

These considerations implement the pillars of the Azure Well-Architected Framework, which is a set of guiding tenets that you can use to improve the quality of a workload. For more information, see [Microsoft Azure Well-Architected Framework](#).

## Reliability

Reliability ensures your application can meet the commitments you make to your customers. For more information, see [Overview of the reliability pillar](#).

This solution's resiliency depends on the failure modes of individual services like Azure Virtual Machines, Azure Database for MySQL, and Azure Load Balancer. For more information, see these articles:

- [Design for reliability](#)
- [Resiliency in Azure](#)
- [Resiliency checklist for specific Azure services](#)

For information about disaster recovery, see [Business continuity and disaster recovery](#).

For Azure VMs in the web tier, you can use [availability sets](#) to create a logical grouping of VMs that provides redundancy and availability. We recommend at least two VMs per availability set to create a high-availability application and meet the [Azure SLA](#).

Because Azure Database for MySQL is a managed database as a service, its architecture is optimized for built-in high availability. For information about SLAs, see [SLAs for Azure Database for MySQL](#).

## Security

Security provides assurances against deliberate attacks and the abuse of your valuable data and systems. For more information, see [Overview of the security pillar](#).

You might want to add [Azure Web Application Firewall](#) to this solution. It helps protect your application from common vulnerabilities. This Azure Application Gateway option uses Open Web Application Security Project (OWASP) rules to help prevent attacks like cross-site scripting, session hijacking, and other exploits.

As an added layer of protection, be sure to use [Azure network security groups](#) to filter network traffic traveling to and from Azure resources in the Azure virtual network.

You should also consider [Private Link for Azure Database for MySQL](#). You can use Private Link to connect to platform as a service (PaaS) services in Azure via a private endpoint. Private Link essentially brings Azure services inside your private virtual network. PaaS

resources can be accessed via the private IP address just like any other resource in the virtual network.

Finally, follow [these security guidelines](#) when you implement this solution.

## Cost optimization

Cost optimization is about finding ways to reduce unnecessary expenses and improve operational efficiencies. For more information, see [Overview of the cost optimization pillar](#).

To better understand the cost of running this scenario on Azure, use the [pricing calculator](#).

For more information about the cost of this solution, see:

- [Linux Virtual Machines pricing](#)
- [Azure Database for MySQL pricing](#)
- [Load Balancer pricing](#)
- [Azure Reserved Virtual Machine Instances](#)

## Operational excellence

Operational excellence covers the operations processes that deploy an application and keep it running in production. For more information, see [Overview of the operational excellence pillar](#).

Operational excellence applies reliability, predictability, and automated operations to your architecture to keep your application running in production. Deployments must be reliable and predictable. Automated deployments reduce the chance of human error.

Implement software engineering disciplines across your entire environment, including these practices:

- Implement [Infrastructure as Code](#).
- Build and release with [continuous integration](#) and [continuous delivery](#) (CI/CD) pipelines.
- Use automated testing.

For more information, see [Operational excellence design principles](#).

## Performance efficiency

Performance efficiency is the ability of your workload to scale to meet the demands placed on it by users in an efficient manner. For more information, see [Performance efficiency pillar overview](#).

To accomplish this goal, consider using [Azure Virtual Machine Scale Sets](#), which you can use to create and manage a group of load-balanced VMs. The number of VM instances can automatically increase or decrease in response to demand or according to a defined schedule.

## Deploy this scenario

We recommend that you use the Bash environment in [Azure Cloud Shell](#) to deploy this solution. If you'd rather run commands in your own Windows, Linux, or macOS environment, [install the Azure CLI](#).

For deployment steps, see [Deploying Apache Guacamole on Azure](#) .

## Contributors

*This article is maintained by Microsoft. It was originally written by the following contributors.*

Principal author:

- [Ricardo Macedo Martins](#)  | Sr. Customer Engineer

Other contributor:

- [Mick Alberts](#)  | Technical Writer

*To see non-public LinkedIn profiles, sign in to LinkedIn.*

## Next steps

- [Apache Guacamole documentation](#) 
- [Azure Bastion documentation](#)
- [What is Azure Load Balancer?](#)
- [Azure Database for MySQL](#)
- [Learn module: Introduction to Azure Bastion](#)
- [Learn module: Introduction to Azure Load Balancer](#)

## Related resources

- Build solutions for high availability using availability zones
- Build high availability into your BCDR strategy
- Baseline zone-redundant web application
- Highly available multi-region web application

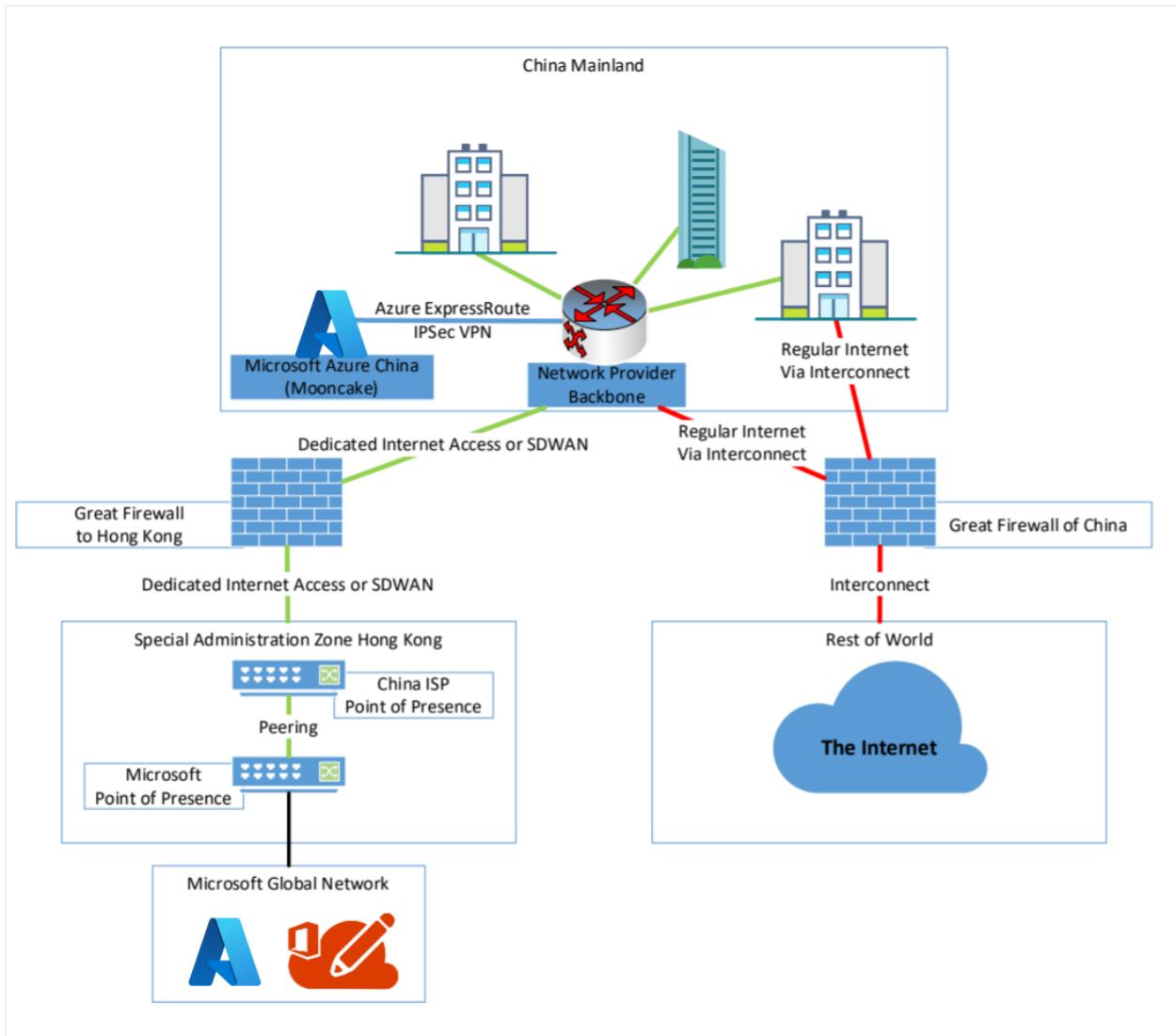
# Interconnect with China using Azure Virtual WAN and Secure Hub

Article • 02/15/2023

When looking at common automotive, manufacturing, logistics industries, or other institutes like embassies, there is often the question about how to improve interconnection with China. Those improvements are mostly relevant for using Cloud Services like Microsoft 365, Azure Global Services, or interconnect branches inside of China with a customer backbone.

In most of the cases, customers are struggling with high latencies, low bandwidth, unstable connection, and high costs connecting to outside of China (for example, Europe or the United States).

A reason for these struggles is the "Great Firewall of China", which protects the Chinese part of the Internet and filters traffic to China. Nearly all traffic running from People's Republic of China to outside of China, except the special administration zones like Hong Kong and Macau, passes the Great Firewall. The traffic running through Hong Kong and Macau doesn't hit the Great Firewall in full force, it's handled by a subset of the Great Firewall.



Using Virtual WAN, a customer can establish a more performant and stable connection to Microsoft Cloud Services and a connection to their enterprise network without breaking the Chinese cybersecurity law.

## Requirements and workflow

If you want to stay compliant to the Chinese cybersecurity law, you need to meet a set of certain conditions.

First, you need to work together with a network and ISP who owns an ICP (Internet Content Provider) license for China. In most cases, you'll end up with one of the following providers:

- China Telecom Global Ltd.
- China Mobile Ltd.
- China Unicom Ltd.
- PCCW Global Ltd.
- Hong Kong Telecom Ltd.

Depending on the provider and your needs, you now need to purchase one of the following network connectivity services to interconnect your branches within China.

- A MPLS/IPVPN Network
- A Software Defined WAN (SDWAN)
- Dedicated Internet Access

Next, you need to agree with that provider to give a breakout to the Microsoft Global Network and its Edge Network in Hong Kong, not in Beijing or Shanghai. In this case, Hong Kong is very important because of its physical connection and location to China.

While most customers think using Singapore for interconnect is the best case because it looks nearer to China when looking on the map, this isn't true. When you follow network fiber maps, nearly all network connects go through Beijing, Shanghai, and Hong Kong. This makes Hong Kong a better location choice to interconnect to China.

Depending on the provider, you may get different service offerings. The table below shows an example of providers and the service they offer, based on information at the time this article was written.

Service	Provider examples
MPLS/IPVPN Network	PCCW, China Telecom Global
SDWAN	PCCW, China Telecom Global
Dedicated Internet Access	PCCW, Hong Kong Telecom, China Mobil

With your provider, you can agree on which of the following two solutions to use to reach the Microsoft global backbone:

- Getting a Microsoft Azure ExpressRoute terminated in Hong Kong. That would be the case for the use of MPLS/IPVPN. Currently, only the only ICP license provider with ExpressRoute to Hong Kong is China Telecom Global. However, they can also talk to the other providers if they leverage Cloud Exchange Providers like Megaport or InterCloud. For more information, see [ExpressRoute connectivity providers](#).
- Using a Dedicated Internet Access directly at one of the following Internet Exchange Points, or using a private network interconnect.

The following list shows Internet Exchanges possible in Hong Kong:

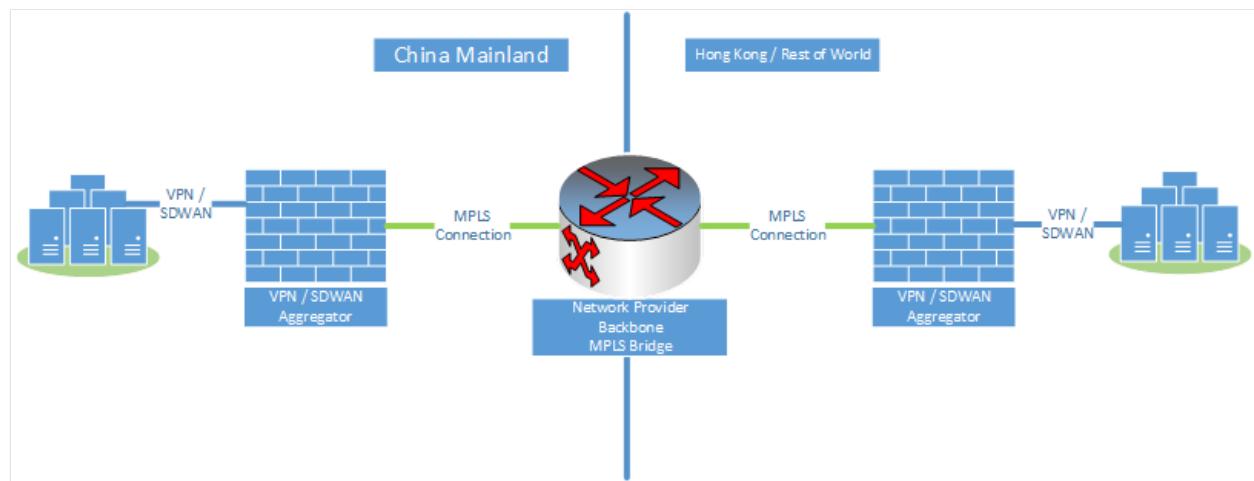
- AMS-IX Hong Kong
- BBIX Hong Kong

- Equinix Hong Kong
- HKIX

When using this connect, your next BGP hop for Microsoft Services must be Microsoft Autonomous System Number (AS#) 8075. If you use a single location or SDWAN solution, that would be the choice of connection.

With the current changes regarding interconnects between China and Hong Kong SAR, most of these network providers build an MPLS bridge between China and Hong Kong SAR.

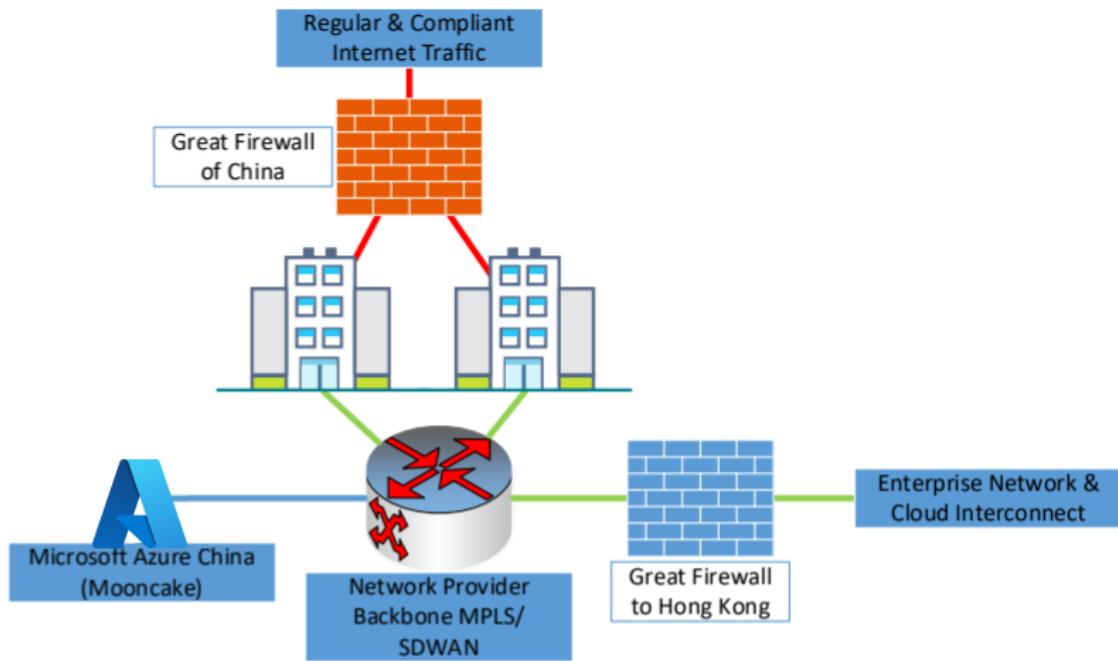
You can see that site-to-site VPN connections inside China are allowed and are mostly stable. The same applies for the site-to-site connections between branches in the rest of the world. Providers now create a VPN/SDWAN Aggregation on both sides and bridge via MPLS between them.



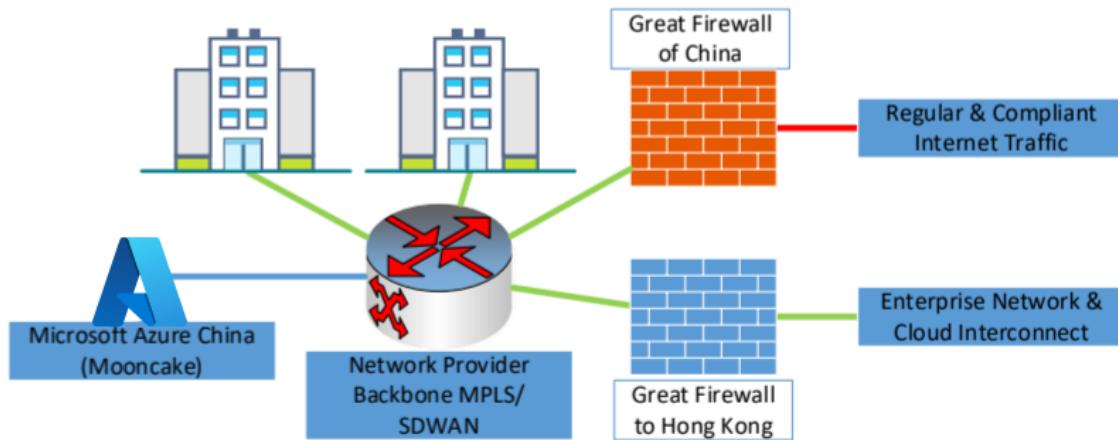
Either way, we still recommend that you have a second and regular internet breakout into China. This is to split the traffic between enterprise traffic to cloud services like Microsoft 365 and Azure, and by-law regulated internet traffic.

