Zero Trust Security for Azure Workloads with Zscaler Cloud Connector

Reference Architecture

Contents

©2023 Zscaler, Inc. All rights reserved.

About Zscaler Reference Architectures Guides

The Zscaler™ Reference Architecture series delivers best practices based on real-world deployments. The recommendations in this series were developed by Zscaler's transformation experts from across the company.

Each guide steers you through the architecture process and provides technical deep dives into specific platform functionality and integrations.

The Zscaler Reference Architecture series is designed to be modular. Each guide shows you how to configure a different aspect of the platform. You can use only the guides that you need to meet your specific policy goals.

Who Is This Guide For?

The Overview portion of this guide is suitable for all audiences. It provides a brief refresher on the platform features and integrations being covered. A summary of the design follows, along with a consolidated summary of recommendations.

The rest of the document is written with a technical reader in mind, covering detailed information on the recommendations and the architecture process. For configuration steps, we provide links to the appropriate Zscaler Help site articles or configuration steps on integration partner sites.

A Note for Federal Cloud Customers

This series assumes you are a Zscaler public cloud customer. If you are a Federal Cloud user, please check with your Zscaler account team on feature availability and configuration requirements.

Conventions Used in This Guide

The product name ZIA Service Edge is used as a reference to the following Zscaler products: ZIA Public Service Edge, ZIA Private Service Edge, and ZIA Virtual Service Edge. Any reference to ZIA Service Edge means that the features and functions being discussed are applicable to all three products. Similarly, ZPA Service Edge is used to represent ZPA Public Service Edge and ZPA Private Service Edge where the discussion applies to both products.

Notes call out important information that you need to complete your design and implementation.

Warnings indicate that a configuration could be risky. Read the warnings carefully and exercise caution before making your configuration changes.

Finding Out More

You can find our guides on the Zscaler website at **Reference Architectures** (**https://www.zscaler.com/resources/reference-architectures**).

You can join our user and partner community and get answers to your questions in the **Zenith Community** (**https://community.zscaler.com/**).

©2023 Zscaler, Inc. All rights reserved.

Terms and Acronyms Used in This Guide

Acronym	Definition
ACL	Access Control Lists
AWS	Amazon Web Services
AZ	availability zone
CA	Certificate Authority
DLP	Data Loss Prevention
DTLS	Datagram Transport Layer Security
IaaS	Infrastructure as a Service
IPS	Intrusion Prevention System
LSS	Log Streaming Service
MITM	Man-in-the-Middle
NIST	National Institute of Standards and Technology
NSS	Nanolog Streaming Service
PaaS	Platform as a Service
SaaS	Software as a Service
SIEM	Security Information and Event Management
SSL	Secure Sockets Layer
TLS	Transport Layer Security
VM	Virtual Machine
VNets	Virtual Networks
VPN	Virtual Private Network
ZIA	Zscaler Internet Access
ZPA	Zscaler Private Access
ZTE	Zero Trust Exchange

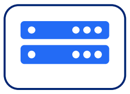
Icons Used in This Guide

The following icons are used in the diagrams contained in this guide.

Zscaler Zero Trust Exchange

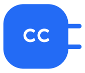
Zero Trust
Exchange

ZIA or ZPA Service Edge



Zscaler App Connector

Zscaler Cloud Connector



Azure Load Balancer

Azure Application Gateway

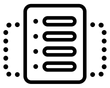
Azure Virtual Machine

Azure Application or Workload

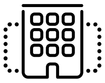
AWS Application or Workload

Generic Application or Workload

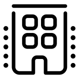
Private Data Center Location



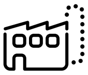
Headquarters Office Location



Branch Office Location



Factory Location



Authorized Use

Bad Actor

Data Tunnel

©2023 Zscaler, Inc. All rights reserved.

Introduction

The shift to cloud services has rebuilt the enterprise data center off-premises and outside of traditional security boundaries. Infrastructure as a Service (IaaS), Platform as a Service (PaaS), and Software as a Service (SaaS) enable organizations to quickly build out and scale their platforms and services. Securing services across multiple clouds, vendors, and support features requires a different approach than that of the traditional data center.

Securing this communication through the layering of legacy Access Control Lists (ACL), on-premises firewalls, and service-chaining has always been both complicated to build and difficult to maintain. Private applications were accessed via Virtual Private Networks (VPNs) that extended the network to locations in an any-to-any access model. This large, flat network gave users a single location to connect to for access to private applications.

Leveraging the cloud breaks these models. You now have multiple vendors across different clouds, products, and services. Your policy must be interpreted at each cloud and application to determine how best to implement it with the tools available. This risk goes up given a mistake, potentially exposing your organization to a host of network-born attack vectors.

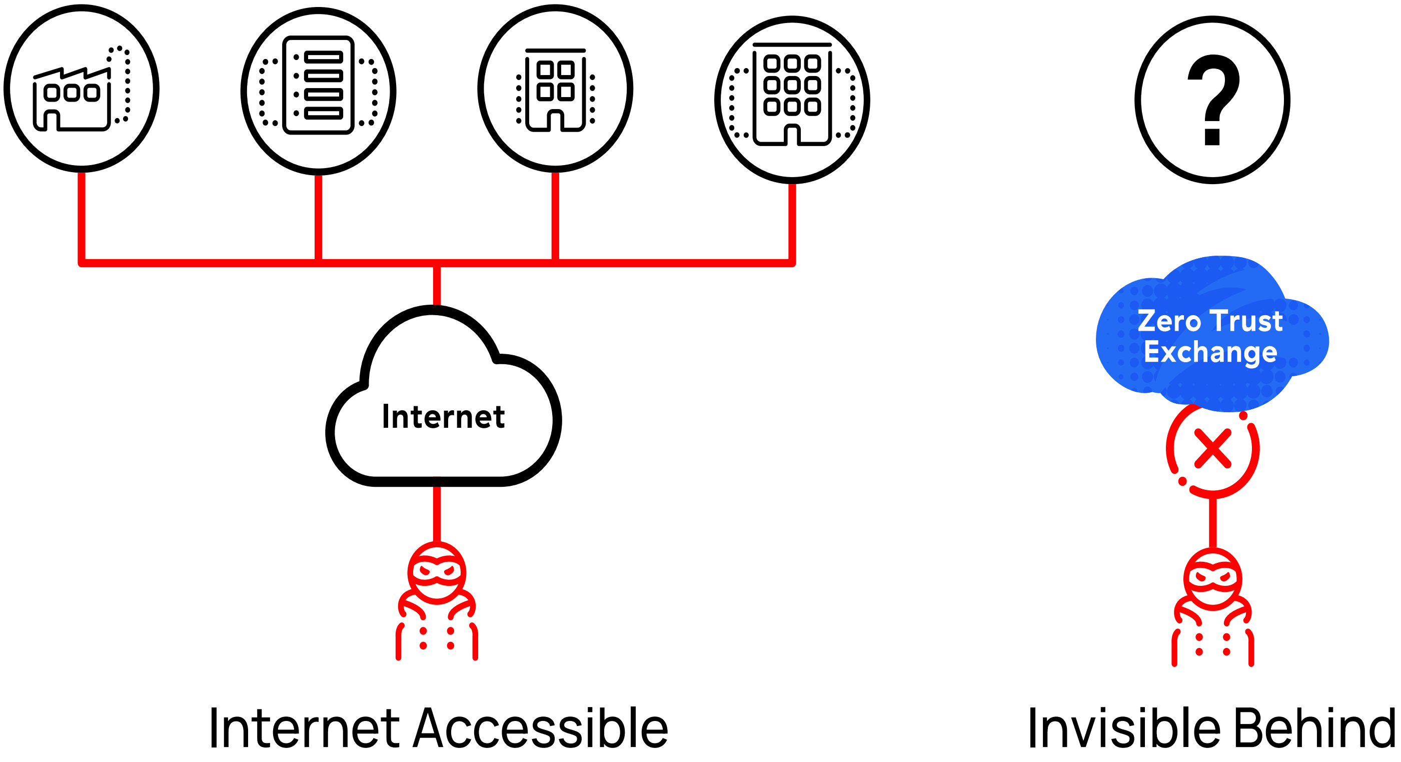


Figure 1. Zero trust moves from exposed network security to user-centric policy enforcement

Ideally, an organization's security policy should be at the foundation of its network design. Connectivity to and from devices happens as a product of the security policy and not the other way around. This is the heart of the Zscaler Zero Trust Exchange (ZTE) model. Users must be authorized before they can connect to that service. Even knowing the application's hostname and the services it provides won't give the attacker any information, as that service won't resolve until the user authenticates. Your applications are effectively hidden from the internet and each other until you define policy to allow access.

©2023 Zscaler, Inc. All rights reserved.

