** check box on your WorkSpaces directory, you prevent the attachment of public IP addresses to your WorkSpaces. This is the recommended setting when using a NAT Gateway since individual WorkSpaces do not require a public IP address.

Further considerations of the NAT Gateway can be found here:
https://docs.aws.amazon.com/vpc/latest/userguide/vpc-nat-gateway.html

# Monitoring

The monitoring of the network infrastructure that underpins your WorkSpaces environment is critical to ensure that users are not being impacted when there is a degradation of service or potential impact to end users because of the loss of a network link. The topic of monitoring is too broad for this document and your existing monitoring tools for monitoring networks should be used where possible.

## Internet access monitoring

If existing on-premises internet proxies are employed within your design to provide internet access to WorkSpaces, then these should be monitored for availability and response times to ensure that users have a consistent experience.

If a NAT Gateway has been employed as an alternative, then the AWS NAT Gateway provides both CloudWatch Alarms and Metrics. Alarms can be defined for specific thresholds based on your use case and metrics can be used for real time monitoring and defining alarms.

Beyond the use of a NAT Gateway, proxy servers can be used to monitor traffic and whitelist or blacklist internet access from WorkSpaces. This can be done within a virtual network appliance from the AWS Marketplace or built on an EC2 instance. However, the NAT Gateway itself does not provide the capability to create user logs.

### Connectivity monitoring

General connectivity of your WorkSpaces environment to Amazon VPC networking components can be monitored by using VPC Flow Logs. In addition, Amazon CloudWatch Alarms (https://docs.aws.amazon.com/directconnect/latest/UserGuide/monitoring-cloudwatch.html) and AWS CloudTrail Log Monitoring is available for AWS Direct Connect and Site-to-Site VPN (https://docs.aws.amazon.com/vpn/latest/s2svpn/monitoring-cloudwatch-vpn.html) connections back to your on-premises environment.

# WorkSpaces service networking controls

While the Amazon VPC service and associated services provide a number of networking capabilities, the Amazon WorkSpaces service also provides a native networking capability that can control access to the streaming interface (eth1) of your WorkSpaces environment from the internet.

## IP access control groups

An IP access control group acts as a virtual firewall that controls the IP addresses from which users are allowed to access their WorkSpaces. You can associate each IP access control group with one or more directories. You can create up to 100 IP access control groups per AWS account. However, you can only associate up to 25 IP access control groups with a single directory.

A default IP access control group is associated with each directory and enables all traffic. To specify the IP addresses and ranges of IP addresses for your trusted networks, add rules to your IP access control groups. If your users access their

WorkSpaces through a NAT firewall or VPN, you must create rules that allow traffic from the IP addresses for the NAT firewall or VPN.

Amazon WorkSpaces Design Considerations:

- **All IP address ranges from where clients connect from must be known**. By enabling an IP access control group and associating a set of IP addresses, users that have an IP address outside this range will be prevented from connecting. Therefore, if users are permitted to connect from their own homes, public Wi-Fi etc. the use of this feature may not be practical since the range of IP addresses to define in the group will be unknown. However, if users are not permitted to connect from home and are to be restricted to connecting from a well-known set of IP addresses associated with a small set of physical locations this feature can be very powerful in preventing access to your WorkSpaces environment from untrusted networks.
- **PCoIP Connection Manager (see here) cannot be used**. The use of this feature does not work with the PCoIP Connection Manager.

> **Note:** There is no equivalent if you are using the WSP protocol.

Further information regarding the use of IP Access Control Groups for your WorkSpaces environment can be found here:

https://docs.aws.amazon.com/workspaces/latest/adminguide/amazon-workspaces-ip-access-control-groups.html

# Other considerations

Amazon WorkSpaces can host both Linux and Windows-based WorkSpaces. Therefore, if either of these types are being used then it is possible to further impose some network security controls to tighten the security posture of the environment. If both types are employed, then the segregation of the two types into separate subnets and the application of a distinct security group for Windows and another security group for Linux WorkSpaces instances should be considered. This is because the networking and therefore port and IP address connectivity requirements between the two operating systems are distinct and different from each other.

# Conclusion

Throughout this document, guidance has been provided from different perspectives around the best practices for setting up VPCs and networking for Amazon WorkSpaces deployment. We have explored the networking requirements for an individual WorkSpace, VPC designs for different scenarios, the physical, and logical networking components of AWS VPCs. Different requirements will guide the design choices that need to be made within your WorkSpaces deployment. The guidance within this document can be applied to your WorkSpaces environment to provide users with a consistent, reliable and secure end-user experience.

Finally, it is worth considering contacting AWS for assistance with your VPC design because while this document outlines best practices for WorkSpaces, the addition and use of other AWS services will influence the design and will need to be factored into a wider VPC architecture design.

# Appendix A: Table of design considerations and document locations

| Design Area | AWS Consideration | Description |
|---|---|---|
| Physical Location | Region | AWS Regions provide multiple, physically separated and isolated Availability Zones which are connected with low latency, high throughput, and highly redundant networking. |
| | Availability Zones | Availability Zones offer AWS customers an easier and more effective way to design and operate applications and databases, making them more highly available, fault tolerant, and scalable than traditional single data center infrastructures or multi-datacenter infrastructures. |
| **External Connectivity to Amazon WorkSpaces** | AWS Direct Connect | Establishes a dedicated network connection from your premises to AWS. |
| | VPN | Establishes a secure and private tunnel from your network or device to the AWS global network. |
| **Internet Connectivity from Amazon WorkSpaces** | Internet gateway | Internet gateway is a horizontally scaled, redundant, and highly available VPC component that allows communication between instances in your VPC and the internet. |
| | NAT Gateway | A network address translation (NAT) gateway to enable instances in a private subnet to connect to the internet or other AWS services, but prevent the internet from initiating a connection with those instances. |

| Design Area | AWS Consideration | Description |
|---|---|---|
| **Internal Connectivity within Amazon WorkSpaces environment** | VPC | A Virtual Private Cloud (Amazon VPC) enables you to launch AWS resources into a virtual network that you've defined. |
| | VPC Peering | A VPC peering connection is a networking connection between two VPCs that enables you to route traffic between them privately. |
| | AWS Transit Gateway | AWS Transit Gateway is a service that enables customers to connect their Amazon Virtual Private Clouds (VPCs) and their on-premises networks to a |
| | Subnets | A new subnet in your VPC requires an IPv4 CIDR block to be specified from the range associated with the subnet's VPC. The location of the subnet within an Availability Zone must also be defined. |
| | Route Table | A route table contains a set of rules, called routes. These are used to determine where network traffic is directed. |
| | Network ACLs | A network access control list (ACL) is an optional layer of security for your VPC that acts as a firewall for controlling traffic in and out of one or more subnets. |
| Instance | Security groups | A security group acts as a virtual firewall for your instance to control inbound and outbound traffic. |

# Author


- Mark Homer, Senior Architect, Amazon Web Services


# Contributors

Contributors to this document include:

- Ivan Levchenko, Senior Consultant, Amazon Web Services
- Perry Wald, Senior Specialist Solution Architect, Amazon Web Services
- Dan Garibay, Specialist Technical Account Manager, Amazon Web Services
- Brett Looney, Global Solution Architect, Amazon Web Services
- Payoj Mistry, Technical Account Manager, Amazon Web Services
- Navi Magee, Senior Specialist Solution Architect, Amazon Web Services
- Andrew Wood, Senior Specialist Solution Architect, Amazon Web Services

# Document revisions

| Date | Description |
|---|---|
| **July 2020** | Initial publication |