A compliant network architecture within China could look like the following example:

### Multiple Branches using WAN Backbone & Internet Breakout per Branch



### Multiple Branches using WAN Backbone & central managed Internet Breakout

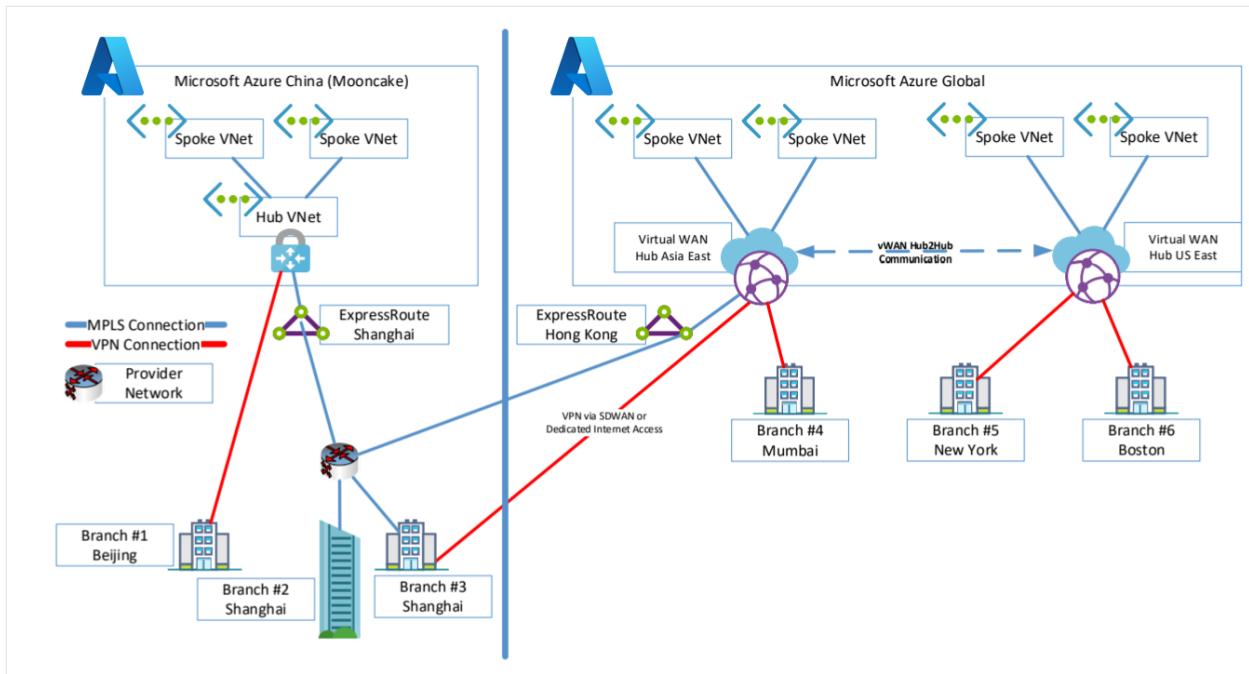


In this example, having an interconnect with the Microsoft Global Network in Hong Kong, you can now start to leverage the [Azure Virtual WAN Global Transit Architecture](#) and additional services, like Azure secure Virtual WAN hub, in order to consume services and interconnect to your branches and datacenter outside China.

## Hub-to-hub communication

In this section, we use Virtual WAN hub-to-hub communication to interconnect. In this scenario, you create a new Virtual WAN hub resource to connect to a Virtual WAN hub in Hong Kong, other regions you prefer, a region where you already have Azure resources, or where want to connect.

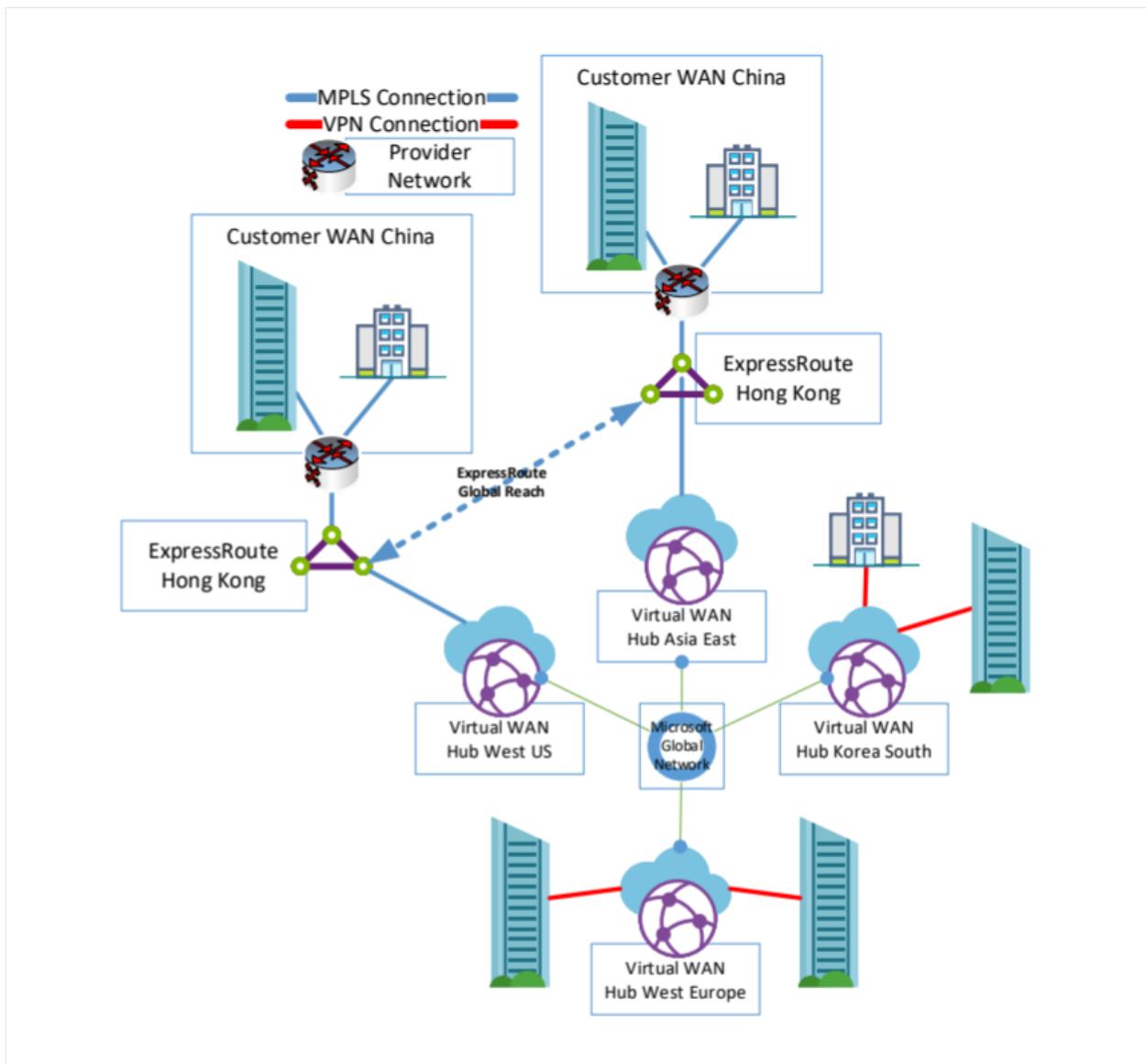
A sample architecture could look like following example:



In this example, the China branches connect to Azure Cloud China and each other by using VPN or MPLS connections. Branches that need to be connected to Global Services use MPLS or Internet-based services that are connected directly to Hong Kong. If you want to use ExpressRoute in Hong Kong and in the other region, you need to configure [ExpressRoute Global Reach](#) to interconnect both ExpressRoute Circuits.

ExpressRoute Global Reach isn't available in some regions. If you need to interconnect with Brazil or India, for example, you need to leverage [Cloud Exchange Providers](#) to provide the routing services.

The figure below shows both examples for this scenario.

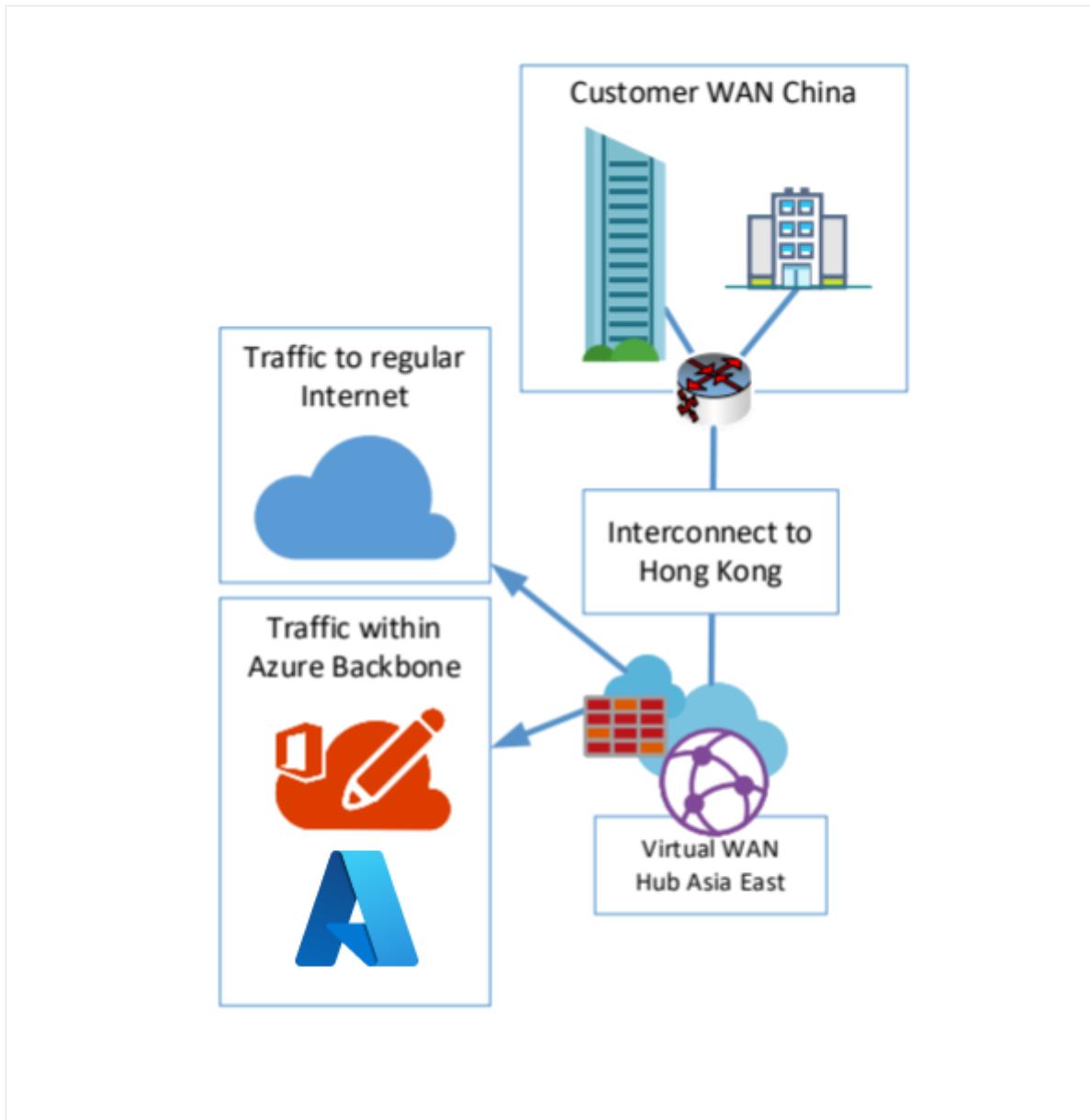


## Secure Internet breakout for Microsoft 365

Another consideration is network security and logging for the entry point between China and the Virtual WAN established backbone component, and the customer backbone. In most cases, there is a need to breakout to the Internet in Hong Kong to directly reach the Microsoft Edge Network and, with that, the Azure Front Door Servers used for Microsoft 365 Services.

For both scenarios with Virtual WAN, you would leverage the [Azure Virtual WAN secured hub](#). Using Azure Firewall Manager, you can change a regular Virtual WAN hub to a secured hub, and then deploy and manage an Azure Firewall within that hub.

The following figure shows an example of this scenario:



## Architecture and traffic flows

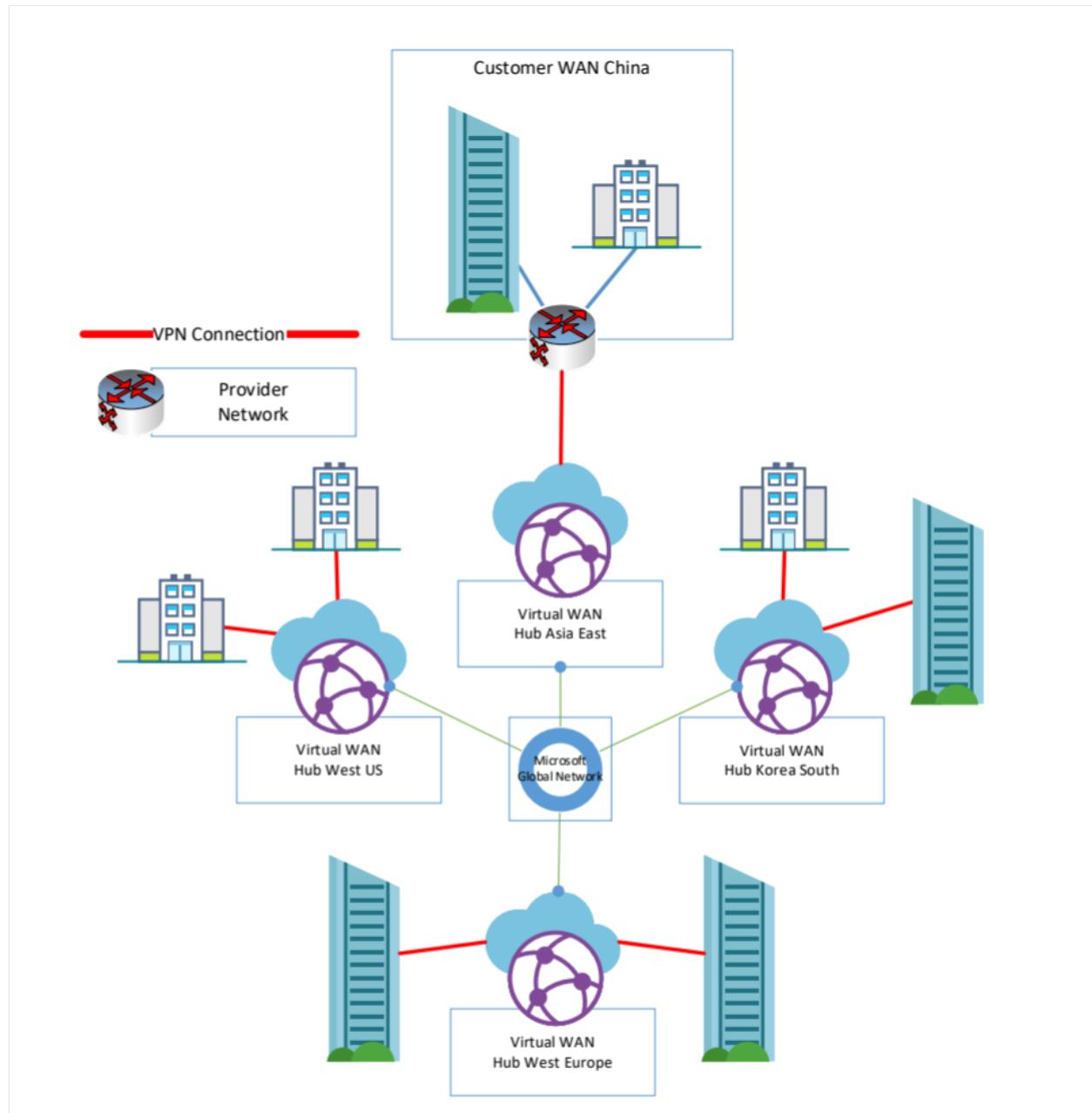
Depending on your choice regarding the connection to Hong Kong, the overall architecture may change slightly. This section shows three available architectures in different combination with VPN or SDWAN and/or ExpressRoute.

All of these options make use of Azure Virtual WAN secured hub for direct Microsoft 365 connectivity in Hong Kong. These architectures also support the compliance requirements for [Microsoft 365 Multi-Geo](#) and keep that traffic near the next Azure Front Door location. As a result, it's also an improvement for the usage of Microsoft 365 out of China.

When using Azure Virtual WAN together with Internet connections, every connection can benefit from additional services like [Microsoft Azure Peering Services \(MAPS\)](#). MAPS was built to optimize traffic coming to the Microsoft Global Network from 3rd Party Internet Service Providers.