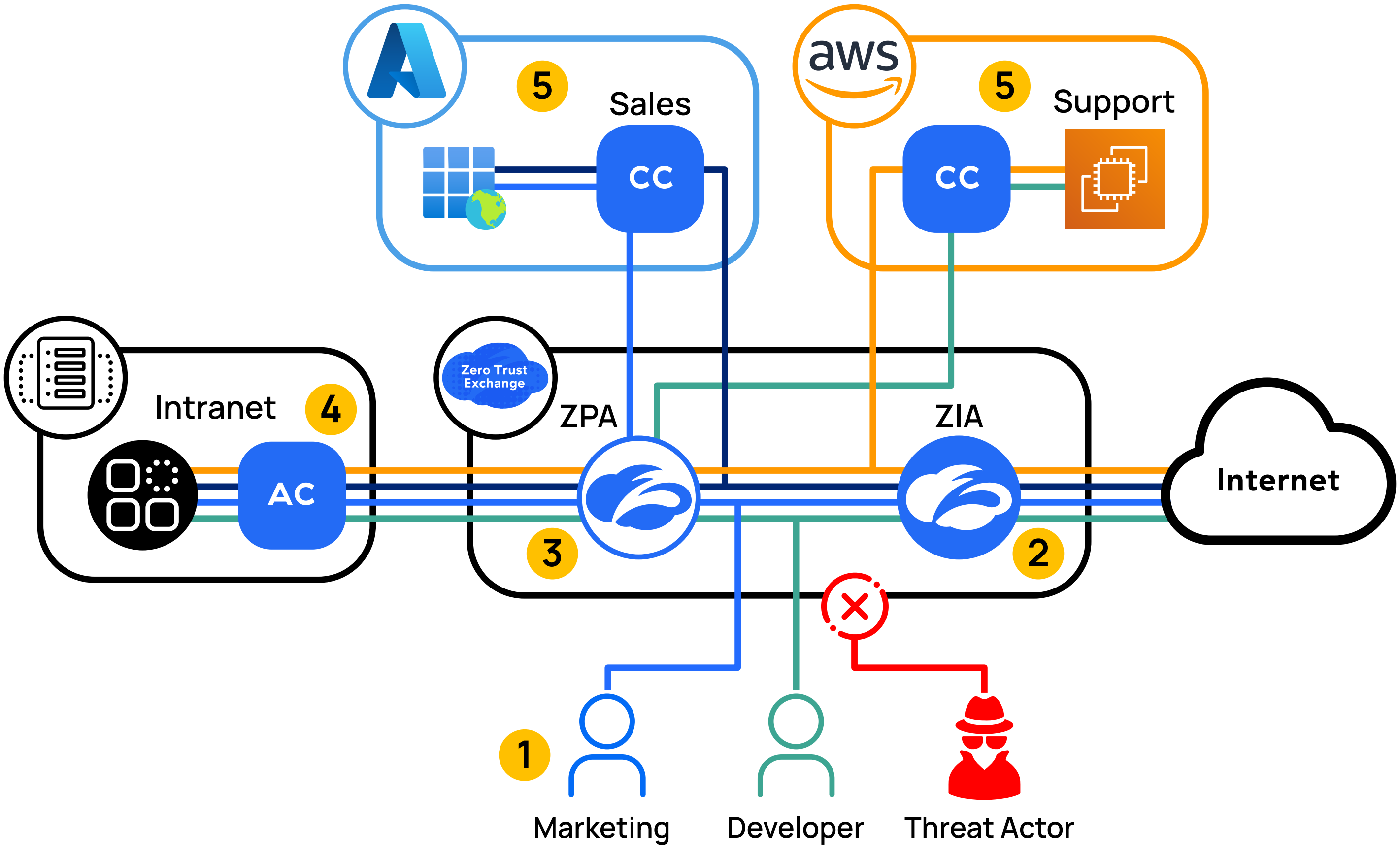


Figure 2. Zero trust principles applied to users and workloads

1. **Authentication** – All users must first authenticate to Zscaler. Based on multiple criteria such as user group membership, device posture, and location, the user is assigned a set of policies. These include the ability to see internal applications.
2. **ZIA Service Edge** – When traffic from users or workloads needs to be routed to the internet, a ZIA Service Edge inspects the traffic. If your policy allows the traffic out, the return traffic is also scanned for malicious content on its way back to the user.
3. **ZPA Service Edge** – Traffic from users or workloads bound for other internal applications is handled by ZPA. Based on the user's authentication and assigned policy, only approved resources are resolved. All other resources are hidden from unauthorized users as if the services do not exist.
4. **App Connector** – Sitting in front of internal applications, App Connector allows ZPA connections to applications for authorized users.
5. **Cloud Connector** – Deployed in front of your internal applications, Cloud Connector creates a set of outbound tunnels to ZIA and ZPA. They decide where the tunnel connects based on policy.

In the previous model, your users in marketing have workloads running in Microsoft Azure, and your developers have workloads in Amazon Web Services (AWS). All users have access to internal applications in the data center, as well as general internet access. In this case, each user and workload are limited to which applications they can resolve and access, based on the policy applied to them.

Our marketing user (blue) can access their workloads in Azure, the data center, and the internet based on policy. Your developer (green) can access their workloads in AWS, the data center, and the internet. Finally, your workloads in Azure, AWS, or your data center can all reach one another and the internet via the ZTE, without the need to set up additional VPN links. All these connections are subject to the policy you set.

Cloud Connector ensures that cloud workloads adhere to organizational security policy when accessing both public and private endpoints. This is achieved by intelligently forwarding traffic to the Zscaler Internet Access (ZIA) and Zscaler Private Access (ZPA) platforms. Cloud Connector also enables multi-cloud connectivity and enforces a security policy for cloud-to-cloud traffic.

Key Features and Benefits

- Reduce complexity by connecting directly to the internet and eliminate the need for complex routing configurations through SNAT, transit gateways, and transit hubs.
- Total visibility, control, and awareness for workload communications. Centralized logging and real-time streaming can also be used with third-party monitoring solutions.
- Flexible scalability with elastic, horizontal scaling made possible through the Zero Trust Exchange architecture, which operates in over 150 global data centers.
- High availability with built-in automatic failover with N+2 redundancy is provided for forwarding and security.
- Lower operational costs by removing expenses associated with complex network configurations, network service replication, and hidden costs for cloud connectivity.

New to Zscaler Cloud Connector or Microsoft Azure?

If this is your first time reading about Zscaler Cloud Connector, we encourage you to explore the following resources:

- If you are new to Azure, an Azure Fundamentals course is available at **Microsoft Azure Fundamentals: Describe core Azure concepts** (**https://docs.microsoft.com/en-us/learn/paths/az-900-describe-cloud-concepts/**).
- To watch a demonstration of Zscaler Cloud Connector, see **Zero Trust Your Cloud Workloads** (**https://www.youtube.com/watch?v=S-g_qmuxnqU&t=1845s**).
- Zscaler Internet Access (ZIA) provides outbound internet protection for users. Learn more at **Zscaler Internet Access** (**https://www.zscaler.com/products/zscaler-internet-access**).
- Zscaler Private Access (ZPA) provides private access to applications, not networks. Learn more at **Zscaler Private Access** (**https://www.zscaler.com/products/zscaler-private-access**).
- To learn more about zero trust, see **It Starts With Zero** (**https://www.zscaler.com/it-starts-with-zero**).
- To learn more about the zero trust architecture, we recommend the National Institute of Standards and Technology (NIST) paper **Zero Trust Architecture** (**https://www.nist.gov/publications/zero-trust-architecture**).

Cloud Infrastructure Protection Using Cloud Connector

As organizations began moving workloads to the cloud, securing those resources has always been a challenge. Securing applications against attack both from outside actors and malicious content on legitimate sites led some organizations to provide access to applications over VPN. Attempting to leverage legacy security products adds latency and frustration for your users. As cloud usage expanded, data now moves between the cloud and the user, between systems in the cloud vendor, and between applications across cloud vendors and your legacy data center.

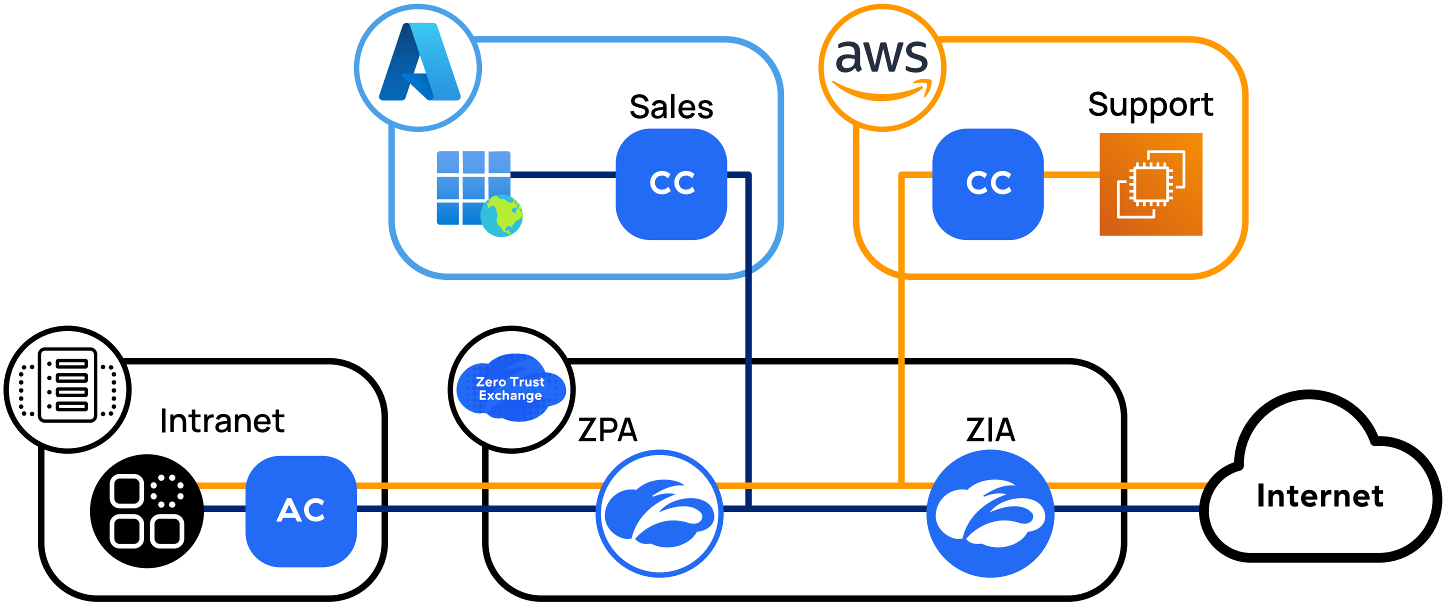


Figure 3. Workload communication between private and public applications

The communication can be from private workloads (IaaS or physical DC) to public workloads (SaaS internet application) or between private workloads (IaaS to IaaS, or physical DC to IaaS). Securing these communications channels with physical or virtual appliances is cumbersome and can lead to inconsistent configuration.

In the previous example, our application sales.azure.internal.safemarch.com sits behind a Cloud Connector with access to both ZPA and ZIA platforms. In this model, the workload can reach out to the support workloads in AWS, allowing the sales team to file support and product requests without logging into the support portal. The sales portal is accessed by our intranet workload in our data center to pull deals and rankings for the organization's dashboard. Finally, our sales workload can reach the internet to update our cloud CRM, which in turn only accepts connections from Zscaler IPs for our tenant.

©2023 Zscaler, Inc. All rights reserved.

Zscaler Cloud Connector virtual machines extend the security of ZIA and ZPA to cloud native workloads. ZIA protects your workload traffic communicating with a public application. ZPA protects your communications between private workloads. This allows organizations to secure all workload communications over any network. The Zscaler Zero Trust Exchange allows workloads to communicate with each other with a granular security policy applied.