## Option 1: SDWAN or VPN

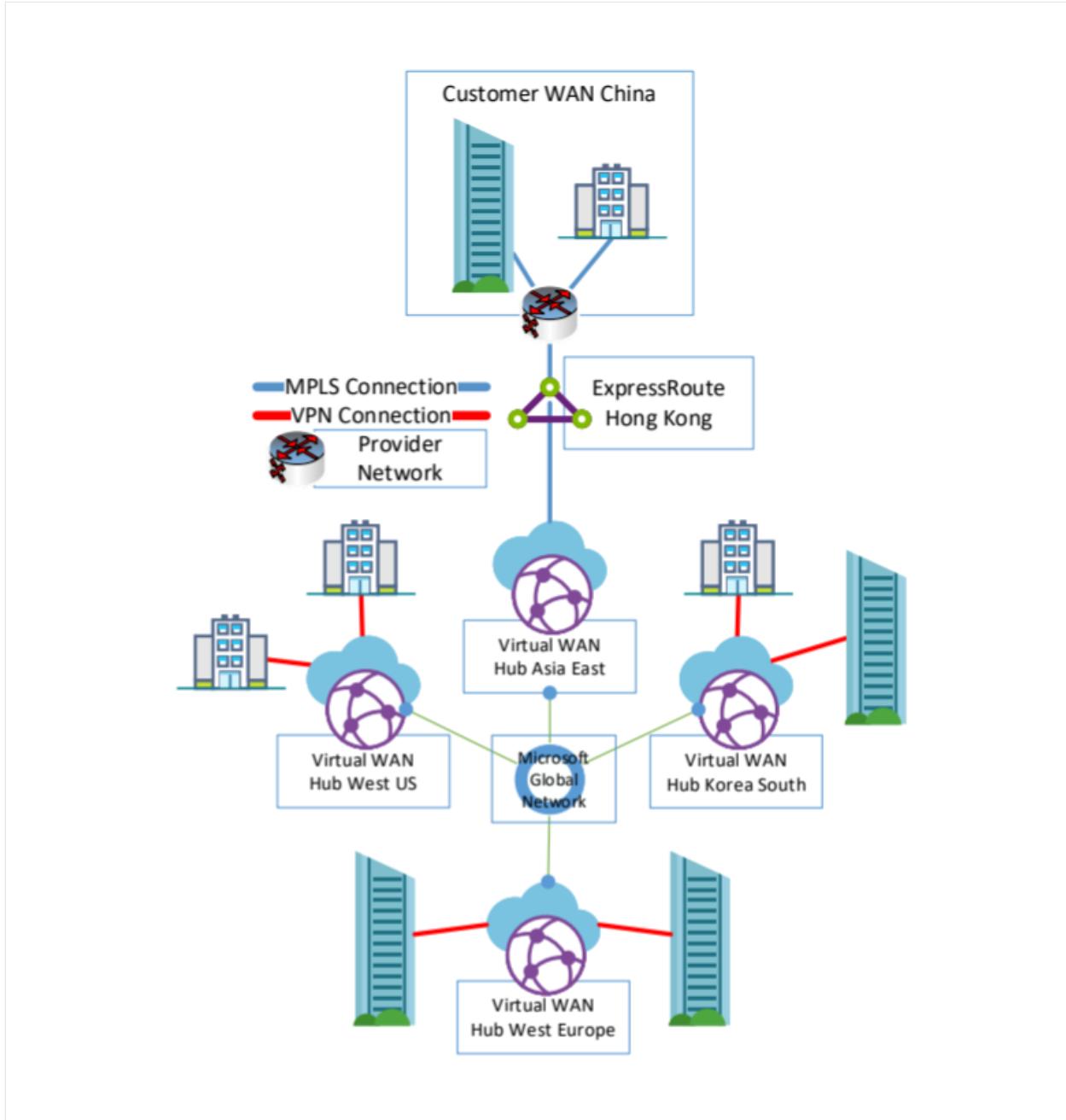
This section discusses a design that uses SDWAN or VPN to Hong Kong and to other branches. This option shows the use and traffic flow when using pure Internet connection on both sites of the Virtual WAN backbone. In this case, the connection is brought to Hong Kong using dedicated Internet access, or an ICP provider SDWAN solution. Other branches are using pure Internet or SDWAN Solutions as well.



In this architecture, every site is connected to the Microsoft Global Network by using VPN and Azure Virtual WAN. The traffic between the sites and Hong Kong is transmitted through the Microsoft Network and only uses regular Internet connection on the last mile.

## Option 2: ExpressRoute and SDWAN or VPN

This section discusses a design that uses ExpressRoute in Hong Kong and other Branches with VPN/SDWAN Branches. This option shows the use of ExpressRoute terminated in Hong Kong and other branches connected via SDWAN or VPN. ExpressRoute in Hong Kong is currently limited to a short list of Providers, which you can find in the list of [Express Route Partners](#).



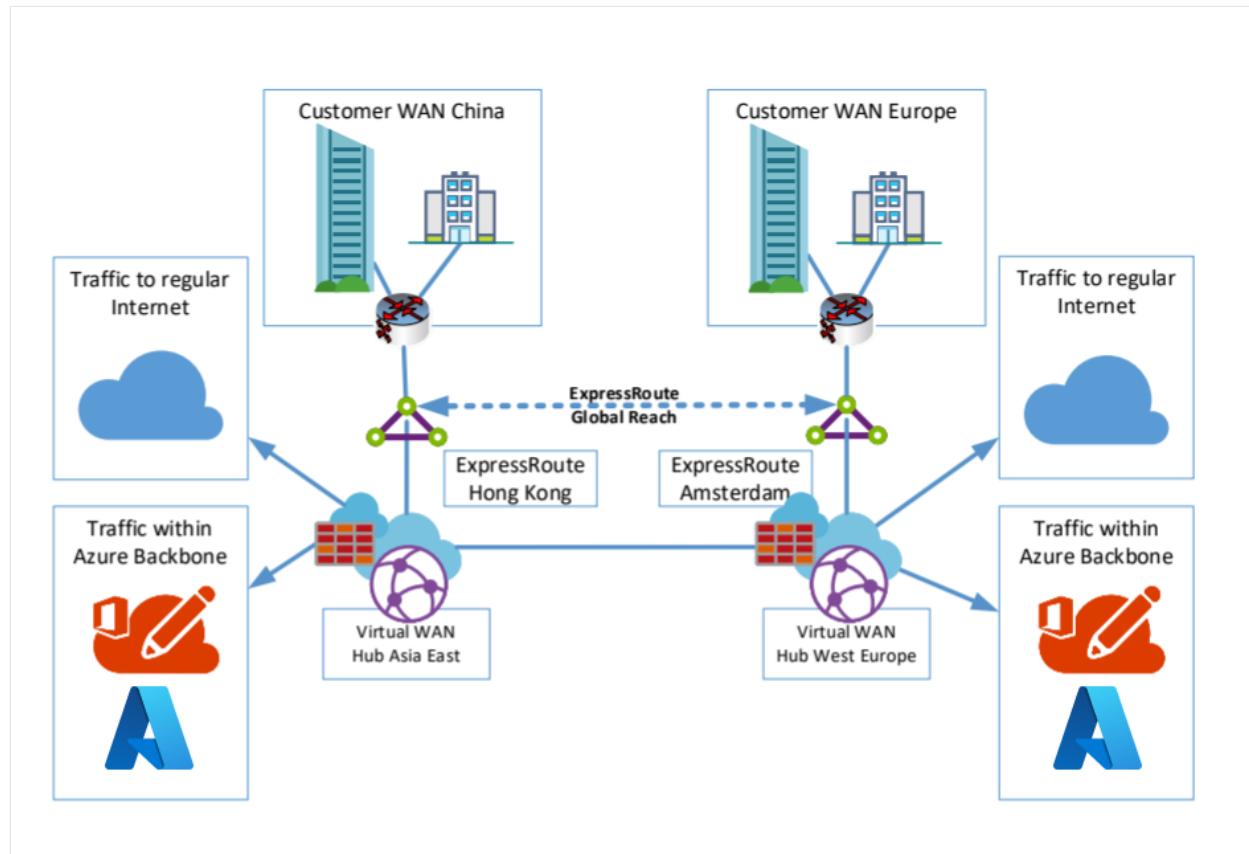
There are also options to terminate ExpressRoute from China, for example, in South Korea or Japan. But, given compliance, regulation, and latency, Hong Kong is currently the best choice.

## Option 3: ExpressRoute only

This section discusses a design that where ExpressRoute is used for Hong Kong and other Branches. This option shows the interconnect using ExpressRoute on both ends.

Here you have a different traffic flow than the other. The Microsoft 365 traffic will flow to the Azure virtual WAN secured hub and from there to the Microsoft Edge Network and the Internet.

The traffic that goes to the interconnected branches or from them to the locations in China will follow a different approach within that architecture. Currently virtual WAN doesn't support ExpressRoute to ExpressRoute transit. The traffic will leverage ExpressRoute Global Reach or the 3rd Party interconnect without passing the virtual WAN Hub. It will directly flow from one Microsoft Enterprise Edge (MSEE) to another.



Currently ExpressRoute Global Reach isn't available in every country/region, but you can configure a solution using Azure Virtual WAN.

You can, for example, configure an ExpressRoute with Microsoft Peering and connect a VPN tunnel through that peering to Azure Virtual WAN. Now you have enabled, again, the transit between VPN and ExpressRoute without Global Reach and 3rd party provider and service, such as Megaport Cloud.

## Next steps

See the following articles for more information:

- [Global Transit network architecture with Azure Virtual WAN](#)

- Create a Virtual WAN hub
- Configure a Virtual WAN secured hub
- Azure Peering Service Preview Overview

# Migrate to Azure Virtual WAN

Article • 12/14/2022

Azure Virtual WAN lets companies simplify their global connectivity in order to benefit from the scale of the Microsoft global network. This article provides technical details for companies that want to migrate from an existing customer-managed hub-and-spoke topology, to a design that leverages Microsoft-managed Virtual WAN hubs.

For information about the benefits that Azure Virtual WAN enables for enterprises adopting a cloud-centric modern enterprise global network, see [Global transit network architecture and Virtual WAN](#).

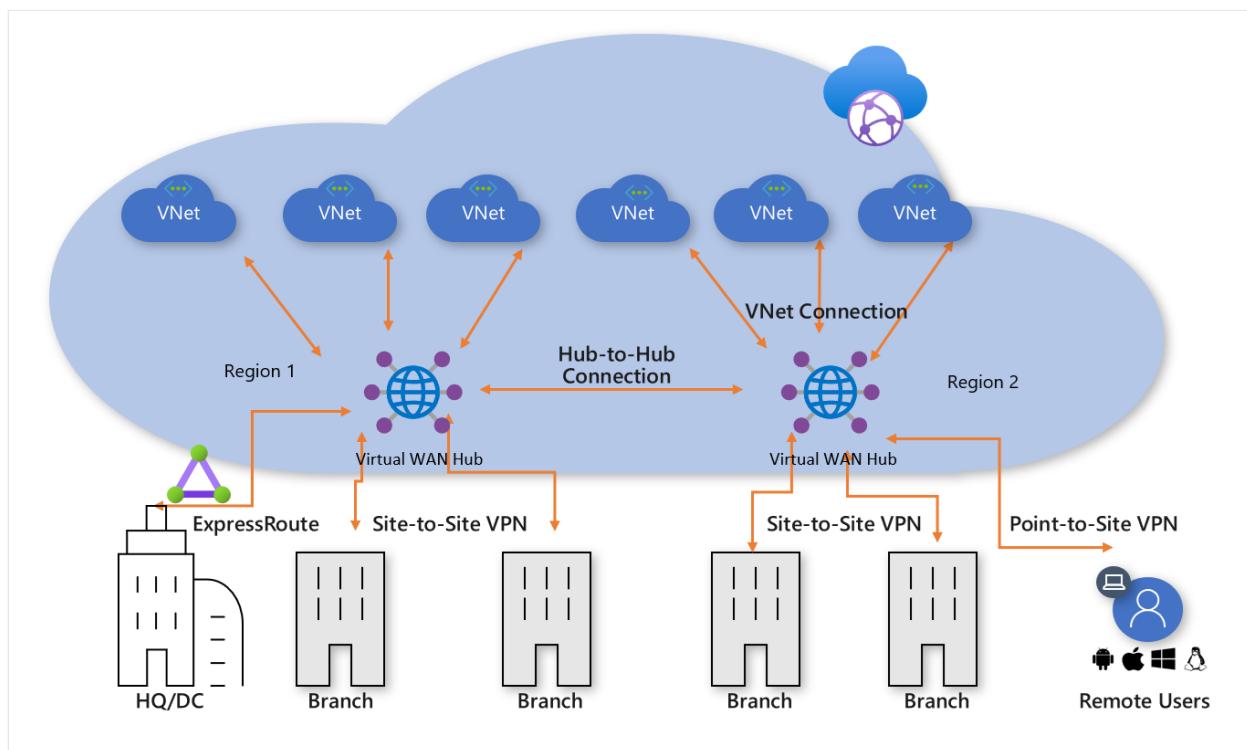


Figure: Azure Virtual WAN

The Azure hub-and-spoke connectivity model has been adopted by thousands of our customers to leverage the default transitive routing behavior of Azure Networking in order to build simple and scalable cloud networks. Azure Virtual WAN builds on these concepts and introduces new capabilities that allow global connectivity topologies, not only between on-premises locations and Azure, but also allowing customers to leverage the scale of the Microsoft network to augment their existing global networks.

This article shows how to migrate an existing customer-managed hub-and-spoke environment, to a topology that is based on Azure Virtual WAN.

## Scenario

Contoso is a global financial organization with offices in both Europe and Asia. They are planning to move their existing applications from an on-premises data center in to Azure and have built out a foundation design based on the customer-managed hub-and-spoke architecture, including regional hub virtual networks for hybrid connectivity. As part of the move to cloud-based technologies, the network team has been tasked with ensuring that their connectivity is optimized for the business moving forward.

The following figure shows a high-level view of the existing global network including connectivity to multiple Azure regions.

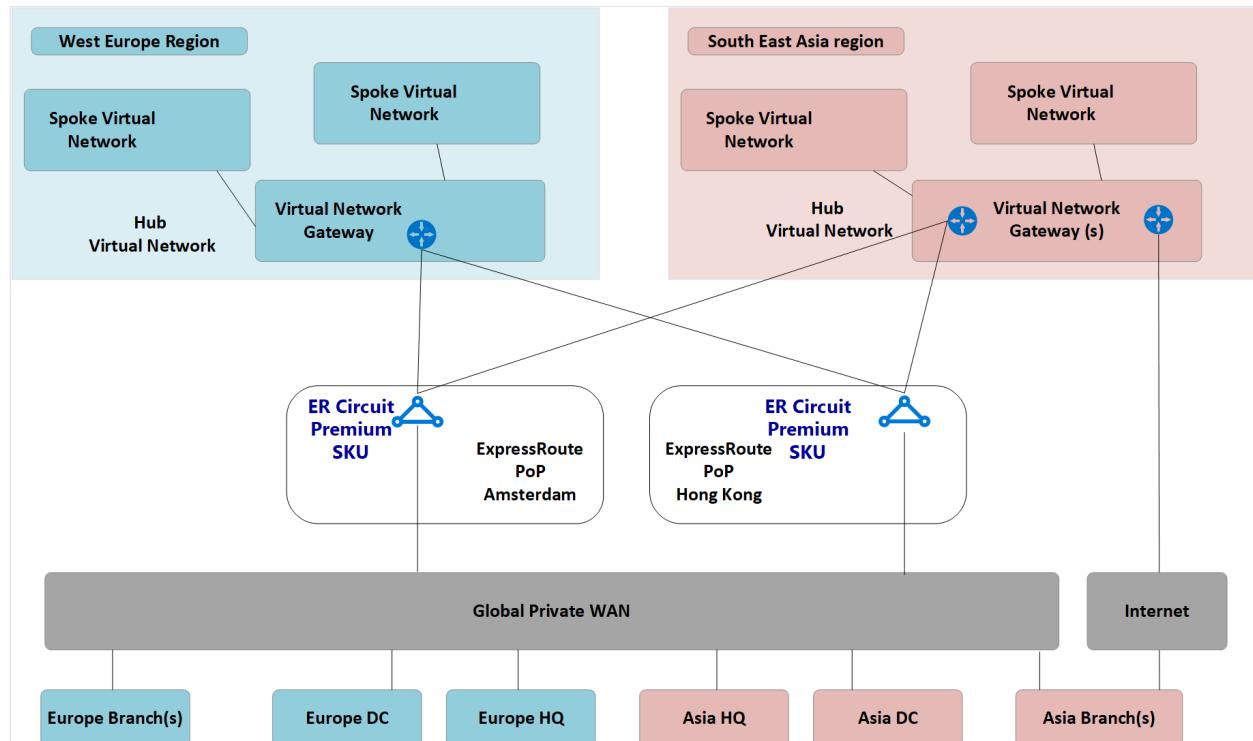


Figure: Contoso existing network topology

The following points can be understood from the existing network topology:

- A hub-and-spoke topology is used in multiple regions including ExpressRoute circuits for connectivity back to a common private Wide Area Network (WAN).
- Some of these sites also have VPN tunnels directly in to Azure to reach applications hosted within the cloud.

## Requirements

The networking team has been tasked with delivering a global network model that can support the Contoso migration to the cloud and must optimize in the areas of cost, scale, and performance. In summary, the following requirements are to be met:

- Provide both head quarter (HQ) and branch offices with optimized path to cloud hosted applications.

- Remove the reliance on existing on-premises data centers (DC) for VPN termination while retaining the following connectivity paths:
  - **Branch-to-VNet:** VPN connected offices must be able to access applications migrated to the cloud in the local Azure region.
  - **Branch-to-Hub to Hub-to-VNet:** VPN connected offices must be able to access applications migrated to the cloud in the remote Azure region.
  - **Branch-to-branch:** Regional VPN connected offices must be able to communicate with each other and ExpressRoute connected HQ/DC sites.
  - **Branch-to-Hub to Hub-to-branch:** Globally separated VPN connected offices must be able to communicate with each other and any ExpressRoute connected HQ/DC sites.
  - **Branch-to-Internet:** Connected sites must be able to communicate with the Internet. This traffic must be filtered and logged.
  - **VNet-to-VNet:** Spoke virtual networks in the same region must be able to communicate with each other.
  - **VNet-to-Hub to Hub-to-VNet:** Spoke virtual networks in the different regions must be able to communicate with each other.
- Provide the ability for Contoso roaming users (laptop and phone) to access company resources while not on the corporate network.

## Azure Virtual WAN architecture

The following figure shows a high-level view of the updated target topology using Azure Virtual WAN to meet the requirements detailed in the previous section.

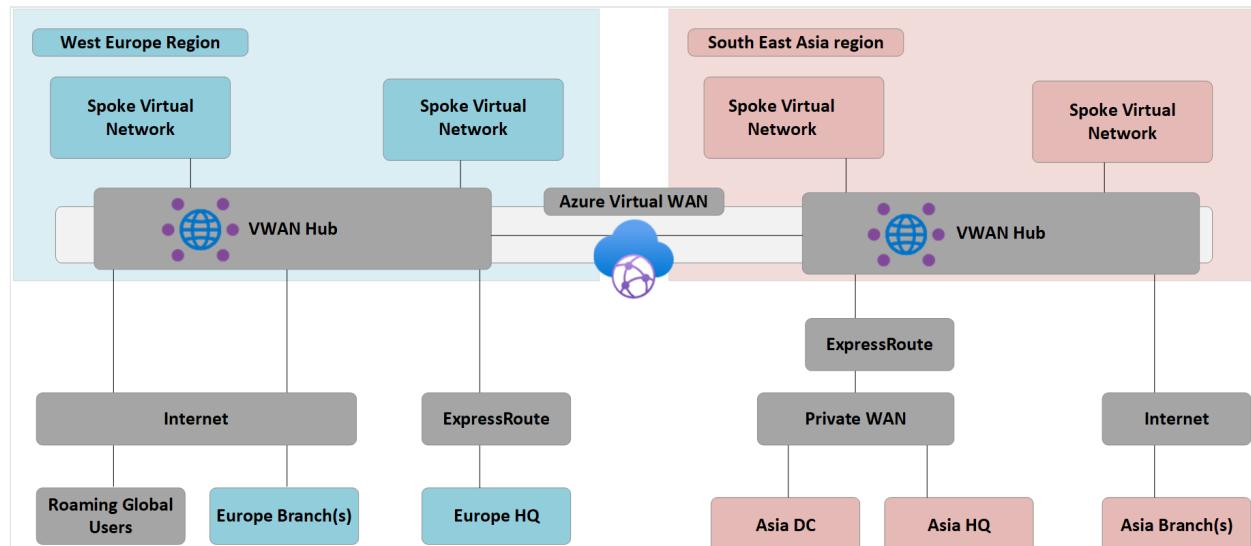


Figure: Azure Virtual WAN architecture

Summary:

- HQ in Europe remains ExpressRoute connected, Europe on-premises DC are fully migrated to Azure and now decommissioned.
- Asia DC and HQ remain connected to Private WAN. Azure Virtual WAN now used to augment the local carrier network and provide global connectivity.
- Azure Virtual WAN hubs deployed in both West Europe and South East Asia Azure regions to provide connectivity hub for ExpressRoute and VPN connected devices.
- Hubs also provide VPN termination for roaming users across multiple client types using OpenVPN connectivity to the global mesh network, allowing access to not only applications migrated to Azure, but also any resources remaining on-premises.
- Internet connectivity for resources within a virtual network provided by Azure Virtual WAN.

Internet connectivity for remote sites also provided by Azure Virtual WAN. Local internet breakout supported via partner integration for optimized access to SaaS services such as Microsoft 365.

## Migrate to Virtual WAN

This section shows the various steps for migrating to Azure Virtual WAN.

### Step 1: Single region customer-managed hub-and-spoke

The following figure shows a single region topology for Contoso prior to the rollout of Azure Virtual WAN:

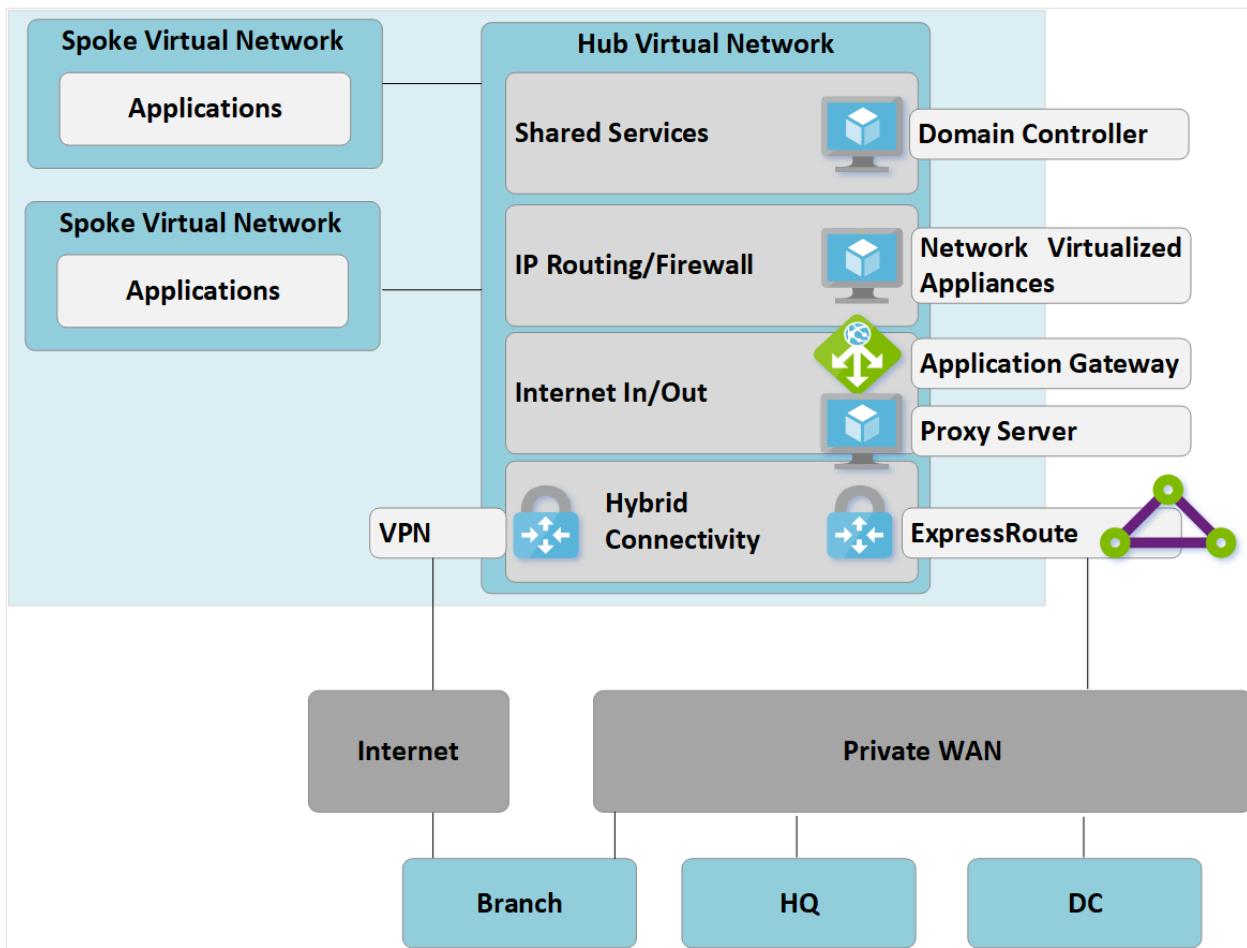


Figure 1: Single region manual hub-and-spoke

In keeping with the hub-and-spoke approach, the customer-managed hub virtual network contains several function blocks:

- Shared services (any common function required by multiple spokes). Example: Contoso uses Windows Server domain controllers on Infrastructure-as-a-service (IaaS) virtual machines.
- IP/Routing firewall services are provided by a third-party network virtual appliance, enabling spoke-to-spoke layer-3 IP routing.
- Internet ingress/egress services including Azure Application Gateway for inbound HTTPS requests and third-party proxy services running on virtual machines for filtered outbound access to internet resources.
- ExpressRoute and VPN virtual network gateway for connectivity to on-premises networks.

## Step 2: Deploy Virtual WAN hubs

Deploy a Virtual WAN hub in each region. Set up the Virtual WAN hub with VPN and ExpressRoute functionality as described in the following articles:

- Tutorial: Create a Site-to-Site connection using Azure Virtual WAN
- Tutorial: Create an ExpressRoute association using Azure Virtual WAN

## ⚠ Note

Azure Virtual WAN must be using the Standard SKU to enable some of the traffic paths shown in this article.

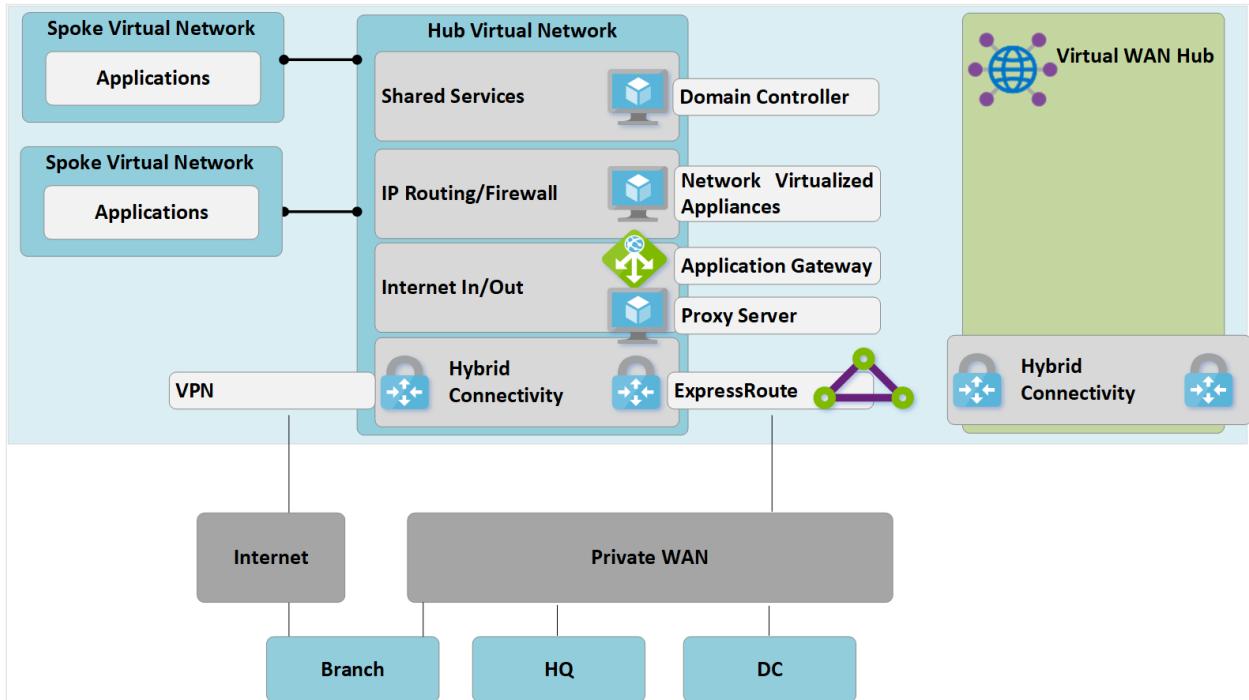


Figure 2: Customer-managed hub-and-spoke to Virtual WAN migration

## Step 3: Connect remote sites (ExpressRoute and VPN) to Virtual WAN

Connect the Virtual WAN hub to the existing ExpressRoute circuits and set up Site-to-site VPNs over the Internet to any remote branches.

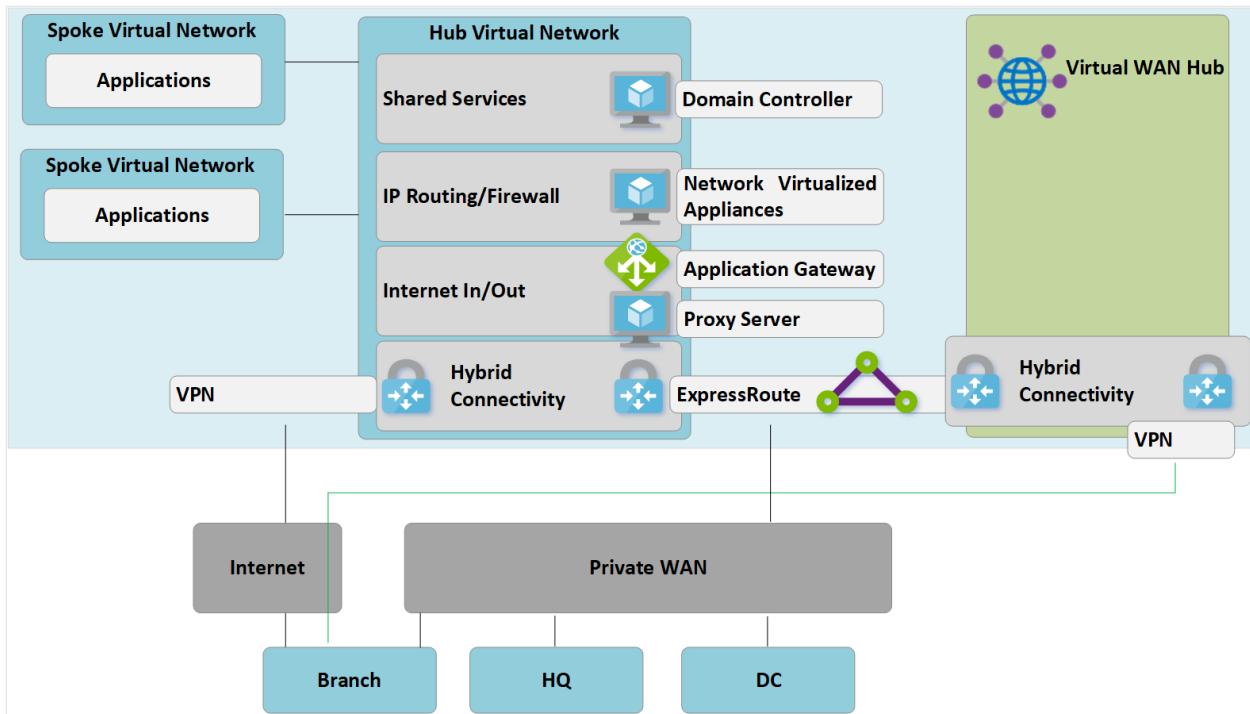
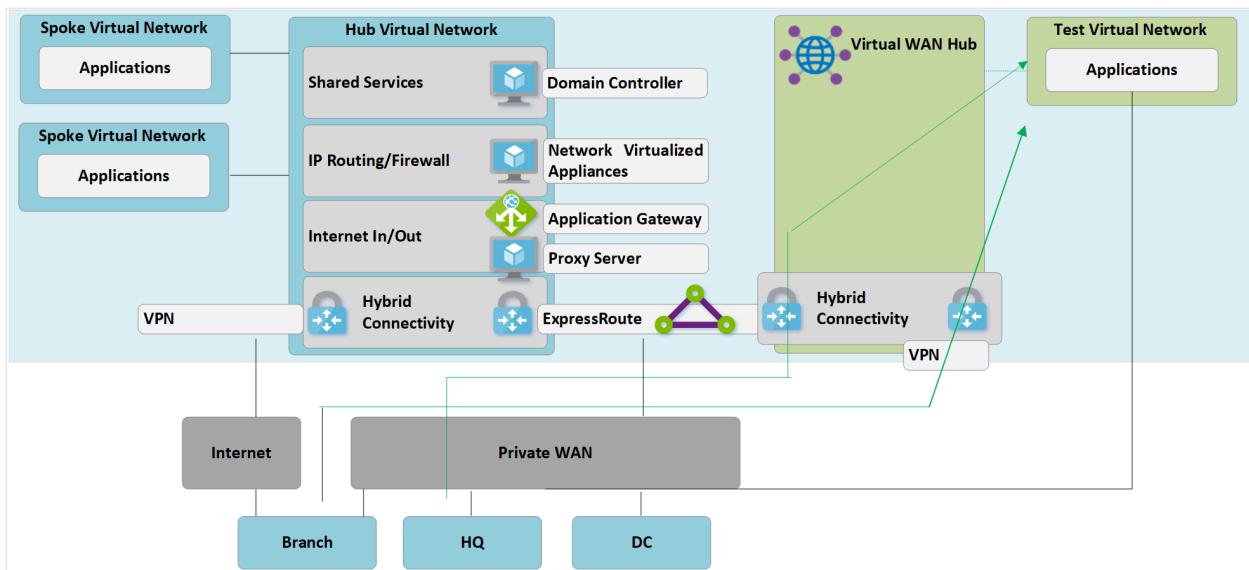


Figure 3: Customer-managed hub-and-spoke to Virtual WAN migration

At this point, on-premises network equipment will begin to receive routes reflecting the IP address space assigned to the Virtual WAN-managed hub VNet. Remote VPN-connected branches at this stage will see two paths to any existing applications in the spoke virtual networks. These devices should be configured to continue to use the tunnel to the customer-managed hub to ensure symmetrical routing during the transition phase.