- Applications-to-Internet Communications for applications that might need to access any internet or SaaS destination, such as third-party APIs, software updates, etc. A scalable, reliable security solution that inspects all transactions and applies advanced threat prevention and data loss protection controls.
- Application-to-Application Communication to other public clouds and corporate data centers for multi/hybrid cloud connectivity. Delivered with better security and a dramatically simplified operational model, as compared with traditional solutions like proxies, virtual firewalls, and IDS/IPS.
- Application-to-Application Communications within a Virtual Private Cloud by securing process-to-process communications. This achieves microsegmentation of traffic with no changes to the application or the network.

Cloud Connector is delivered in several form factors. It is available as a virtual appliance in both Amazon Web Services and Microsoft Azure, as well as VMs for on-premises deployment.

If you are deploying on Microsoft Azure:

- Zscaler recommends the *Standard D2s v3* instance size to support Cloud Connector as it offers the best performance of 300 Mbps (unidirectional).
- The appliance is available on the Azure Marketplace at **Zscaler Cloud Connector Application** (**https://azuremarketplace.microsoft.com/en-us/marketplace/apps/zscaler1579058425289.zia_cloud_connector_app**).

For on-premises deployments, the image requires:

- VMware ESXi and CentOS/Linux (KVM) images
- 2 virtual CPUs
- 4 GB of RAM

Deploying Cloud Connector VMs via Scripts

Cloud Connector can be deployed in AWS, leveraging scripts to simplify deployments and ensure consistency. Zscaler supports two scripting methods for deploying Cloud Connector on Azure marketplace: leveraging Azure Resource Manager (ARM) templates or Terraform scripts. ARM templates allow a user to deploy the appliance directly from the Azure Marketplace via guided workflow, are user-friendly, and work well with existing brownfield deployments. Terraform is the most flexible option. Its goal is to be as "hands-off" as possible by automatically configuring items without user intervention. However, Terraform is more complex in its initial setup. Both options allow you to automate your deployment and achieve the same results. For a detailed look at these deployments, see **Deploying Cloud Connector via Terraform Scripts** and **Deploying Cloud Connector via Marketplace Application** later in this guide.

High Availability Deployment Design

Cloud Connector leverages Azure Load Balancer functionality to achieve high availability and horizontal scalability. In this model, inbound traffic from workload Virtual Networks (VNets) are directed to the front-end IP address of the load balancer. When traffic returns from the internet, the Cloud Connector appliance strips off Datagram Transport Layer Security (DTLS) encapsulation and forwards the traffic back to the originating workload VNet.

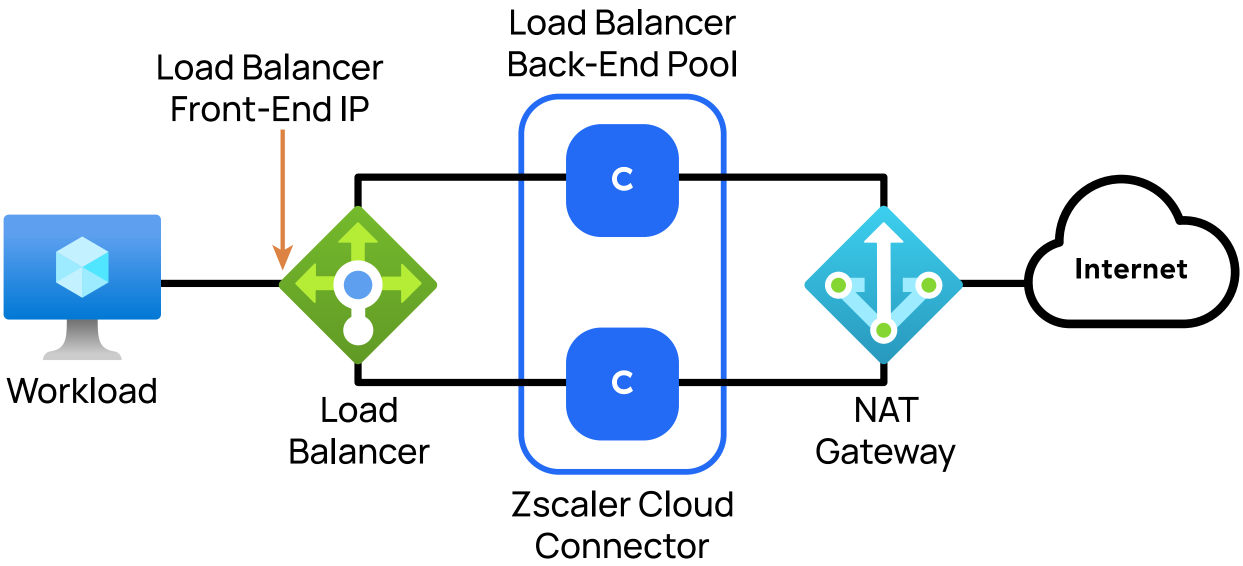


Figure 4. Cloud Connectors receive outbound traffic from the load balancer

Zscaler recommends a minimum of two Cloud Connector appliances, each in a different availability zone (AZ). Workloads within those same availability zones should then leverage their respective Cloud Connector appliances. If a Cloud Connector appliance fails, load balancer functionality automatically redirects traffic to the active appliance in the adjacent AZ.

©2023 Zscaler, Inc. All rights reserved.

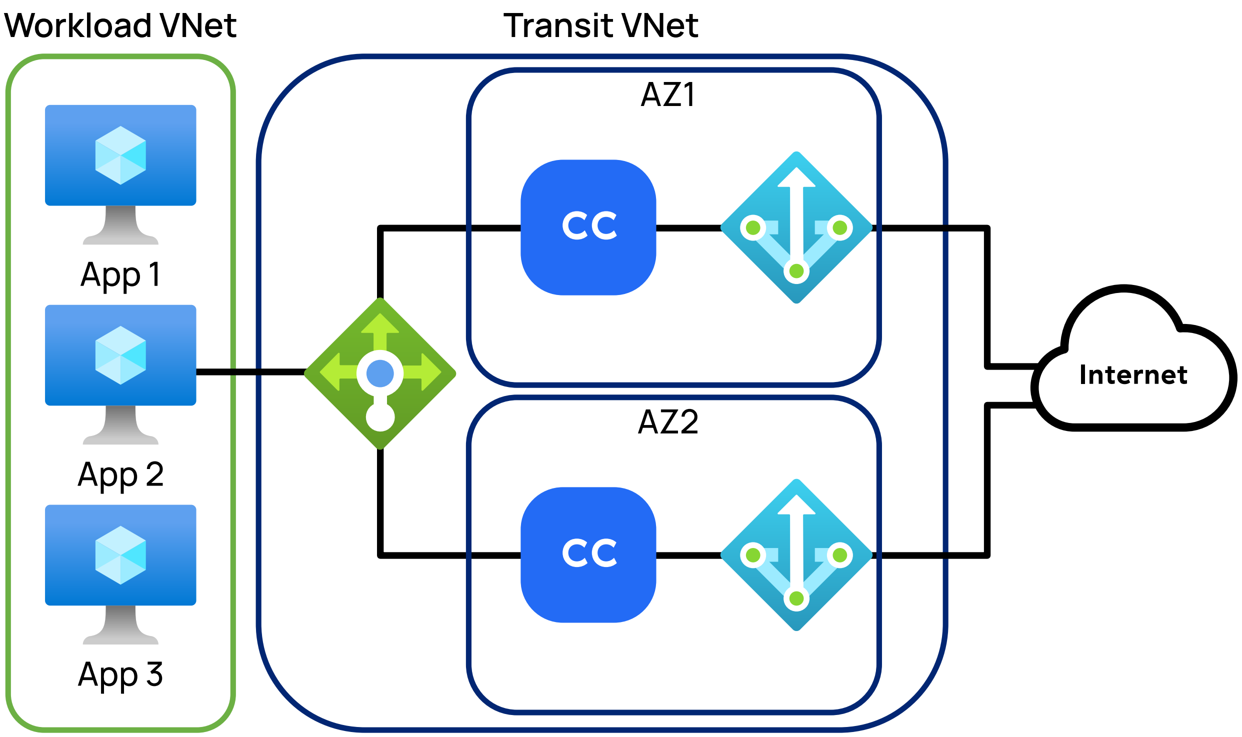


Figure 5. Cloud Connectors deployed in redundant pairs across availability zones

By default, Azure Load Balancer uses a 2-tuple hash (source and destination IP address) to balance traffic. This ensures that traffic from a workload communicating with another resource always uses the same Cloud Connector instance.

Azure Load Balancer, by default, probes the HTTP Probe Port every 15 seconds. Two probes must fail before considering an appliance down (for a total outage of 30 seconds). However, Microsoft Azure allows tuning down to 5 seconds with two failed attempts as necessary (for a total outage of 10 seconds).

In addition to the appliance-level redundancy, Cloud Connector also maintains redundant DTLS tunnels to the Zscaler cloud. Primary and secondary/backup nodes can be set within the Cloud Connector portal for ZIA, or left as automatic (as is the case with ZPA), wherein the Cloud Connector chooses geographically proximate brokers to connect to.

Terraform scripts can be used to automate deployments, or a script can be built manually. Zscaler provides a Terraform template for your use at **About Cloud Automation Scripts** (**https://help.zscaler.com/cloud-connector/about-cloud-automation-scripts**).

Learn more at **What Is Azure Load Balancer?** (**https://docs.microsoft.com/en-us/azure/load-balancer/load-balancer-overview**).

©2023 Zscaler, Inc. All rights reserved.

Scalability of Cloud Connector Instances

Cloud Connector supports two methods of scaling: vertical and horizontal. With vertical scaling, the Cloud Connector can be deployed with a higher footprint of vCPU and RAM. However, throughput and connection capacity scales linearly with additional resources. You will eventually hit the maximum throughput limit for the appliance. Additionally, the failure of a larger appliance requires more connections to fail over to the backup solution, which faces the same throughput restrictions.

The following image is simplified for clarity. Redundant instances of Zscaler Cloud Connector should be deployed in all instances.