## Step 4: Test hybrid connectivity via Virtual WAN

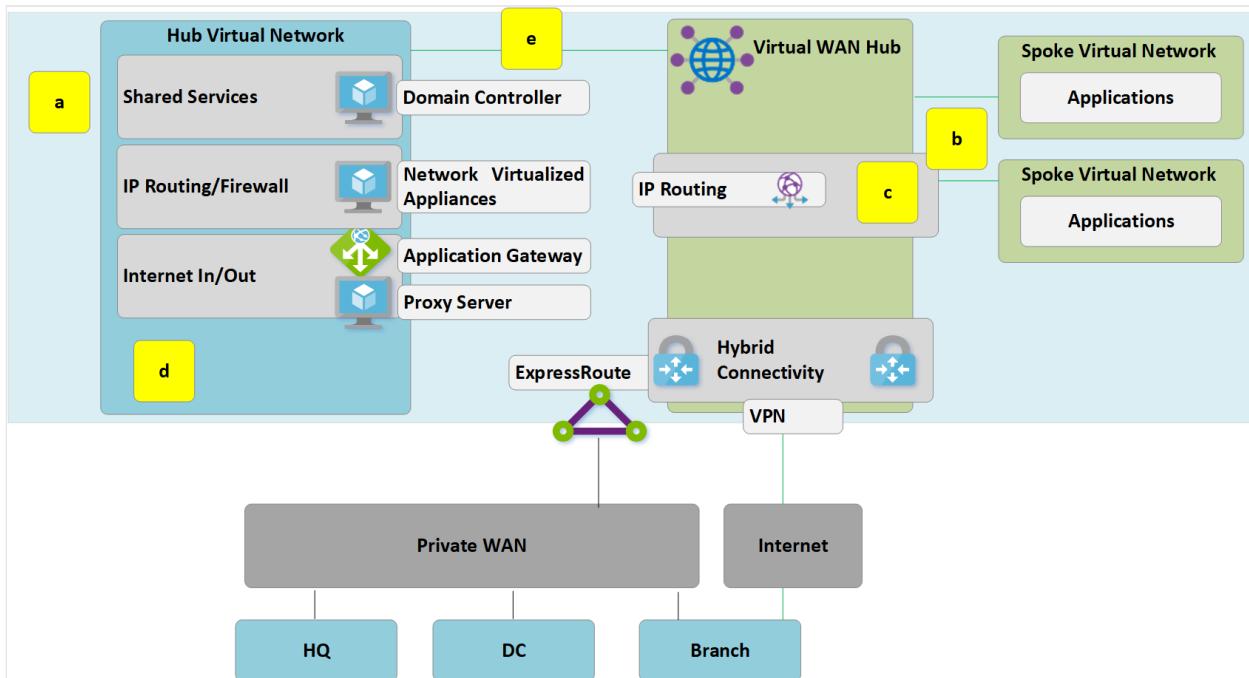
Prior to using the managed Virtual WAN hub for production connectivity, we recommend that you set up a test spoke virtual network and Virtual WAN VNet connection. Validate that connections to this test environment work via ExpressRoute and Site to Site VPN before continuing with the next steps.



**Figure 4: Customer-managed hub-and-spoke to Virtual WAN migration**

At this stage, it's important to recognize that both the original customer-managed hub virtual network and the new Virtual WAN Hub are both connected to the same ExpressRoute circuit. Due to this, we have a traffic path that can be used to enable spokes in both environments to communicate. For example, traffic from a spoke that is attached to the customer-managed hub virtual network will traverse the MSEE devices used for the ExpressRoute circuit to reach any spoke connected via a VNet connection to the new Virtual WAN hub. This allows a staged migration of spokes in Step 5.

## Step 5: Transition connectivity to virtual WAN hub



**Figure 5: Customer-managed hub-and-spoke to Virtual WAN migration**

- Delete the existing peering connections from Spoke virtual networks to the old customer-managed hub. Access to applications in spoke virtual networks is unavailable

until steps a-c are complete.

- b. Connect the spoke virtual networks to the Virtual WAN hub via VNet connections.
- c. Remove any user-defined routes (UDR) previously used within spoke virtual networks for spoke-to-spoke communications. This path is now enabled by dynamic routing available within the Virtual WAN hub.
- d. Existing ExpressRoute and VPN Gateways in the customer-managed hub are now decommissioned to permit the next step (e).
- e. Connect the old customer-managed hub (hub virtual network) to the Virtual WAN hub via a new VNet connection.

## Step 6: Old hub becomes shared services spoke

We have now redesigned our Azure network to make the Virtual WAN hub the central point in our new topology.

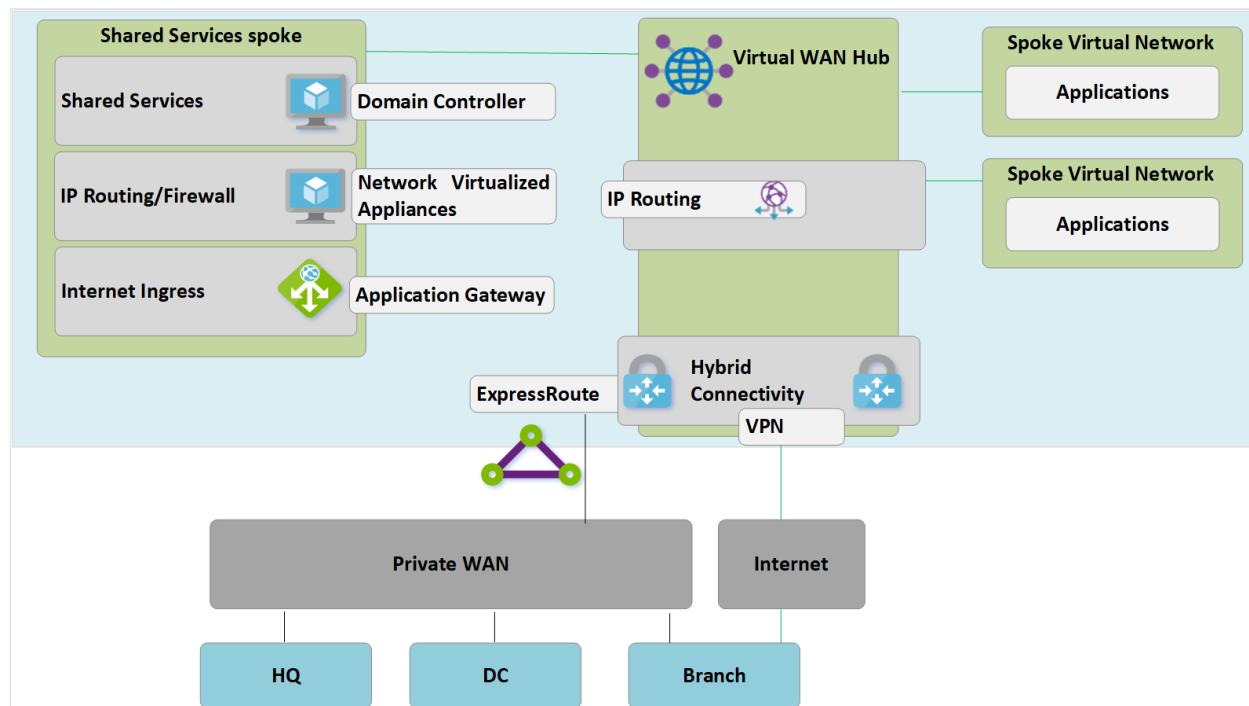


Figure 6: Customer-managed hub-and-spoke to Virtual WAN migration

Because the Virtual WAN hub is a managed entity and doesn't allow deployment of custom resources such as virtual machines, the shared services block now exists as a spoke virtual network and hosts functions such as internet ingress via Azure Application Gateway or network virtualized appliance. Traffic between the shared services environment and backend virtual machines now transits the Virtual WAN-managed hub.

## Step 7: Optimize on-premises connectivity to fully utilize Virtual WAN

At this stage, Contoso has mostly completed their migrations of business applications into the Microsoft Cloud, with only a few legacy applications remaining within the on-premises DC.

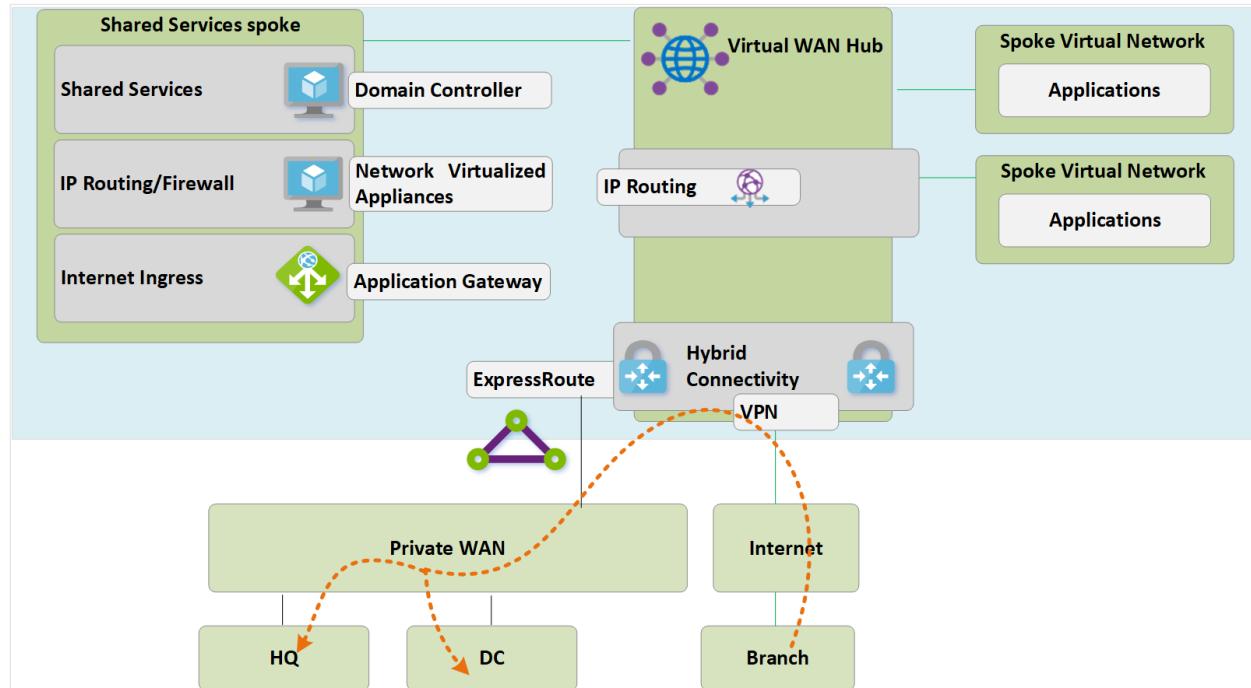


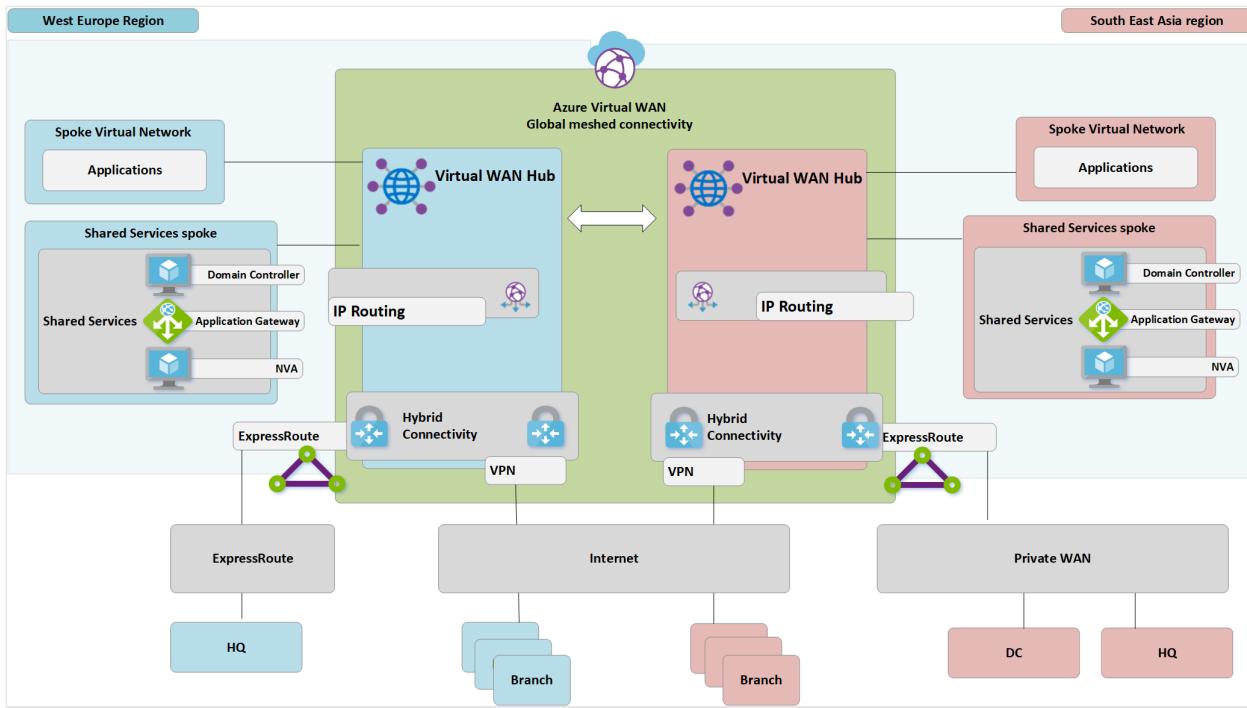
Figure 7: Customer-managed hub-and-spoke to Virtual WAN migration

To leverage the full functionality of Azure Virtual WAN, Contoso decides to decommission their legacy on-premises VPN connections. Any branches continuing to access HQ or DC networks are able to transit the Microsoft global network using the built-in transit routing of Azure Virtual WAN.

### ⓘ Note

ExpressRoute Global Reach is required for customers that want to leverage the Microsoft backbone to provide ExpressRoute to ExpressRoute transit (not shown Figure 7.).

## End-state architecture and traffic paths



**Figure: Dual region Virtual WAN**

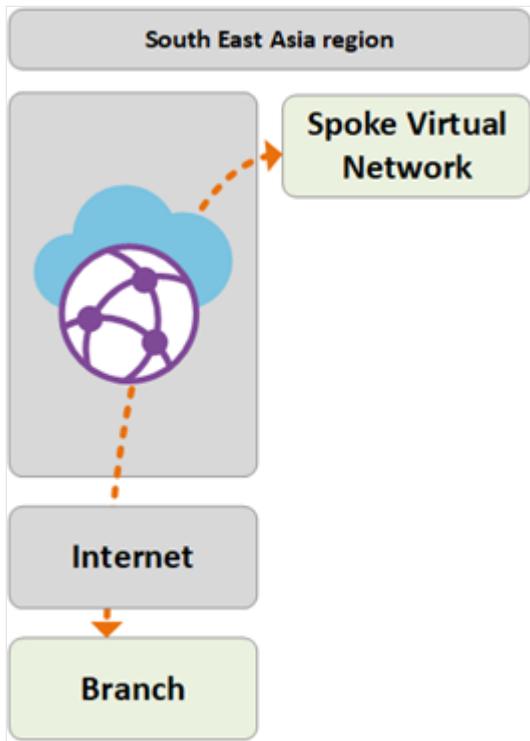
This section provides a summary of how this topology meets the original requirements by looking at some example traffic flows.

## Path 1

Path 1 shows traffic flow from a S2S VPN connected branch in Asia to an Azure VNet in the South East Asia region.

The traffic is routed as follows:

- Asia branch is connected via resilient S2S BGP enabled tunnels into South East Asia Virtual WAN hub.
- Asia Virtual WAN hub routes traffic locally to connected VNet.

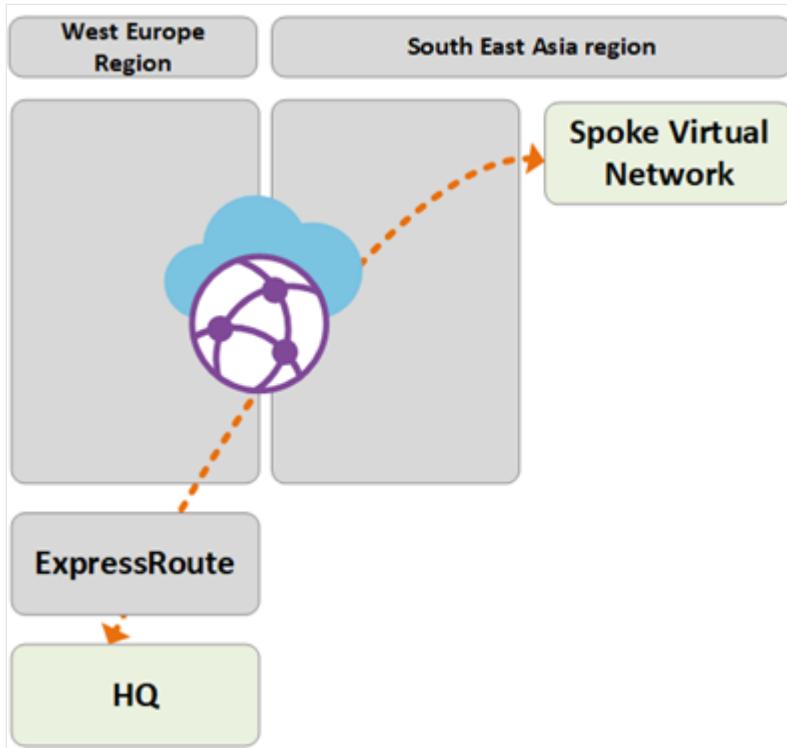


## Path 2

Path 2 shows traffic flow from the ExpressRoute connected European HQ to an Azure VNet in the South East Asia region.

The traffic is routed as follows:

- European HQ is connected via ExpressRoute circuit into West Europe Virtual WAN hub.
- Virtual WAN hub-to-hub global connectivity enables transit of traffic to VNet connected in remote region.

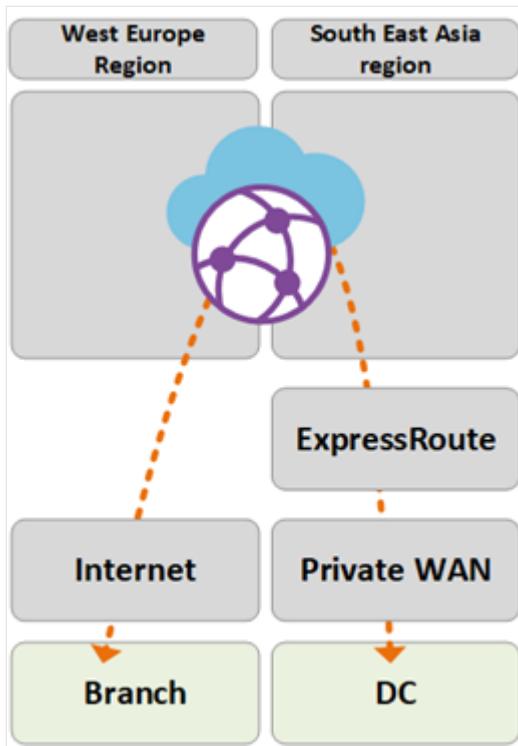


## Path 3

Path 3 shows traffic flow from the Asia on-premises DC connected to Private WAN to a European S2S connected Branch.

The traffic is routed as follows:

- Asia DC is connected to local Private WAN carrier.
- ExpressRoute circuit locally terminates in Private WAN connects to the South East Asia Virtual WAN hub.
- Virtual WAN hub-to-hub global connectivity enables transit of traffic.

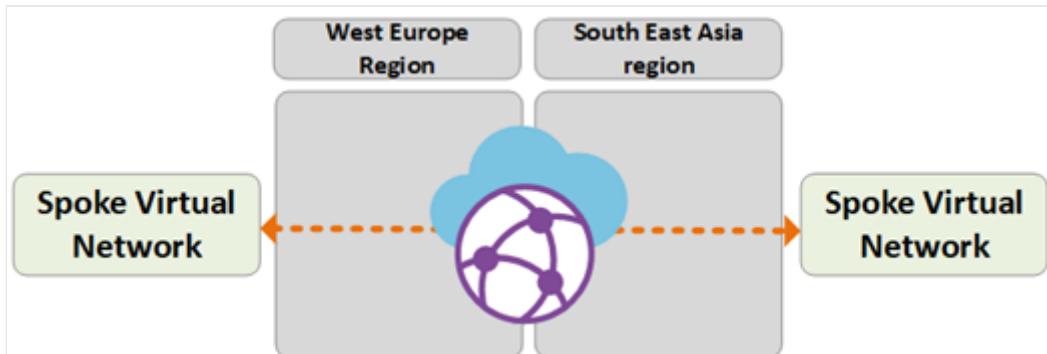


## Path 4

Path 4 shows traffic flow from an Azure VNet in South East Asia region to an Azure VNet in West Europe region.

The traffic is routed as follows:

- Virtual WAN hub-to-hub global connectivity enables native transit of all connected Azure VNets without further user config.

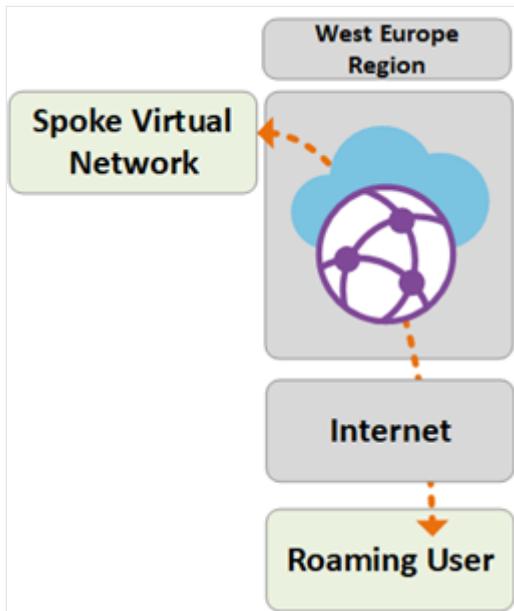


## Path 5

Path 5 shows traffic flow from roaming VPN (P2S) users to an Azure VNet in the West Europe region.

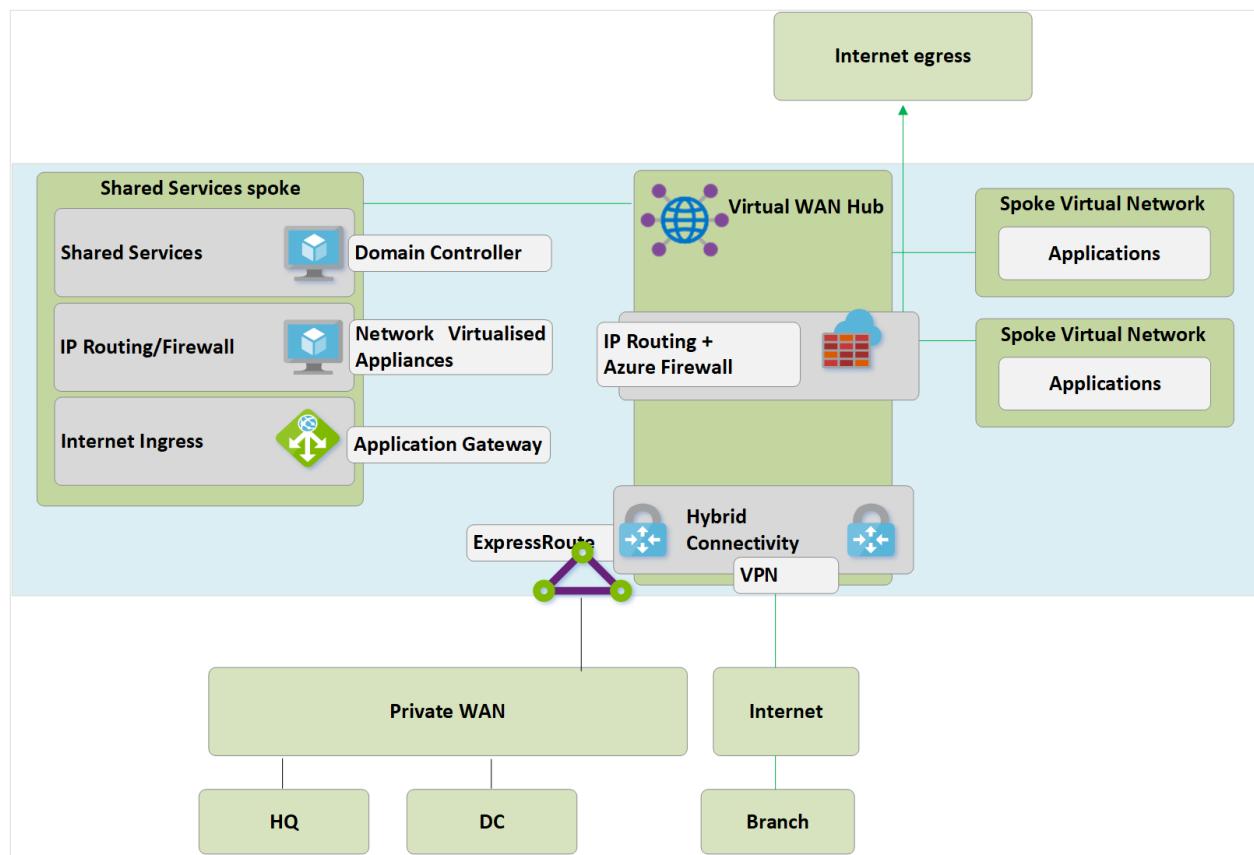
The traffic is routed as follows:

- Laptop and mobile device users use the OpenVPN client for transparent connectivity in to the P2S VPN gateway in West Europe.
- West Europe Virtual WAN hub routes traffic locally to connected VNet.



## Security and policy control via Azure Firewall

Contoso has now validated connectivity between all branches and VNets in line with the requirements discussed earlier in this article. To meet their requirements for security control and network isolation, they need to continue to separate and log traffic via the hub network. Previously this function was performed by a network virtual appliance (NVA). Contoso also wants to decommission their existing proxy services and utilize native Azure services for outbound Internet filtering.



**Figure: Azure Firewall in Virtual WAN (Secured Virtual hub)**

The following high-level steps are required to introduce Azure Firewall into the Virtual WAN hubs to enable a unified point of policy control. For more information about this process and the concept of Secure Virtual Hubs, see [Azure Firewall Manager](#).

1. Create Azure Firewall policy.
2. Link firewall policy to Azure Virtual WAN hub. This step allows the existing Virtual WAN hub to function as a secured virtual hub, and deploys the required Azure Firewall resources.

**① Note**

There are constraints relating to use of secured virtual hubs, including inter-region traffic. For more information, see [Firewall Manager - known issues](#).

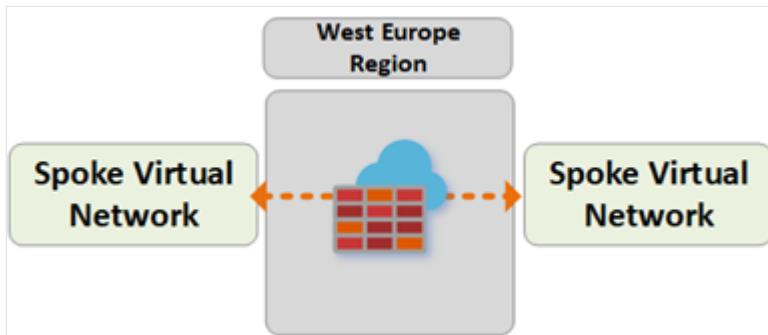
The following paths show the connectivity paths enabled by using Azure secured virtual hubs:

## Path 6

Path 6 shows secure traffic flow between VNets within the same region.

The traffic is routed as follows:

- Virtual Networks connected to the same Secured Virtual Hub now route traffic to via the Azure Firewall.
- Azure Firewall can apply policy to these flows.

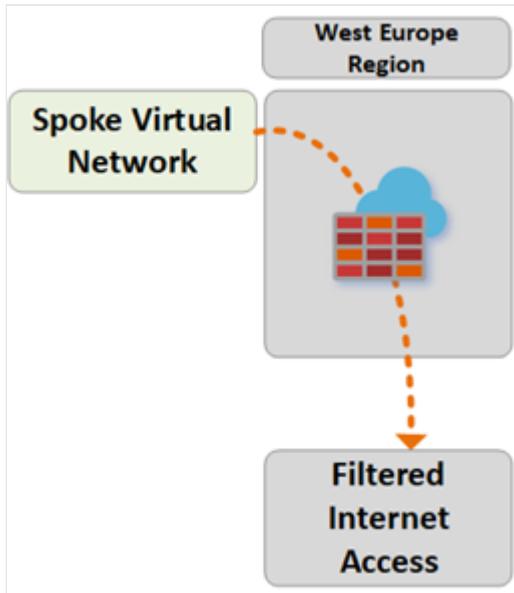


## Path 7

Path 7 shows traffic flow from an Azure VNet to the Internet or third-party Security Service.

The traffic is routed as follows:

- Virtual Networks connected to the Secure Virtual Hub can send traffic to public, destinations on the Internet, using the Secure Hub as a central point of Internet access.
- This traffic can be filtered locally using Azure Firewall FQDN rules, or sent to a third-party security service for inspection.

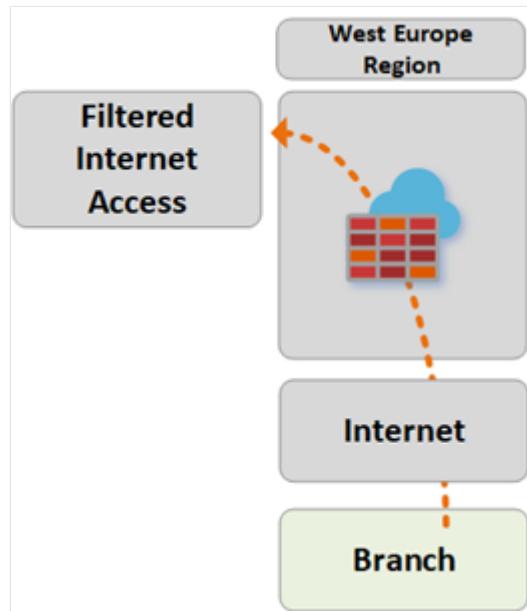


## Path 8

Path 8 shows traffic flow from branch-to-Internet or third-party Security Service.

The traffic is routed as follows:

- Branches connected to the Secure Virtual Hub can send traffic to public destinations on the Internet by using the Secure Hub as a central point of Internet access.
- This traffic can be filtered locally using Azure Firewall FQDN rules, or sent to a third-party security service for inspection.



## Next steps

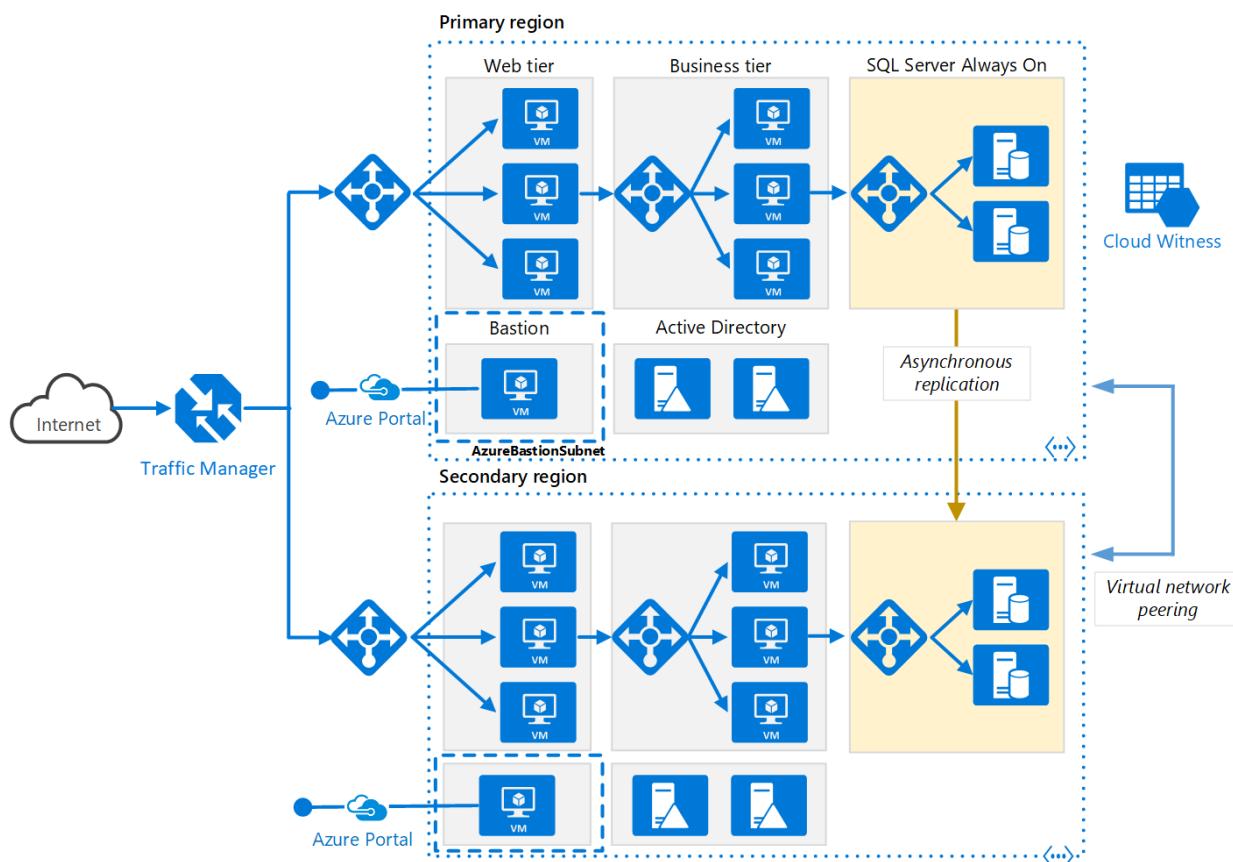
- Learn more about [Azure Virtual WAN](#).
- [Configure Virtual WAN for Azure NetApp Files](#)

# Multi-region N-tier application

Azure Monitor   Azure Traffic Manager   Azure SQL Database   Azure Virtual Machines

This reference architecture shows a set of proven practices for running an N-tier application in multiple Azure regions, in order to achieve availability and a robust disaster recovery infrastructure.

## Architecture



Download a [Visio file](#) of this architecture.

## Workflow

- **Primary and secondary regions.** Use two regions to achieve higher availability. One is the primary region. The other region is for failover.
- **Azure Traffic Manager.** [Traffic Manager](#) routes incoming requests to one of the regions. During normal operations, it routes requests to the primary region. If that region becomes unavailable, Traffic Manager fails over to the secondary region. For more information, see the section [Traffic Manager configuration](#).

- **Resource groups.** Create separate [resource groups](#) for the primary region, the secondary region, and for Traffic Manager. This method gives you the flexibility to manage each region as a single collection of resources. For example, you could redeploy one region, without taking down the other one. [Link the resource groups](#), so that you can run a query to list all the resources for the application.
- **Virtual networks.** Create a separate virtual network for each region. Make sure the address spaces don't overlap.
- **SQL Server Always On Availability Group.** If you use SQL Server, we recommend [SQL Always On Availability Groups](#) for high availability. Create a single availability group that includes the SQL Server instances in both regions.

 **Note**

Also consider [Azure SQL Database](#), which provides a relational database as a cloud service. With SQL Database, you don't need to configure an availability group or manage failover.

- **Virtual network peering.** Peer the two virtual networks to allow data replication from the primary region to the secondary region. For more information, see [Virtual network peering](#).

## Components

- [Availability sets](#) ensure that the VMs you deploy on Azure are distributed across multiple isolated hardware nodes in a cluster. If a hardware or software failure occurs within Azure, only a subset of your VMs are affected, and your entire solution remains available and operational.
- [Availability zones](#) protect your applications and data from datacenter failures. Availability zones are separate physical locations within an Azure region. Each zone consists of one or more datacenters equipped with independent power, cooling, and networking.
- [Azure Traffic Manager](#) is a DNS-based traffic load balancer that distributes traffic optimally. It provides services across global Azure regions, with high availability and responsiveness.
- [Azure Load Balancer](#) distributes inbound traffic, according to defined rules and health probes. A load balancer provides low latency and high throughput, scaling up to millions of flows for all TCP and UDP applications. A public load balancer is used in this scenario, to distribute incoming client traffic to the web tier. An

internal load balancer is used in this scenario, to distribute traffic from the business tier to the back-end SQL Server cluster.

- [Azure Bastion](#) provides secure RDP and SSH connectivity to all of the VMs, in the virtual network in which it's provisioned. Use Azure Bastion to protect your virtual machines from exposing RDP/SSH ports to the outside world, while still providing secure access using RDP/SSH.