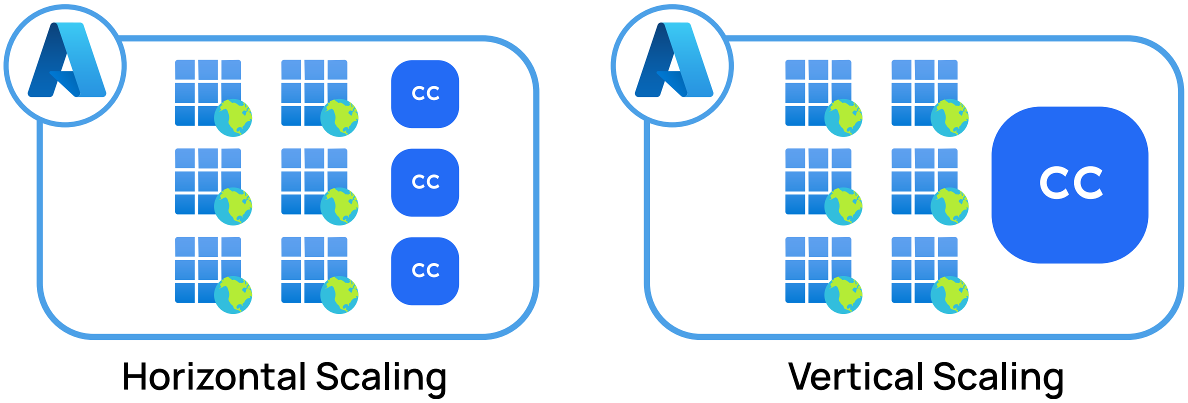


Figure 6. Horizontal and vertical scaling of Cloud Connectors

Cloud Connector can also be scaled horizontally, wherein multiple appliances are deployed within multiple availability zones around a region. Inbound traffic to the Cloud Connector appliance can then be load-balanced across all available paths. Either or both methods are supported when considering current and future throughput requirements. Typically, horizontal scaling is more scalable and fault-tolerant, and avoids any cloud provider platform limits.

Cloud Connector Logging and Service Dashboards

Cloud Connector can use built-in logging functionality through the Insights page of the portal. Zscaler streams all logs to centralized log locations, allowing you to view logs from across your organization. The dashboard has views for Session Insights, DNS Insights, and ZIA Tunnel Insights. All three facilities allow you to review traffic that passes through the Cloud Connector appliance from a different perspective.

- Learn more about Analyzing Traffic Using Insights (https://help.zscaler.com/cloud-connector/analyzing-traffic-using-insights).
- Learn more about ZIA Dashboards (https://help.zscaler.com/zia/about-dashboards).
- Learn more about ZPA Dashboard and Diagnostics (https://help.zscaler.com/zpa/dashboard-diagnostics).

Cloud Connector supports both the Nanolog Streaming Service (NSS) for ZIA use cases and Log Streaming Service (LSS) for ZPA use cases. NSS uses a Virtual Machine (VM) to stream traffic logs in real time to your Security Information and Event Management (SIEM) system, such as Splunk or ArcSight. LSS operates in a similar way, with the deployment of a ZPA App Connector VM that receives the log stream and then forwards it to the log receiver.

Both services enable real-time alerting and correlation of logs with your other devices. NSS and LSS can be configured from the Cloud Connector portal.

NSS and LSS require separate subscriptions for each virtual machine.

- Learn more about Nanolog Streaming Service (https://help.zscaler.com/zia/about-nanolog-streaming-service).
- Learn more about Log Streaming Service (https://help.zscaler.com/zpa/about-log-streaming-service).

Upgrading Your Cloud Connectors

Cloud Connector runs the Zscaler OS in the virtual machine. Software updates and OS updates are provided by Zscaler via automatic upgrades. When a Cloud Connector is deployed, the software is automatically upgraded to the latest version. Cloud Connector instances check for new software daily. If a new version is available the Cloud Connector will upgrade itself automatically at midnight local time, based on the deployed cloud region.

This automatic check and update means it is critical that your Cloud Connector locations are accurate. An inaccurate location can lead to upgrades occurring in the middle of the day. Always specify exactly where the Cloud Connector is located when deploying the virtual machine.

As a matter of redundancy during upgrades, Cloud Connector is installed in pairs within an availability zone. Multiple pairs of Cloud Connectors should be instantiated within different availability zones, thereby minimizing the impact of service upgrades or infrastructure failures.

Cloud Connector is based on the Zscaler OS, and therefore the software updates and OS updates are provided and automatically applied by Zscaler. When a Cloud Connector is deployed, the software is automatically updated to the latest version. Cloud Connector then checks for new software daily and upgrades itself automatically at midnight (local time, based on the deployed cloud region).

You can configure this upgrade window from the Cloud Connector Portal. As mentioned throughout this document, Zscaler recommends that Cloud Connector appliances be deployed as redundant, high-availability instances. Specifically, we recommend deploying two appliances per availability zone with a minimum regional cluster size of four (two in AZ1 and two in AZ2). The Zscaler software upgrade process upgrades one instance of a pair at a time, providing availability for the AZ from the remaining instance.

Zscaler recommends that Cloud Connector appliances be deployed as redundant, high-availability appliances. Specific to software upgrades performed by Zscaler, this ensures that you incur no downtime. When an appliance is rebooted to accept a new update, Azure Load Balancer automatically moves traffic over to the redundant, active appliance.

Although cloud IaaS providers such as Azure are responsible for ensuring the security and availability of their infrastructure, organizations are ultimately still responsible for the security of their workloads, applications, and data. To learn more about the shared responsibility model, see **Shared responsibility in the cloud** (**https://docs.microsoft.com/en-us/azure/security/fundamentals/shared-responsibility**).

©2023 Zscaler, Inc. All rights reserved.

Deployment and Design Options

The following section outlines your options when deploying Cloud Connector. You can design your network using the tools that best match your cloud deployment. We recommend that you review each use case to familiarize yourself with the various options, which can be combined to meet your organization's deployment needs. For example, in many production environments Cloud Connector would be deployed in a Transit/Egress VNet to perform outbound internet *and* Zscaler Private Access. Although the use cases do not depend on one another, the concepts and logic depicted within them progressively build on one another.

Pre-Deployment Considerations

The following sections provide some general design recommendations common to all deployment types.

Cloud Connectors and Availability Zones

Zscaler recommends that Cloud Connector appliances be installed in pairs for high availability. When building high-availability pairs of Cloud Connector appliances, Zscaler recommends that each appliance be deployed in different availability zones within the same region. This ensures that individual Cloud Connector appliances exist on physically separate pieces of underlying hardware from one another and provide failover access.

To learn more about Azure availability zones, see **Regions and availability zones** (**https://docs.microsoft.com/en-us/azure/availability-zones/az-overview**).

Network Connectivity

Zscaler recommends employing NAT Gateways for internet access. Just like Cloud Connector, a NAT Gateway is also deployed in an availability zone. NAT Gateways can be shared across availability zones. Zscaler recommends that NAT Gateways also be operated in pairs with one gateway in each of the Cloud Connector availability zones.

To learn more about Azure NAT Gateway, see **Design virtual networks with NAT gateway** (**https://docs.microsoft.com/en-us/azure/virtual-network/nat-gateway/nat-gateway-resource**).

©2023 Zscaler, Inc. All rights reserved.

Deploying Cloud Connector via Terraform Scripts

Zscaler Terraform scripts provide complete end-to-end automation to not only deploy Cloud Connector appliances, but all the secondary and tertiary components as well in a repeatable and predictable way. Terraform scripts can be downloaded from the Cloud Connector portal in two versions:

- **Starter Deployment Template** – Deploy a Resource Group containing Cloud Connector appliance, VNet, Route Tables, Subnet, NAT Gateway, and Network Security Groups for use cases where only ZIA is required. In addition, Terraform also creates a VM instance for use as a Management/Bastion host in the VNet that Cloud Connector is deployed in. This host is not a requirement long term but is recommended for easier troubleshooting and testing.
- **Starter Deployment Template with Load Balancer** – Deploy Cloud Connectors in high-availability mode using Azure Load Balancer, along with required cloud constructs mentioned previously for use cases where high availability is a requirement. In addition, Terraform also creates a VM instance for use as a Management/Bastion host in the VNet that Cloud Connector is deployed in. This host is not a requirement long term, but is recommended for easier troubleshooting and testing. Learn more about **Azure Load Balancer** (**https://azure.microsoft.com/en-us/services/load-balancer/**).

It is important to note that Terraform does not modify brownfield deployments. When executing Terraform scripts, new VNets, Route Tables, Subnets, and VM instances are spawned to support the current workflow. It is your responsibility to integrate the new deployment into your existing environment. This can mean that the new Cloud Connector VNet is peered with existing VNets, or that new workloads are installed within the Cloud Connector VNet. Bear this in mind when considering whether Terraform is the correct option to use when integrating with a brownfield environment.

For detailed deployment instructions and to find the templates listed previously, see **About Cloud Automation Scripts** (**https://help.zscaler.com/cloud-connector/about-cloud-automation-scripts**).

Deploying Cloud Connector via Marketplace Application

A more native automation option for deploying Cloud Connectors, Zscaler offers an Azure Marketplace application. This application is built using Azure Resource Manager (ARM) templates. The ARM template is a JSON file that defines infrastructure and configuration for your devices.

Though the Marketplace application can be used in greenfield situations, its value shines when you are deploying in a brownfield deployment, as many items you would build out using Terraform likely already exist in your current Azure buildout. Unlike the Terraform scripts where an entire network is deployed, Marketplace Applications build only the application itself.

If you want to deploy Cloud Connector into existing infrastructure, the cloud Marketplace might be the correct choice. The Marketplace will prompt you for your account information, network settings, and Zscaler account information. After you complete the walkthrough, the system proceeds to build and deploy your Cloud Connectors based on the inputs.

If you have a small number of Cloud Connector deployments, the Azure Marketplace route might be sufficient for your organization. If you plan to build out more robust infrastructure routinely, Terraform scripts might be a better option.

Learn more about **Zscaler Cloud Connector on the Azure Marketplace** (**https://azuremarketplace.microsoft.com/en-us/marketplace/apps/zscaler1579058425289.zia_cloud_connector?tab=overview**).

Learn more about **ARM templates** (**https://docs.microsoft.com/en-us/azure/azure-resource-manager/templates/overview**).

©2023 Zscaler, Inc. All rights reserved.