## Recommendations

A multi-region architecture can provide higher availability than deploying to a single region. If a regional outage affects the primary region, you can use [Traffic Manager](#) to fail over to the secondary region. This architecture can also help if an individual subsystem of the application fails.

There are several general approaches to achieving high availability across regions:

- Active/passive with hot standby. Traffic goes to one region, while the other waits on hot standby. Hot standby means the VMs in the secondary region are allocated and are always running.
- Active/passive with cold standby. Traffic goes to one region, while the other waits on cold standby. Cold standby means the VMs in the secondary region aren't allocated until needed for failover. This approach costs less to run, but will generally take longer to come online during a failure.
- Active/active. Both regions are active, and requests are load balanced between them. If one region becomes unavailable, it's taken out of rotation.

This reference architecture focuses on active/passive with hot standby, using Traffic Manager for failover. You could deploy a few VMs for hot standby and then scale out as needed.

## Regional pairing

Each Azure region is paired with another region within the same geography. In general, choose regions from the same regional pair (for example, East US 2 and US Central). Benefits of doing so include:

- If there's a broad outage, recovery of at least one region out of every pair is prioritized.
- Planned Azure system updates are rolled out to paired regions sequentially, to minimize possible downtime.
- Pairs reside within the same geography, to meet data residency requirements.

However, make sure that both regions support all of the Azure services needed for your application (see [Services by region](#)). For more information about regional pairs, see [Business continuity and disaster recovery \(BCDR\): Azure Paired Regions](#).

## Traffic Manager configuration

Consider the following points when configuring Traffic Manager:

- **Routing.** Traffic Manager supports several [routing algorithms](#). For the scenario described in this article, use *priority* routing (formerly called *failover* routing). With this setting, Traffic Manager sends all requests to the primary region, unless the primary region becomes unreachable. At that point, it automatically fails over to the secondary region. See [Configure Failover routing method](#).
- **Health probe.** Traffic Manager uses an HTTP (or HTTPS) [probe](#) to monitor the availability of each region. The probe checks for an HTTP 200 response for a specified URL path. As a best practice, create an endpoint that reports the overall health of the application, and use this endpoint for the health probe. Otherwise, the probe might report a healthy endpoint when critical parts of the application are actually failing. For more information, see [Health Endpoint Monitoring pattern](#).

When Traffic Manager fails over, there's a period of time when clients can't reach the application. The duration is affected by the following factors:

- The health probe must detect that the primary region has become unreachable.
- DNS servers must update the cached DNS records for the IP address, which depends on the DNS time-to-live (TTL). The default TTL is 300 seconds (5 minutes), but you can configure this value when you create the Traffic Manager profile.

For details, see [About Traffic Manager Monitoring](#).

If Traffic Manager fails over, we recommend performing a manual failback rather than implementing an automatic failback. Otherwise, you can create a situation where the application flips back and forth between regions. Verify that all application subsystems are healthy before failing back.

Traffic Manager automatically fails back by default. To prevent this issue, manually lower the priority of the primary region after a failover event. For example, suppose the primary region is priority 1 and the secondary is priority 2. After a failover, set the primary region to priority 3, to prevent automatic failback. When you're ready to switch back, update the priority to 1.

The following [Azure CLI](#) command updates the priority:

Azure CLI

```
az network traffic-manager endpoint update --resource-group <resource-group>
--profile-name <profile>
--name <endpoint-name> --type azureEndpoints --priority 3
```

Another approach is to temporarily disable the endpoint until you're ready to fail back:

Azure CLI

```
az network traffic-manager endpoint update --resource-group <resource-group>
--profile-name <profile>
--name <endpoint-name> --type azureEndpoints --endpoint-status Disabled
```

Depending on the cause of a failover, you might need to redeploy the resources within a region. Before failing back, perform an operational readiness test. The test should verify things like:

- VMs are configured correctly. (All required software is installed, IIS is running, and so on.)
- Application subsystems are healthy.
- Functional testing. (For example, the database tier is reachable from the web tier.)

## Configure SQL Server Always On Availability Groups

Prior to Windows Server 2016, SQL Server Always On Availability Groups require a domain controller, and all nodes in the availability group must be in the same Active Directory (AD) domain.

To configure the availability group:

- At a minimum, place two domain controllers in each region.
- Give each domain controller a static IP address.
- Peer the two virtual networks to enable communication between them.
- For each virtual network, add the IP addresses of the domain controllers (from both regions) to the DNS server list. You can use the following CLI command. For more information, see [Change DNS servers](#).

Azure CLI

```
az network vnet update --resource-group <resource-group> --name <vnet-
```

```
name> --dns-servers "10.0.0.4,10.0.0.6,172.16.0.4,172.16.0.6"
```

- Create a [Windows Server Failover Clustering](#) (WSFC) cluster that includes the SQL Server instances in both regions.
- Create a SQL Server Always On Availability Group that includes the SQL Server instances in both the primary and secondary regions. See [Extending Always On Availability Group to Remote Azure Datacenter \(PowerShell\)](#) for the steps.
  - Put the primary replica in the primary region.
  - Put one or more secondary replicas in the primary region. Configure these replicas to use *synchronous* commit with automatic failover.
  - Put one or more secondary replicas in the secondary region. Configure these replicas to use *asynchronous* commit, for performance reasons. (Otherwise, all T-SQL transactions have to wait on a round trip over the network to the secondary region.)

 **Note**

Asynchronous commit replicas don't support automatic failover.

## Considerations

These considerations implement the pillars of the Azure Well-Architected Framework, which is a set of guiding tenets that can be used to improve the quality of a workload. For more information, see [Microsoft Azure Well-Architected Framework](#).

## Availability

With a complex N-tier app, you may not need to replicate the entire application in the secondary region. Instead, you might just replicate a critical subsystem that is needed to support business continuity.

Traffic Manager is a possible failure point in the system. If the Traffic Manager service fails, clients can't access your application during the downtime. Review the [Traffic Manager SLA](#), and determine whether using Traffic Manager alone meets your business requirements for high availability. If not, consider adding another traffic management solution as a fallback. If the Azure Traffic Manager service fails, change your CNAME records in DNS to point to the other traffic management service. (This step

must be performed manually, and your application will be unavailable until the DNS changes are propagated.)

For the SQL Server cluster, there are two failover scenarios to consider:

- All of the SQL Server database replicas in the primary region fail. For example, this failure could happen during a regional outage. In that case, you must manually fail over the availability group, even though Traffic Manager automatically fails over on the front end. Follow the steps in [Perform a Forced Manual Failover of a SQL Server Availability Group](#), which describes how to perform a forced failover by using SQL Server Management Studio, Transact-SQL, or PowerShell in SQL Server 2016.

### Warning

With forced failover, there's a risk of data loss. Once the primary region is back online, take a snapshot of the database and use **tablediff** to find the differences.

- Traffic Manager fails over to the secondary region, but the primary SQL Server database replica is still available. For example, the front-end tier might fail, without affecting the SQL Server VMs. In that case, Internet traffic is routed to the secondary region, and that region can still connect to the primary replica. However, there will be increased latency, because the SQL Server connections are going across regions. In this situation, you should perform a manual failover as follows:
  1. Temporarily switch a SQL Server database replica in the secondary region to *synchronous* commit. This step ensures there won't be data loss during the failover.
  2. Fail over to that replica.
  3. When you fail back to the primary region, restore the *asynchronous* commit setting.

## Manageability

When you update your deployment, update one region at a time to reduce the chance of a global failure from an incorrect configuration or an error in the application.

Test the resiliency of the system to failures. Here are some common failure scenarios to test:

- Shut down VM instances.
- Pressure resources such as CPU and memory.

- Disconnect/delay network.
- Crash processes.
- Expire certificates.
- Simulate hardware faults.
- Shut down the DNS service on the domain controllers.

Measure the recovery times and verify they meet your business requirements. Test combinations of failure modes, as well.

## Cost optimization

Cost optimization is about looking at ways to reduce unnecessary expenses and improve operational efficiencies. For more information, see [Overview of the cost optimization pillar](#).

Use the [Azure Pricing Calculator](#) to estimate costs. Here are some other considerations.

## Virtual Machine Scale Sets

Virtual Machine Scale Sets are available on all Windows VM sizes. You're only charged for the Azure VMs you deploy and any added underlying infrastructure resources that are consumed, such as storage and networking. There are no incremental charges for the Virtual Machine Scale Sets service.

For single VMs pricing options, see [Windows VMs pricing](#).

## SQL server

If you choose Azure SQL DBaaS, you can save on cost because don't need to configure an Always On Availability Group and domain controller machines. There are several deployment options starting from single database up to managed instance, or elastic pools. For more information, see [Azure SQL pricing](#).

For SQL server VMs pricing options, see [SQL VMs pricing](#).

## Load balancers

You're only charged for the number of configured load-balancing and outbound rules. Inbound NAT rules are free. There's no hourly charge for the Standard Load Balancer when no rules are configured.

## Traffic Manager pricing

Traffic Manager billing is based on the number of DNS queries received, with a discount for services receiving more than 1 billion monthly queries. You're also charged for each monitored endpoint.

For more information, see the cost section in [Microsoft Azure Well-Architected Framework](#).

## VNET-Peering pricing

A high-availability deployment that uses multiple Azure Regions will make use of VNET-Peering. There are different charges for VNET-Peering within the same region and for Global VNET-Peering.

For more information, see [Virtual Network Pricing](#).

## DevOps

Use a single [Azure Resource Manager template](#) for provisioning the Azure resources and its dependencies. Use the same template to deploy the resources to both primary and secondary regions. Include all the resources in the same virtual network so they're isolated in the same basic workload. By including all the resources, you make it easier to associate the workload's specific resources to a DevOps team, so that the team can independently manage all aspects of those resources. This isolation enables DevOps Team and Services to perform continuous integration and continuous delivery (CI/CD).

Also, you can use different [Azure Resource Manager templates](#) and integrate them with [Azure DevOps Services](#) to provision different environments in minutes, for example to replicate production like scenarios or load testing environments only when needed, saving cost.

Consider using the [Azure Monitor](#) to Analyze and optimize the performance of your infrastructure, Monitor and diagnose networking issues without logging into your virtual machines. Application Insights is actually one of the components of Azure Monitor, which gives you rich metrics and logs to verify the state of your complete Azure landscape. Azure Monitor will help you to follow the state of your infrastructure.

Make sure not only to monitor your compute elements supporting your application code, but your data platform as well, in particular your databases, since a low performance of the data tier of an application could have serious consequences.

In order to test the Azure environment where the applications are running, it should be version-controlled and deployed through the same mechanisms as application code, then it can be tested and validated using DevOps testing paradigms too.

For more information, see the Operational Excellence section in [Microsoft Azure Well-Architected Framework](#).

## Contributors

*This article is maintained by Microsoft. It was originally written by the following contributors.*

Principal author:

- [Donnie Trumpower](#) | Senior Cloud Solution Architect

*To see non-public LinkedIn profiles, sign in to LinkedIn.*

## Next steps

- [Deploy Traffic Manager in Azure](#)
- [Deploy Azure Load Balancer](#)

## Related resources

The following architecture uses some of the same technologies:

- [Multitier web application built for high availability and disaster recovery on Azure](#)
- [Multi-region load balancing](#)

# Multi-region load balancing with Traffic Manager, Azure Firewall, and Application Gateway

Azure Firewall

Azure Application Gateway

Azure Bastion

Azure Load Balancer

Azure Traffic Manager

This architecture is for global, internet-facing applications that use HTTP(S) and non-HTTP(S) protocols. It features DNS-based global load balancing, two forms of regional load balancing, and global virtual network peering to create a high availability architecture that can withstand a regional outage. Traffic inspection is provided by both Azure Web Application Firewall (WAF) and Azure Firewall.

## Architecture notes

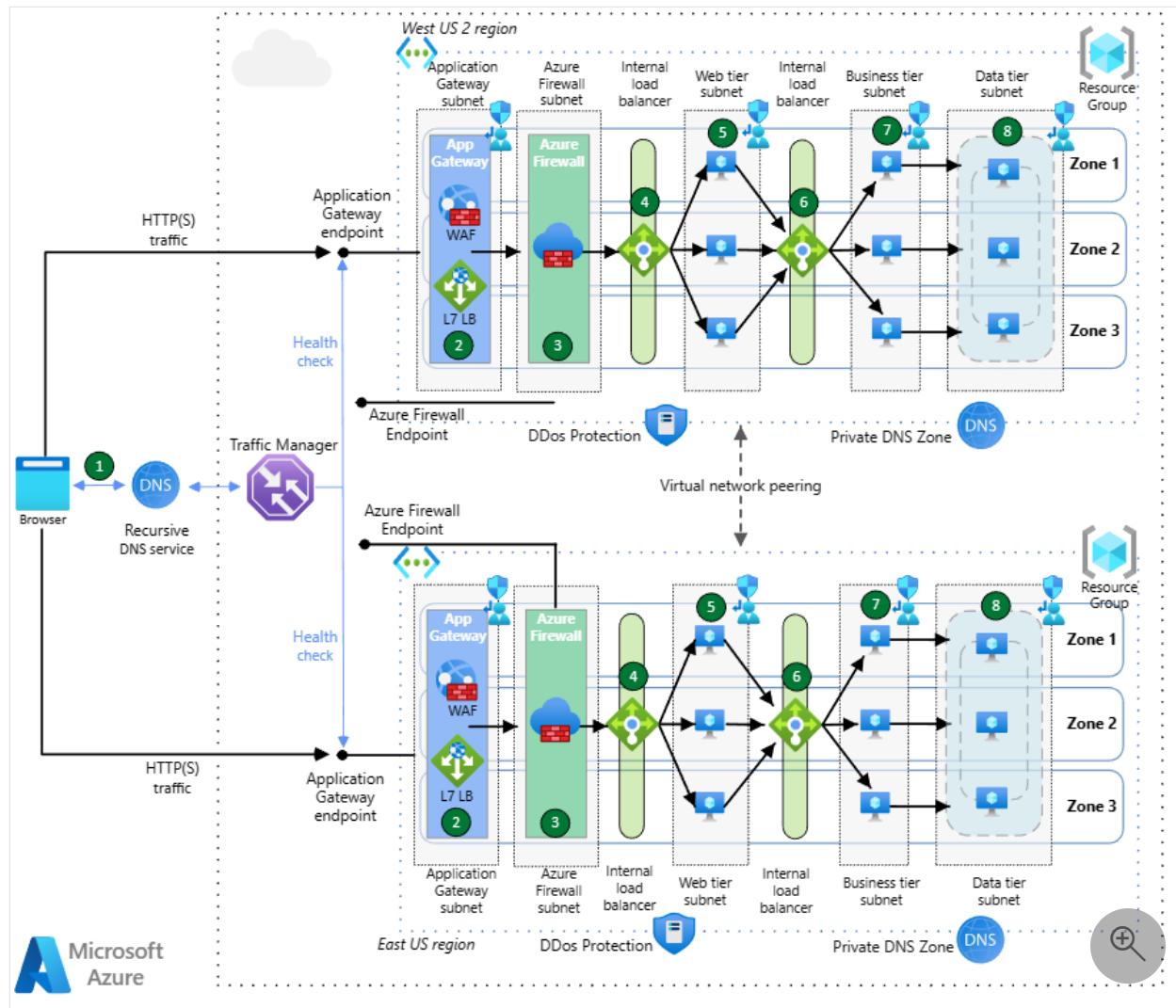
The architecture in this document is easily extensible to a hub-and-spoke virtual network design, where the Azure Firewall would be in the hub network, and the Application Gateway either in the hub network as well or in a spoke. If the Application Gateway is deployed in the hub, you still want multiple Application Gateways, each for a given set of applications, to avoid RBAC conflicts and prevent hitting Application Gateway limits (see [Application Gateway Limits](#)).

In a Virtual WAN environment Application Gateways cannot be deployed in the hub, so they would be installed in spoke virtual networks.

The proposed architecture opts for double inspection of web content through both a Web Application Firewall (based on Application Gateway) in front of Azure Firewall. Other options exist, as documented in [Firewall and Application Gateway for virtual networks](#), but this option is the most flexible and complete one: it exposes the client's IP address in the HTTP header `X-Forwarded-For` for the end application, it provides end-to-end encryption, and it prevents clients from bypassing the WAF to access the application.

If only web applications are exposed (no non-HTTP(S) applications), and the double inspection by WAF and Azure Firewall of this web traffic is not required, Azure Front Door would be a better global load balancing solution than Traffic Manager. Front Door is a layer-7 load balancer for HTTP(S) content that also provides caching, traffic acceleration, SSL/TLS termination, certificate management, health probes, and other capabilities. However, Application Gateway offers better integration with Azure Firewall for a layered protection approach.