Directing Traffic to Cloud Connector

Cloud Connector acts as a gateway to cloud workloads. Directing traffic through the Cloud Connector is as simple as modifying the default gateway route of the workload route table to point to the appliance, or to the Azure Load Balancer IP. In most circumstances, this ensures that both internet-bound traffic destined for ZIA, and DNS traffic that requires modification for **ZPA Use Cases** where redirection to an App Connector is necessary, are appropriately handled.

For example, with a single instance of Cloud Connector the workload route table can be updated with a default route using the IP address of the service interface of the Cloud Connector appliance as the target. The Cloud Connector appliance uses the service subnet and route table created during the deployment process. The default route for this route table should point towards the NAT Gateway, also created in the deployment process. A public subnet and route table should have also been created during the deployment process and reference the corresponding internet gateway with its default route.

The following image is simplified for clarity. Redundant instances of Zscaler Cloud Connector should be deployed in all instances.

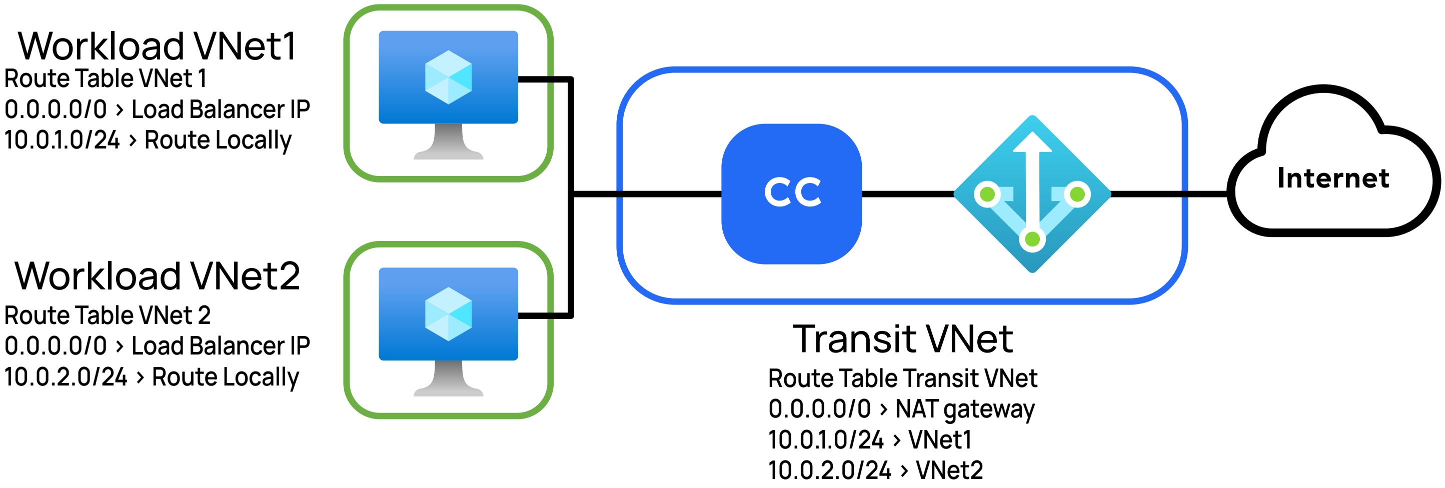


Figure 7. Default routes for workloads go through the Cloud Connector

In the case of hub and spoke, wherein the Transit/Egress VNet is the “hub” and workload VNets are the “spokes,” the Transit VNet should be peered with all workload VNets. A default route in the service subnet route table of the Cloud Connector appliance directs traffic towards the NAT Gateway by default. In the workload VNet, a default route is present to direct traffic across the VNet peering towards the Cloud Connector appliance. As with all network traffic, ensure you have routing set up as well so that returning traffic from the internet is correctly directed back towards the initiating host.

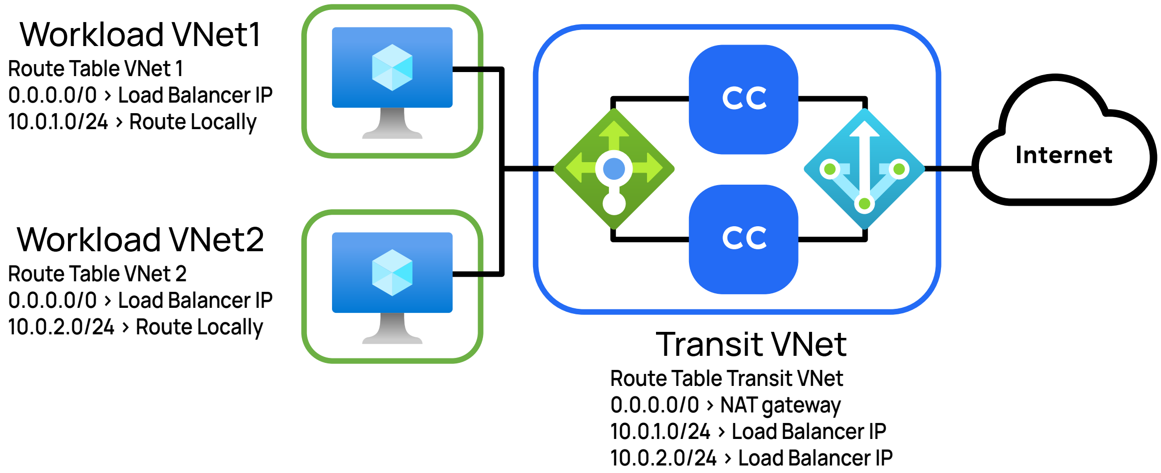

Workload VNet1
Route Table VNet 1
0.0.0.0/0 › Load Balancer IP
10.0.1.0/24 › Route Locally
Workload VNet2
Route Table VNet 2
0.0.0.0/0 › Load Balancer IP
10.0.2.0/24 › Route Locally
CC
CC
Internet
Transit VNet
Route Table Transit VNet
0.0.0.0/0 › NAT gateway
10.0.1.0/24 › Load Balancer IP
10.0.2.0/24 › Load Balancer IP

Figure 8. Transit VNets provide access to one or more private subnets

When the Azure Load Balancer is in use in high-availability use cases, Transit/Egress VNet route tables do not change. However, workload route tables are adjusted to use the load balancer front-end IP address as their default gateway.

Forwarding Options

When traffic has reached the Cloud Connector, there are four Traffic Forwarding options available to direct traffic out of the Azure cloud:

- **Direct** – Traffic matching the criteria defined bypasses the Cloud Connector and is routed out of the service interface, where it follows AWS route tables towards the destination.
- **Zscaler Internet Access (ZIA)** – Traffic matching the criteria defined is forwarded to the ZIA cloud for inspection.
- **Zscaler Private Access (ZPA)** – Traffic matching the criteria defined is forwarded to the ZPA cloud for inspection.
- **Drop** – Traffic matching the criteria is dropped by the Cloud Connector.

Each of the three options permits the administrator to define a range of match criteria. In general, macro forwarding logic can be defined within the Cloud Connector portal, whereas ZIA or ZPA can perform more granular inspection.

Traffic Forwarding policy is in the Policy Management section of the Cloud Connector portal. Rule creation and assessment models ZIA and ZPA workflows. More specific rules should be ordered near the top, while more broad rules ordered towards the bottom. Match criteria is as follows:

General

- **Location** – Locations identify the various VNets from which your workloads send traffic. As Cloud Connector appliances are brought online, the VNet they are installed within automatically populates this menu. It should be noted, however, that in a Transit/Egress VNet scenario, downstream VNets do not automatically populate. In such a case, you must use Source or Destination IP/FQDN as match criteria. In ZIA, if the traffic is from a known location, the service processes the traffic based on the location settings. For example, the service checks whether the location has authentication enabled and proceeds accordingly. It also applies any location policies that you configure and logs internet activity by location.
- **Location Group** – If necessary, location groups can be created to organize various cloud Vnets, such as a “Dev VNets” location group, “Prod VNets” location group, etc. If there are many locations and associated sub-locations within your organization, consider using location groups.
- **Branch and Cloud Connector Groups** – Branch and Cloud Connector groups allow you to match traffic transiting specific Cloud Connector appliances.

Source

- **Source IP Groups** – When multiple source IP addresses must be matched across multiple policy rules, it is operationally more efficient to create source IP groups. These groups allow you to organize IP addresses for easier rule creation and visualization.
- **Source IP Addresses** – This match criteria allows you to specify the source IP address of the workload.

Destination

- **Destination IP Address / FQDN** – For individual IP address/FQDN matching, enter the value you want to be matched in this field.
- **Destination IP / FQDN Group** – You can group together destination IP addresses and FQDNs that you want to control in a Forwarding Policy rule by specifying IP addresses, countries where servers are located, and URL categories.

Wildcard domain identifiers (“*”) are not currently supported.

- **Destination Country** – This match criteria allows you to specify the destination country of the remote machine.

Destination criteria is not supported when Zscaler Private Access is selected as the Forwarding Method.

After configuring a Forwarding Method and match criteria, you must choose an action. By default, for ZIA use cases, the Cloud Connector appliance uses geolocation to locate a ZIA Enforcement Node in geographic proximity to the appliance. Alternatively, you can manually specify which Enforcement Node to use by configuring a gateway under the Forwarding Methods section of the Administration menu.

Gateway selection criteria is not supported when Zscaler Private Access is selected as the Forwarding Method. Cloud Connector automatically selects a broker.

Lastly, specifically for ZPA use cases, Cloud Connector also allows for the filtering of DNS requests/responses. In the Administration menu within DNS Control, administrators can add additional rules to permit or deny specific DNS requests from workload segments. More importantly, this functionality can be used to determine which traffic gets consumed by ZPA, and therefore which synthetic IP Pool is used to address traffic within Microtunnels.

To view configuration instructions visit **Configuring Traffic Forwarding Rules** (**https://help.zscaler.com/cloud-branch-connector/configuring-traffic-forwarding-rule**).

©2023 Zscaler, Inc. All rights reserved.

Choosing the Correct Design Model

Cloud Connector is extremely flexible in the ways in which it can be deployed: directly adjacent to the workloads it services, or in a dedicated island by itself where traffic can be directed through it via Azure networking constructs like VNet Peering. There is no single design model that fits every environment. Many organizations pull elements from all design models to suit their goals. There are three main questions to ask when determining how best to get started:

Is ZPA a requirement?

Zscaler Private Access requires workload DNS queries to transit the Cloud Connector so a synthetic IP Address can be assigned to the connection. Consider how DNS is employed within the cloud. If using cloud-hosted DNS servers, it is possible that DNS resolution requests are never directed across the Cloud Connector which would break ZPA. For this reason, consider how DNS resolution requests transit Cloud Connector, such as if a public DNS server outside of the cloud is used (or if a custom DNS server is used that forwards these requests). Additionally, if this cloud implementation also services inbound requests from remote clouds, consider pointing App Connectors towards real DNS servers in this scenario.

Is high availability a requirement?

Zscaler recommends that high availability be employed in all use cases. However, when deploying directly into the workload VNet, compute costs can quickly spiral out of control, particularly if there are many VNets requiring appliance(s). For this reason, you might consider using a dedicated Transit/Egress VNet with VNet Peering. This allows you to maintain high availability without a large compute footprint.

Will Cloud Connector be deployed within the workload VNet, or in a dedicated VNet?

For small environments with only a handful of VNets, Cloud Connector instances can be deployed directly within the workload VNet. However, the number of VNets and VM instances tend to increase as an organization grows larger and invests further in the cloud. As new VNets are added, they require new appliances. As you consider where the Cloud Connector appliances will be installed, ensure you plan for adequate growth in the number of workloads and VNets that Cloud Connector protects. If the future state of the environment becomes operationally cumbersome, or if the environment already contains several VNets, it might be best to consider a Transit/Egress VNet approach for Cloud Connector.

©2023 Zscaler, Inc. All rights reserved.

Use Case: Direct to Internet Using Zscaler Internet Access

Implementing Cloud Connector to provide outbound internet access through ZIA is one of the first steps to cloud workload protection. The following deployment model represents a recommended option that can be leveraged to satisfy this business requirement and offer a foundation to build on when looking to implement services like ZPA.

In this model, Cloud Connectors can be installed directly into the workload VNet adjacent to the individual workloads they service. As with all deployment models, Zscaler highly recommends deploying Cloud Connector in high availability. The following image assumes redundant appliances are being deployed with an Azure Load Balancer:

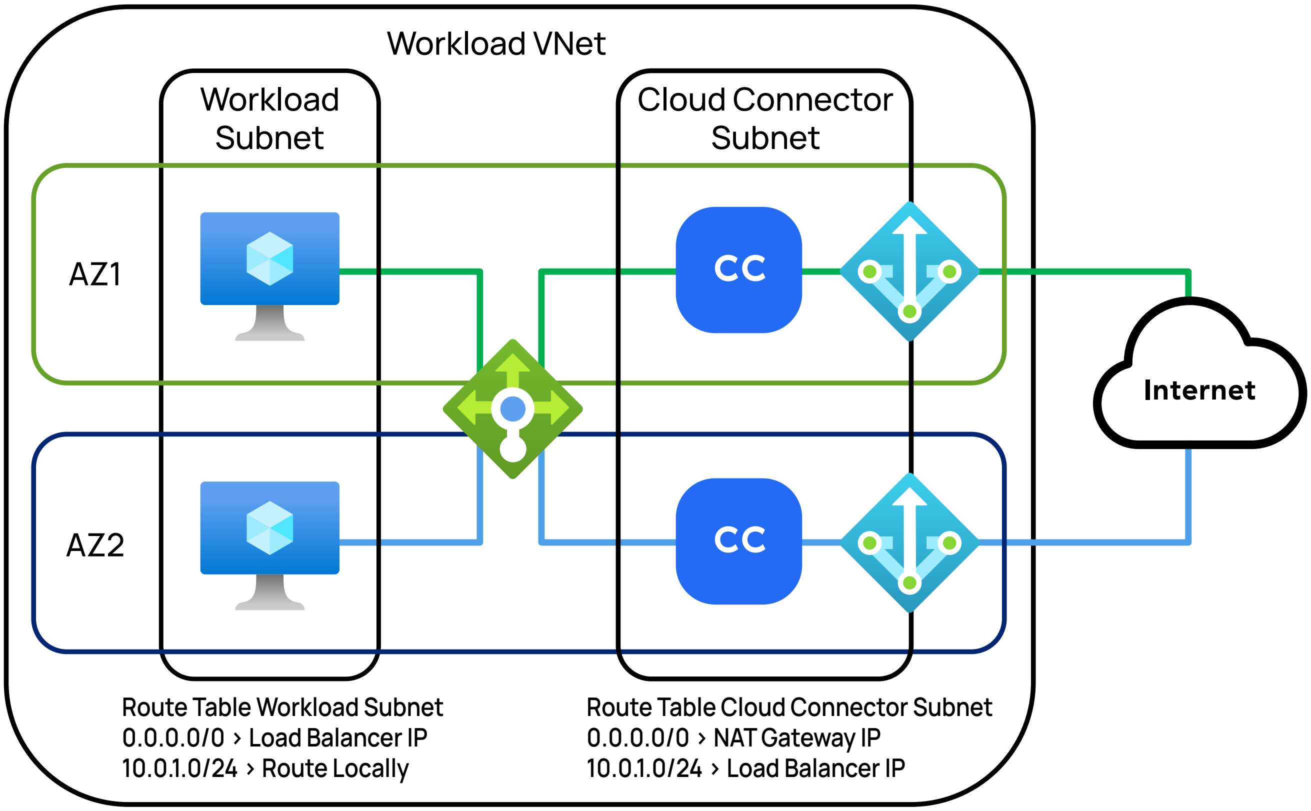

Workload VNet
Workload Subnet
Cloud Connector Subnet
AZ1
AZ2
CC
CC
Internet
Route Table Workload Subnet
0.0.0.0/0 › Load Balancer IP
10.0.1.0/24 › Route Locally
Route Table Cloud Connector Subnet
0.0.0.0/0 › NAT Gateway IP
10.0.1.0/24 › Load Balancer IP

Figure 9. Azure Load Balancer providing redundancy between Cloud Connectors

The primary benefit to this design option is its simplicity and time to implement. Since each Cloud Connector instance is spawned within the workload VNet that it services, routing is made simple. Likewise, whether via Terraform or Marketplace Application, Zscaler automation can implement this model in a matter of minutes. From a cost perspective, you are only paying for egressing data fees one time (as the workload traffic leaves the Cloud Connector), as opposed to the double billing that can occur when using a Transit/Egress VNet.

If you have many workload VNets, however, this design option can be cumbersome. Any cost savings associated with egress fees can be eliminated by the increased compute footprint, since separate Cloud Connector VM instances are required per workload VNet. Additionally, this option requires the modification of many route tables to direct traffic accordingly, which is further complicated when high availability enters the picture.

©2023 Zscaler, Inc. All rights reserved.

When implementing this design option, the first step is to consider which automation technique to employ, leveraging either Terraform scripts or Marketplace Application.

If using a Marketplace Application, consider deploying the HA Cloud Connector Application within a separate availability zone to deploy the Azure Load Balancer. It is recommended to deploy NAT Gateway. Since NAT Gateway(s) operate within a single availability zone, Zscaler recommends creating a second NAT Gateway in a different AZ so that an infrastructure failure of one AZ does not affect both NAT Gateways. Ensure that the Cloud Connector in the first AZ is in a different subnet than the Cloud Connector in the second AZ. Then, associate each NAT Gateway to each respective subnet.

For brownfield implementations, Marketplace Applications might provide more seamless integration.
For greenfield implementations, consider using Terraform.

Use Case: Leveraging VNet Peering

Cloud Connector can also be placed in a dedicated VNet wherein outbound workload traffic is directed through a centralized hub, such as a Transit VNet. Transit VNets with VNet Peerings is a design option that is growing in adoption as organizations seek to address scalability concerns and operational deficiencies imposed by deploying services directly within workload VNets.

This model closely resembles a traditional hub-and-spoke network since the hub Transit/Egress VNet, where Cloud Connector operates, receives traffic from many workload spoke VNets. As with all deployment models, Zscaler highly recommends deploying Cloud Connector in high availability. The following image assumes redundant appliances are being deployed:

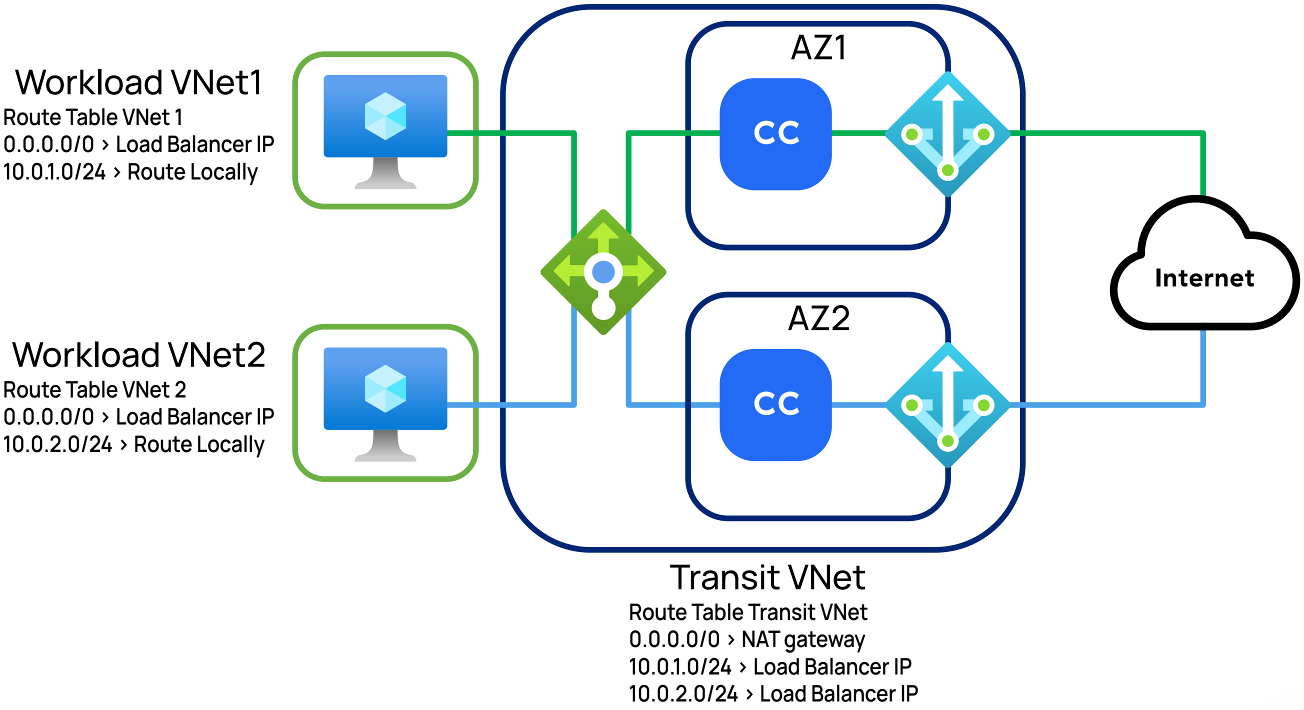

Workload VNet1
Route Table VNet 1
0.0.0.0/0 › Load Balancer IP
10.0.1.0/24 › Route Locally
Workload VNet2
Route Table VNet 2
0.0.0.0/0 › Load Balancer IP
10.0.2.0/24 › Route Locally
AZ1
CC
AZ2
CC
Internet
Transit VNet
Route Table Transit VNet
0.0.0.0/0 › NAT gateway
10.0.1.0/24 › Load Balancer IP
10.0.2.0/24 › Load Balancer IP

Figure 10. Cloud Connectors deployed in a highly available configuration within the transit VNet

Deploying Cloud Connector using a Transit/Egress VNet allows the organization to simplify cloud routing and, more importantly, reduce the compute footprint required when deploying directly to the workload VNets. In this option, only a single pair of Cloud Connector appliances is necessary for a Transit/Egress VNet. Spoke VNet workloads requiring internet or private access are simply directed towards the front-end load balancer IP address using a simple default route, where they can then be directed towards the Cloud Connector appliances for outbound routing.

By default, Terraform installs Cloud Connector using a Transit/Egress VNet model, though it can be customized to suit an organization's deployment needs.

You should be aware that the potential exists for double billing in this model. Microsoft Azure bills you based on egressing traffic out of a VNet. Specifically in this model, the same traffic egresses a VNet twice (once as it travels from the workload VNet to the Transit/Egress VNet, then again as it leaves the Transit/Egress VNet to the internet).

When implementing this design option, the first step is to consider which automation technique to employ. This has been discussed at length in **Deploying Cloud Connector via Terraform Scripts** and **Deploying Cloud Connector via Marketplace Application** earlier in this guide.

If using Marketplace Applications, consider deploying a second Cloud Connector appliance within a separate availability zone. This can be done by simply re-running the Marketplace Application workflow again. Marketplace Applications are not capable of instantiating Microsoft Azure Load Balancer. Azure Load Balancer needs to be set up manually or leverage Terraform scripts downloaded from the Cloud Connector portal.

If using Marketplace Applications scripts, it is recommended to deploy NAT Gateway. Since NAT Gateway(s) operate within a single availability zone, Zscaler recommends creating a second NAT Gateway in a different AZ so that an infrastructure failure of one AZ does not affect both NAT Gateways. Ensure that the Cloud Connector in the first AZ is in a different subnet than the Cloud Connector in the second AZ. Then, associate each NAT Gateway to each respective subnet.

©2023 Zscaler, Inc. All rights reserved.

Use Case: Integrating Zscaler Private Access

After you deploy your Cloud Connectors, you can add support for ZPA. This use case is growing in popularity as organizations seek to depart from legacy VPN technologies to interconnect cloud and on-premises workloads. An important consideration with Cloud Connector is that it is designed to facilitate outbound workload traffic towards a remote destination. When the destination is in a location you control, we must consider how this traffic ingresses into the remote facility. We do this using the Zscaler App Connector appliance, where App Connector VMs sit adjacent to the workloads they provide access to.

This model builds on the foundation provided in the direct-to-internet and VNet Peering use cases discussed previously. Cloud Connector provides outbound connectivity for cloud workloads to an on-premises data center, which uses App Connector appliances (VMs) sitting in an application server segment to provide inbound connectivity. Both appliances build DTLS tunnels to the ZPA Broker and establish a Microtunnel between the source (cloud) workload and the destination data center workload. The traffic within the Microtunnel targets synthetic proxy IP addresses inside the Cloud Connector and App Connector, respectively.

The following image is simplified for clarity. Redundant instances of Zscaler Cloud Connector should be deployed in all instances.

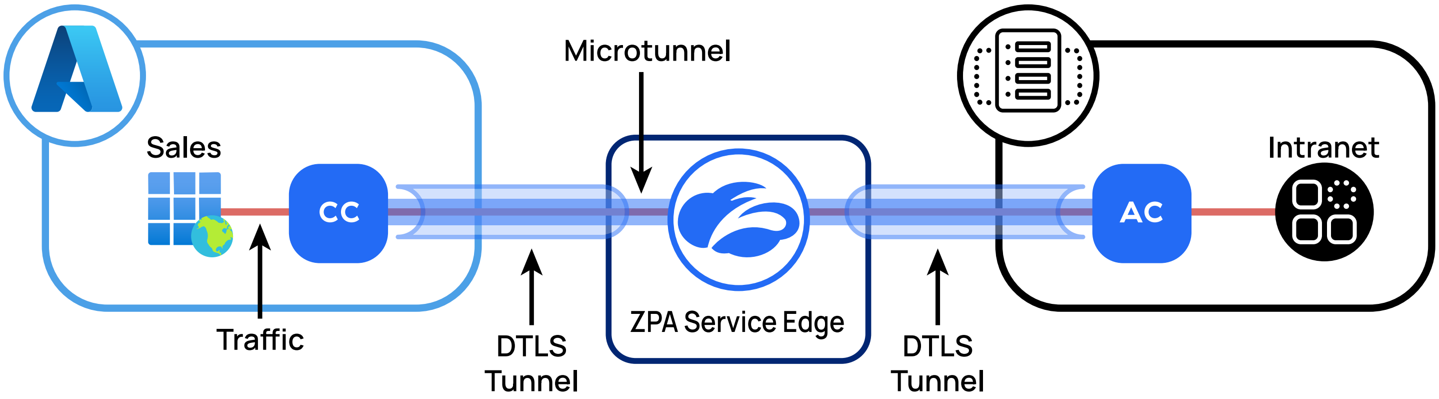


Figure 11. Microsoft Azure access to internal apps in the private data center

All communication between ZPA components travel within a mutually pinned, client and server certificate-verified TLS connection. Within this TLS-encrypted Zscaler Tunnel, a microtunneling protocol (i.e., Microtunnel) exists. Select components of ZPA run through this encrypted Microtunnel end to end. Because the client and server use pinned certificates, it is cryptographically impossible for ZPA to experience a Man-in-the-Middle (MITM) attack. The client certificates are verified against an organization's Certificate Authority (CA) and the server certificates are verified against Zscaler's CA, which cannot be spoofed by any third-party compromised CA.

©2023 Zscaler, Inc. All rights reserved.

ZPA only accepts connections from the Zscaler Cloud Connector and the App Connector instances that present a client certificate signed by a CA associated with each tenant. Zscaler Cloud Connector and App Connector only connect to ZPA service components that present a certificate signed by the ZPA infrastructure PKI.

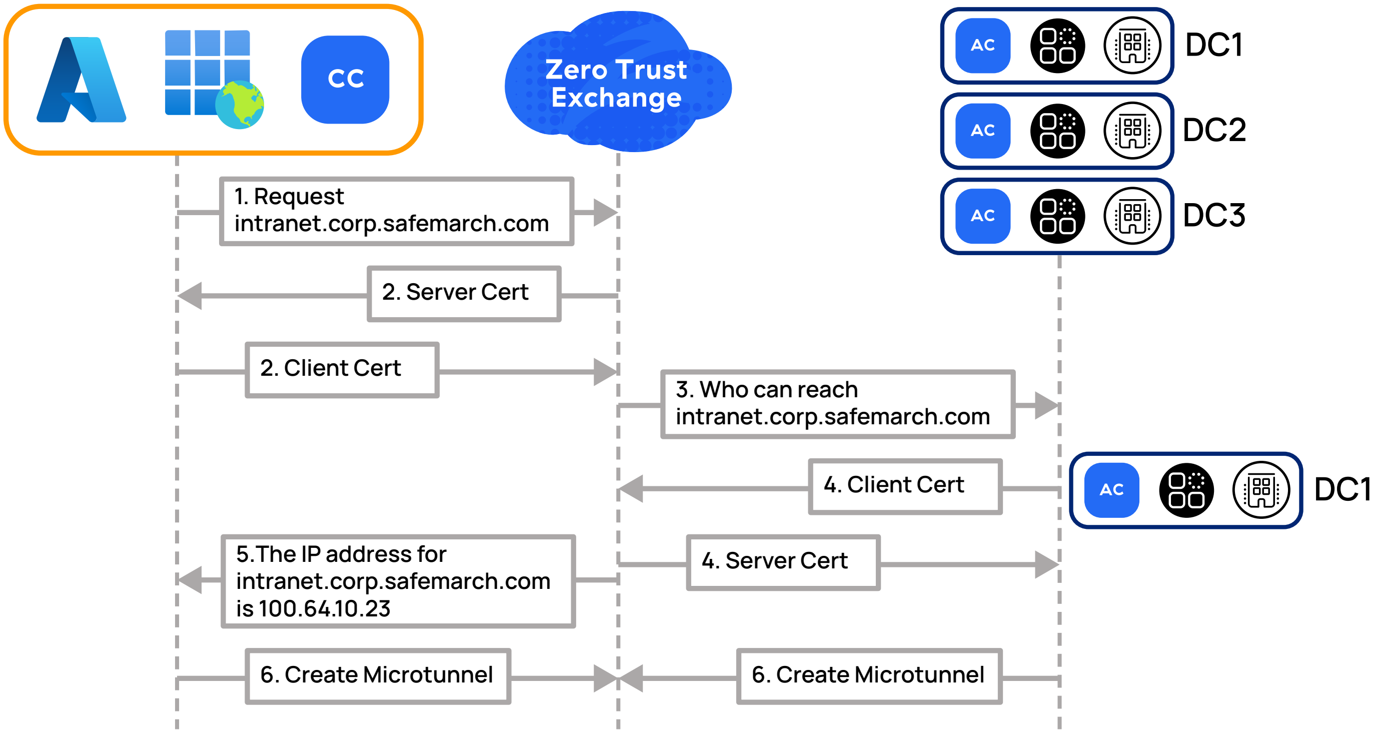

CC
Zero Trust Exchange
AC
DC1
AC
DC2
AC
DC3
1. Request intranet.corp.safemarch.com
2. Server Cert
2. Client Cert
3. Who can reach intranet.corp.safemarch.com
AC
DC1
4. Client Cert
5.The IP address for intranet.corp.safemarch.com is 100.64.10.23
4. Server Cert
6. Create Microtunnel
6. Create Microtunnel

Figure 12. Authentication and tunnel setup between workloads and internal apps

1. A workload requests access to an application.
2. The ZPA Service Edge and Zscaler Cloud Connector authenticate via certificate exchange.
3. If the workload is authorized to access the requested application, the ZPA Service Edge determines which App Connector can service the request.
4. The ZPA Service Edge and Zscaler App Connector authenticate via certificate exchange.
5. The workload is presented with the synthetic IP of the application.
6. A Microtunnel is established between the Zscaler Cloud Connector and Zscaler App Connector.

Zscaler Cloud Connector recognizes the internal applications that are available via ZPA. Access to these applications is defined in ZPA based on policies. Using information received from the ZPA Public Service Edge or ZPA Private Service Edge, Cloud Connector intercepts workload requests for applications, and then forwards those requests to the ZPA cloud.

No network information is required to access available applications. To facilitate secure private connections that are abstracted from the physical network, Cloud Connector associates permitted internal applications with a set of synthetic IP addresses.

When a workload sends out a DNS request, Zscaler Cloud Connector can recognize the domain as an internal application being protected by ZPA. Zscaler Cloud Connector then intercepts the DNS request and delivers a DNS response to the workload that uses the synthetic IP address associated with the internal application.

To intercept and modify DNS requests, Cloud Connector must see the initial request from the cloud workload. To facilitate this, Zscaler recommends adding a custom DNS server within the Azure cloud. Ensure internal domain requests are forwarded across the Cloud Connector.

©2023 Zscaler, Inc. All rights reserved.

When implementing this design option, the first step is to consider which automation technique to employ.

- If using Marketplace Applications, consider deploying a second Cloud Connector appliance within a separate availability zone. This can be done by simply re-running the Marketplace Application workflow again.
- Marketplace Applications are not capable of instantiating Microsoft Azure Load Balancer. Azure Load Balancer needs to be set up manually, or leverage Terraform scripts downloaded from the Cloud Connector portal.
- If using Marketplace Applications scripts, it is recommended to deploy NAT Gateway. Since NAT Gateway(s) operate within a single availability zone, Zscaler recommends creating a second NAT Gateway in a different AZ so that an infrastructure failure of one AZ does not affect both NAT Gateways. Ensure that the Cloud Connector in the first AZ is in a different subnet than the Cloud Connector in the second AZ. Then, associate each NAT Gateway to each respective subnet towards the new NAT Gateway.
- Since cloud workloads initiate requests for resources in an on-premises data center towards App Connector, you must ensure DNS requests from these cloud workloads transit the Cloud Connector.
- By default, a rule already exists for ZPA-bound traffic, but you should ensure that the Cloud Connector **Forwarding Options** are correctly matching and forwarding traffic to ZPA.
- Ensure application segments have been defined within the ZPA portal and, if using a custom DNS server, ensure these same application domains are forwarded through the Cloud Connector.
- Ensure that Zscaler Private Access policy is configured to accept inbound traffic from cloud locations and allowed (or denied) access to the internet.
- When co-located with other Cloud Connectors, App Connectors must sit parallel to the Cloud Connector and not “behind” the Cloud Connector.

©2023 Zscaler, Inc. All rights reserved.

Use Case: Securing Traffic Between Clouds with ZPA

Multi-cloud deployments, where workloads are spread across more than one cloud provider, are becoming more common as organizations look to provide hosting across more than one vendor. You might host your cloud workloads in more than one cloud or, for redundancy or geoproximity, in multiple regions of the same cloud service provider. This use case focuses on how to solve for the challenges faced in this scenario and how we can secure this traffic using the ZPA model discussed previously.

This use case is like that of the ZPA model discussed previously but builds on the fact that remote application destinations secured by ZPA might not be in an on-premises data center. Instead, these applications exist within a different cloud region or in a different cloud service provider altogether. It is common in this scenario to see both Cloud Connector and App Connector co-located in the same workload VNet or Transit/Egress VNet. As originating cloud workloads send requests to remote applications, Cloud Connector routes them to the appropriate App Connectors in the destination cloud.

The following image is simplified for clarity. Redundant instances of Zscaler Cloud Connector should be deployed in all instances.

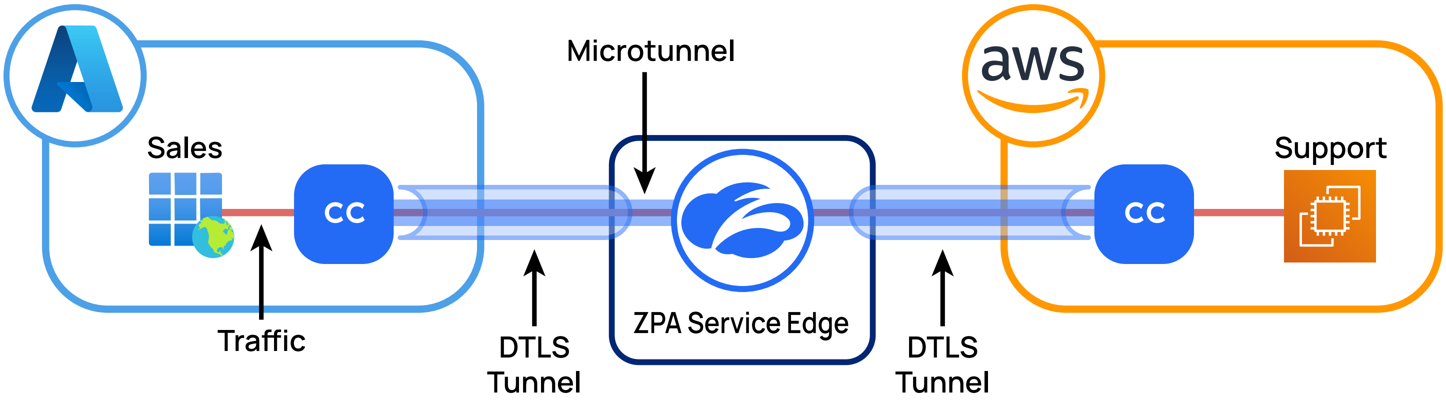


Figure 13. Workload-to-workload communication across cloud providers

When implementing this design option, the first step is to consider which automation technique to employ.

- If using Marketplace Applications, consider deploying a second Cloud Connector appliance within a separate availability zone. This can be done by simply re-running the Marketplace Application workflow again. Marketplace Applications are not capable of instantiating Microsoft Azure Load Balancer. Azure Load Balancer needs to be set up manually or leverage Terraform scripts downloaded from the Cloud Connector portal.
- If using Marketplace Applications scripts, it is recommended to deploy NAT Gateway. Since NAT Gateway(s) operate within a single availability zone, Zscaler recommends creating a second NAT Gateway in a different AZ so that an infrastructure failure of one AZ does not affect both NAT Gateways. Ensure that the Cloud Connector in the first AZ is in a different subnet than the Cloud Connector in the second AZ. Then, associate each NAT Gateway to each respective subnet.

Since cloud workloads initiate requests for resources in an adjacent cloud or region (towards App Connector), you must ensure DNS requests from these cloud workloads transit the Cloud Connector.

By default, a rule already exists for ZPA-bound traffic, but you should ensure that the Cloud Connector **Forwarding Options** are correctly matching and forwarding traffic to ZPA.

Ensure application segments have been defined within the ZPA portal and, if using a custom DNS server, ensure these same application domains are forwarded through the Cloud Connector.

©2023 Zscaler, Inc. All rights reserved.

Ensure that Zscaler Private Access policy is configured to accept inbound traffic from cloud locations and allowed (or denied) access to the internet.

When co-located with other Cloud Connectors, it is imperative that App Connectors sit parallel to the Cloud Connector and not “behind” the Cloud Connector.

For inbound traffic from a remote destination, App Connector must be able to resolve the FQDN of the requested host to the real IP address. So, you must ensure that App Connectors are pointed at a real DNS server that can resolve workload FQDNs. Consider this when deploying custom DNS servers.

Summary

Connecting workloads to the internet across different networks is difficult. What makes this harder is the traditional approach used by organizations to solve this challenge, such as using technologies like VPNs and firewalls. While the outcome of connecting these workloads is achieved, the cost to achieve these goals is significant:

- Risk of lateral threats and internet-based attacks by overextending the trusted network across the internet using VPN and WAN technologies.
- Complexity increases because of complicated route filtering, multiple network hops, and fragmented policy management.
- Poor visibility across application connectivity paths and increased network blind spots.
- Costs rise due to overprovisioning network services and the use of virtual appliances such as firewalls, IPs, routers, and other point products in cloud environments.
- Limited scale and performance from the increase in network and security services used in cloud environments.

As a result, there is a need for a better approach. Zscaler Cloud Connector is a cloud-native zero trust access service that provides fast and secure app-to-app, app-to-internet connectivity across multi-cloud environments. With integrated automated connectivity and security, it reduces complexity and cost, and provides a faster, smarter, and more secure alternative to legacy network solutions.

©2023 Zscaler, Inc. All rights reserved.

About Zscaler

Zscaler (NASDAQ: ZS) accelerates digital transformation so customers can be more agile, efficient, resilient, and secure. The Zscaler Zero Trust Exchange protects thousands of customers from cyberattacks and data loss by securely connecting users, devices, and applications in any location. Distributed across more than 150 data centers globally, the SASE-based Zero Trust Exchange is the world's largest inline cloud security platform.

©2023 Zscaler, Inc. All rights reserved. Zscaler, Zero Trust Exchange, Zscaler Private Access, ZPA, Zscaler Internet Access, ZIA, Zscaler Digital Experience, and ZDX are either (i) registered trademarks or service marks or (ii) trademarks or service marks of Zscaler, Inc. in the United States and/or other countries. Any other trademarks are the properties of their respective owners.