# Inbound HTTP(S) traffic flows

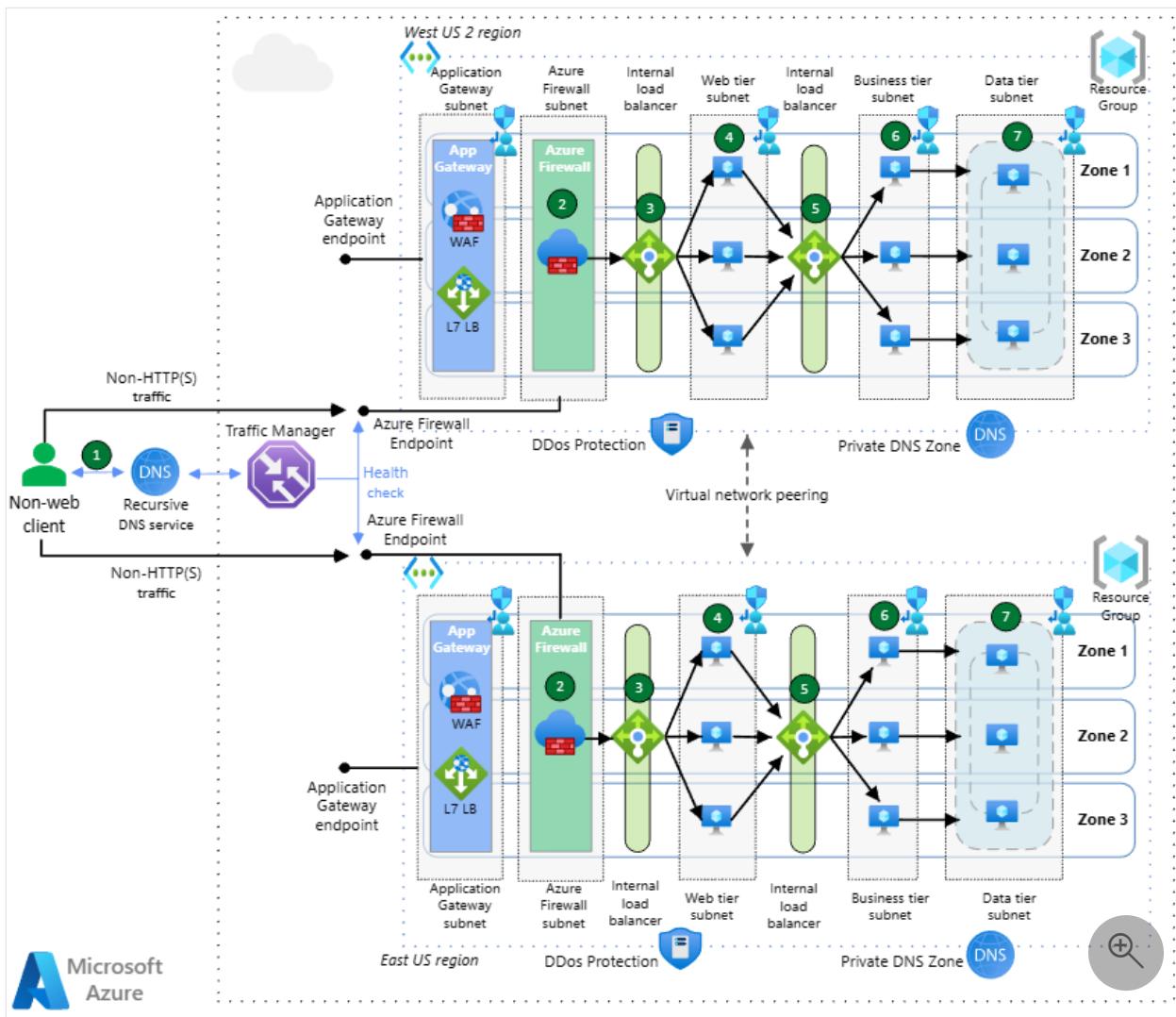


Download a [Visio file](#) of this architecture.

1. Azure Traffic Manager uses DNS-based routing to load balance incoming traffic across the two regions. Traffic Manager resolves DNS queries for the application to the public IP addresses of the Azure Application Gateway endpoints. The public endpoints of the Application Gateways serve as the backend endpoints of Traffic Manager for HTTP(S) traffic. Traffic Manager resolves DNS queries based on a choice of various routing methods. The browser connects directly to the endpoint; **Traffic Manager doesn't see the HTTP(S) traffic.**
2. The Application Gateways deployed across availability zones receive HTTP(S) traffic from the browser, and the Web Application Firewalls Premium inspect the traffic to detect web attacks. The Application Gateways will send traffic to their backend, the internal load balancer for the frontend virtual machines. For this specific flow, the internal load balancer in front of the web servers is not strictly required since the Application Gateway could perform this load balancing itself. However, it is included for consistency with the flow for non-HTTP(S) applications.

3. The traffic between the Application Gateway and the frontend internal load balancer will be intercepted by Azure Firewall Premium via User Defined Routes applied on the Application Gateway subnet. The Azure Firewall Premium will apply TLS inspection to the traffic for additional security. The Azure Firewall is zone-redundant as well. If the Azure Firewall detects a threat in the traffic, it will drop the packets. Otherwise, upon successful inspection the Azure Firewall will forward the traffic to the destination web-tier internal load balancer.
4. The web-tier is the first layer of the three-tier application, it contains the user interface and it also parses user interactions. The web-tier load balancer is spread over all three availability zones, and it will distribute traffic to each of the three web-tier virtual machines.
5. The web-tier virtual machines are spread across all three availability zones, and they will communicate with the business tier via a dedicated internal load balancer.
6. The business tier processes the user interactions and determines the next steps, and it sits between the web and data tiers. The business-tier internal load balancer distributes traffic to the business-tier virtual machines across the three availability zones. The business-tier load balancer is zone-redundant, like the web-tier load balancer.
7. The business-tier virtual machines are spread across availability zones, and they will route traffic to the availability group listener of the databases.
8. The data-tier stores the application data, typically in a database, object storage, or file share. This architecture has SQL server on virtual machines distributed across three availability zones. They are in an availability group and use a distributed network name (DNN) to route traffic to the [availability group listener](#) for load balancing.

## Inbound non-HTTP(S) traffic flows



[Download a Visio file](#) of this architecture.

1. Azure Traffic Manager uses DNS-based routing to load balance incoming traffic across the two regions. Traffic Manager resolves DNS queries for the application to the public IP addresses of the Azure endpoints. The public endpoints of the Application Firewall serve as the backend endpoints of Traffic Manager for non-HTTP(S) traffic. Traffic Manager resolves DNS queries based on a choice of various routing methods. The browser connects directly to the endpoint; **Traffic Manager doesn't see the HTTP(S) traffic.**
2. The Azure Firewall Premium is zone-redundant, and it will inspect the inbound traffic for security. If the Azure Firewall detects a threat in the traffic, it will drop the packets. Otherwise, upon successful inspection the Azure Firewall will forward the traffic to the web-tier internal load balancer performing Destination Network Address Translation (DNAT) on the inbound packets.
3. The web-tier is the first layer of the three-tier application, it contains the user interface and it also parses user interactions. The web-tier load balancer is spread over all three availability zones, and it will distribute traffic to each of the three web-tier virtual machines.

4. The web-tier virtual machines are spread across all three availability zones, and they will communicate with the business tier via a dedicated internal load balancer.
5. The business tier processes the user interactions and determines the next steps, and it sits between the web and data tiers. The business-tier internal load balancer distributes traffic to the business-tier virtual machines across the three availability zones. The business-tier load balancer is zone-redundant, like the web-tier load balancer.
6. The business-tier virtual machines are spread across availability zones, and they will route traffic to the availability group listener of the databases.
7. The data-tier stores the application data, typically in a database, object storage, or file share. This architecture has SQL server on virtual machines distributed across three availability zones. They are in an availability group and use a distributed network name (DNN) to route traffic to the [availability group listener](#) for load balancing.

## Outbound traffic flows (all protocols)

Outbound traffic flows for virtual machine patch updates or other connectivity to the Internet will go from the workload virtual machines to the Azure Firewall through User-Defined Routes. The Azure Firewall will enforce connectivity rules using web categories as well as network and application rules to prevent workloads from accessing inappropriate content or data exfiltration scenarios.

## Components

- [Azure Firewall](#) is a cloud-based, Microsoft-managed next-generation firewall that provides deep packet inspection for both North/South and East/West traffic flows. It can be spread across Availability Zones and it offers automatic autoscaling to cope with application demand changes.
- [Azure Application Gateway](#) is a layer-7 load balancer with optional Web Application Firewall (WAF) functionality. The v2 SKU of Application Gateway supports availability zone redundancy and it is recommended for most scenarios. The Application Gateway includes configurable horizontal autoscaling so that it can react automatically to application demand changes.
- [Azure Traffic Manager](#) is a DNS-based global traffic load balancer that distributes traffic to services across global Azure regions while providing high availability and responsiveness. For more information, see the section [Traffic Manager configuration](#).

- [Azure Load Balancer](#) is a layer-4 load balancer. A zone-redundant load balancer will still distribute traffic with an availability zone failure to the remaining zones.
- [Azure DDoS Protection](#) has enhanced features to protect against distributed denial of service (DDoS) attacks.
- [Azure DNS](#) is a hosting service for DNS domains. It provides name resolution using Microsoft Azure infrastructure. By hosting your domains in Azure, you can manage your DNS records using the same credentials, APIs, tools, and billing as your other Azure services.
- [Azure Private DNS zones](#) are a feature of Azure DNS. Azure DNS Private Zones provide name resolution within a virtual network, and between virtual networks. The records contained in a private DNS zone aren't resolvable from the Internet. DNS resolution against a private DNS zone works only from virtual networks linked to it.
- [Azure Virtual Machines](#) are on-demand, scalable computing resources that give you the flexibility of virtualization but eliminate the maintenance demands of physical hardware. The operating system choices include Windows and Linux. Certain components of the applications can be replaced with platform-as-a-service Azure resources (for example the database and the frontend tier), but the architecture wouldn't change significantly if using [Private Link](#) and [App Service VNet Integration](#) to bring those PaaS services into the virtual network.
- [Azure Virtual Machine Scale Sets](#) is automated and load-balanced virtual machine scaling that simplifies management of your applications and increases availability.
- [SQL Server on VMs](#) lets you use full versions of SQL Server in the cloud without having to manage any on-premises hardware.
- [Azure Virtual Network](#) is a secure private network in the cloud. It connects virtual machines to one another, to the Internet, and to cross-premises networks.
- [User-Defined Routes](#) are a mechanism to override the default routing in virtual networks. In this scenario they are used to force traffic inbound and outbound traffic flows to traverse the Azure Firewall.

## Solution details

*Traffic Manager* - We configured Traffic Manager to use performance routing. It routes traffic to the endpoint that has the lowest latency for the user. Traffic Manager automatically adjusts its load balancing algorithm as endpoint latency changes. Traffic manager provides automatic failover if there's a regional outage. It uses priority routing and regular health checks to determine where to route traffic.

*Availability Zones* - The architecture uses three availability zones. The zones create a high-availability architecture for the Application Gateways, internal load balancers, and

virtual machines in each region. If there is a zone outage, the remaining availability zones in that region would take over the load, which wouldn't trigger a regional failover.

*Application Gateway* - While Traffic Manager provides DNS-based regional load balancing, Application Gateway gives you many of the same capabilities as Azure Front Door but at the regional level such as:

- Web Application Firewall (WAF)
- Transport Layer Security (TLS) termination
- Path-based routing
- Cookie-based session affinity

*Azure Firewall* - Azure Firewall Premium offers network security for generic applications (web and non-web traffic), inspecting three types of flows in this architecture:

- Inbound HTTP(S) flows from the Application Gateway are protected with Azure Firewall Premium TLS inspection.
- Inbound non-HTTP(S) flows from the public Internet are inspected with the rest of [Azure Firewall Premium features](#).
- Outbound flows from Azure Virtual Machines are inspected by Azure Firewall to prevent data exfiltration and access to forbidden sites and applications.

*Virtual network peering* - We call peering between regions "global virtual network peering." Global virtual network peering provides low-latency, high-bandwidth data replication between regions. You can transfer data across Azure subscriptions, Microsoft Entra tenants, and deployment models with this global peering. In hub-spoke environment virtual network peerings would exist between hub and spoke networks.

## Recommendations

The following recommendations adhere to the pillars of the Azure Well-Architected Framework (WAF). The WAF pillars are guiding tenets that help ensure the quality of cloud workloads. For more information, see [Microsoft Azure Well-Architected Framework](#).

## Reliability

*Regions* - Use at least two Azure regions for high availability. You can deploy your application across multiple Azure regions in active/passive or active/active configurations. Multiple regions also help avoid application downtime if a subsystem of the application fails.

- Traffic Manager will automatically fail over to the secondary region if the primary region fails.
- Choosing the best regions for your needs must be based on technical, regulatory considerations, and availability-zone support.

*Region pairs* - Use Region Pairs for the most resiliency. Make sure that both Region Pairs support all the Azure services that your application needs (see [services by region](#) ↗).

Here are two benefits of Region Pairs:

- Planned Azure updates roll out to paired regions one at a time to minimize downtime and risk of application outage.
- Data continues to reside within the same geography as its pair (except for Brazil South) for tax and legal purposes.

*Availability zones* - Use multiple availability zones to support your Application Gateway, Azure Firewall, Azure Load Balancer, and application tiers when available.

*Application gateway autoscaling and instances* - Configure the Application Gateway with a minimum of two instances to avoid downtime, and autoscaling to provide dynamic adaptation to changing application capacity demands.

For more information, see:

- [Regions and availability zones in Azure](#)
- [Business continuity and disaster recovery \(BCDR\): Azure Paired Regions](#)

## Global routing

*Global routing method* - Use the traffic-routing method that best meets the needs of your customers. Traffic Manager supports multiple traffic-routing methods to deterministically route traffic to the various service endpoints.

*Nested configuration* - Use Traffic Manager in a nested configuration if you need more granular control to choose a preferred failover within a region.

For more information, see:

- [Configure the performance traffic routing method](#)
- [Traffic Manager routing methods](#)

## Global traffic view

Use Traffic View in Traffic Manager to see traffic patterns and latency metrics. Traffic View can help you plan your footprint expansion to new Azure regions.

See [Traffic Manager Traffic View](#) for details.

## Application Gateway

Use Application Gateway v2 SKU for out-of-the-box automated resiliency.

- Application Gateway v2 SKU automatically ensures that new instances spawn across fault domains and update domains. If you choose zone redundancy, the newest instances also spawn across availability zones to give fault tolerance.
- Application Gateway v1 SKU supports high-availability scenarios when you've deployed two or more instances. Azure distributes these instances across update and fault domains to ensure that instances don't fail at the same time. The v1 SKU supports scalability by adding multiple instances of the same gateway to share the load.

The Application Gateway needs to trust the CA certificate of Azure Firewall.

## Azure Firewall

The Premium tier of Azure Firewall is required in this design to provide TLS inspection. Azure Firewall will intercept the TLS sessions between Application Gateway and the web-tier virtual machines generating its own certificates, as well as inspect outbound traffic flows from the virtual networks to the public Internet. You can find more information on this design in [Zero-trust network for web applications with Azure Firewall and Application Gateway](#).

## Health probe recommendations

Here are some recommendations for health probes in Traffic Manager, Application Gateway, and Load Balancer.

### Traffic Manager

*Endpoint health* - Create an endpoint that reports the overall health of the application. Traffic Manager uses an HTTP(S) probe to monitor the availability of each region. The probe checks for an HTTP 200 response for a specified URL path. Use the endpoint you created for the health probe. Otherwise, the probe might report a healthy endpoint when critical parts of the application are failing.

For more information, see [health endpoint monitoring pattern](#).

*Failover delay* - Traffic Manager has a failover delay. The following factors determine the duration of the delay:

- Probing intervals: How often the probe checks the health of the endpoint.
- Tolerated number of failures: How many failures the probe tolerates before marking the endpoint unhealthy.
- Probe timeout: how long before Traffic Manager considers the endpoint unhealthy.
- Time-to-live (TTL): DNS servers must update the cached DNS records for the IP address. The time it takes depends on the DNS TTL. The default TTL is 300 seconds (5 minutes), but you can configure this value when you create the Traffic Manager profile.

For more information, see [Traffic Manager monitoring](#).

## Application Gateway and Load Balancer

Familiarize yourself with the health probe policies of the Application Gateway and load balancer to ensure you understand the health of your VMs. Here's a brief overview:

- Application Gateway always uses an HTTP probe.
- Load Balancer can evaluate either HTTP or TCP. Use an HTTP probe if a VM runs an HTTP server. Use TCP for everything else.
- HTTP probes send an HTTP GET request to a specified path and listen for an HTTP 200 response. This path can be the root path (""/"), or a health-monitoring endpoint that implements custom logic to check the health of the application.
- The endpoint must allow anonymous HTTP requests. If a probe can't reach an instance within the timeout period, the Application Gateway or Load Balancer stops sending traffic to that VM. The probe continues to check and will return the VM to the back-end pool if the VM becomes available again.

For more information, see:

- [Load Balancer health probes](#)
- [Application Gateway health monitoring overview](#)
- [Health endpoint monitoring pattern](#)

## Operational excellence

*Resource groups* - Use [resource groups](#) to manage Azure resources by lifetime, owner, and other characteristics.

*Virtual network peering* - Use [virtual network peering](#) to seamlessly connect two or more virtual networks in Azure. The virtual networks appear as one for connectivity purposes. The traffic between virtual machines in peered virtual networks uses the Microsoft backbone infrastructure. Make sure that the address space of the virtual networks doesn't overlap.

*Virtual network and subnets* - Create a separate subnet for each tier of your subnet. You should deploy VMs and resources, such as Application Gateway and Load Balancer, into a virtual network with subnets.

## Security

*Web Application Firewall* - The WAF functionality of Azure Application Gateway will detect and prevent attacks at the HTTP level, such as SQL injection (SQLi) or cross-site scripting (CSS).

*Next-Generation Firewall* - Azure Firewall Premium provides an additional layer of defense by inspecting content for non-web attacks, such as malicious files uploaded via HTTP(S) or any other protocol.

*End-to-end encryption* - Traffic is encrypted at all times when traversing the Azure network. Both Application Gateway and Azure Firewall encrypt traffic before sending it to the corresponding backend system.

*Distributed Denial of Service (DDoS)* - Use [Azure DDoS Network Protection](#) for greater DDoS protection than the basic protection that Azure provides.

*Network security groups (NSGs)* - Use [NSGs](#) to restrict network traffic within the virtual network. For example, in the three-tier architecture shown here, the data tier accepts traffic only from the business tier, not from the web front end. Only the business tier can communicate directly with the database tier. To enforce this rule, the database tier should block all incoming traffic except for the business-tier subnet.

1. Allow inbound traffic from the business-tier subnet.
2. Allow inbound traffic from the database-tier subnet itself. This rule allows communication between the database VMs. Database replication and failover need this rule.
3. Deny all inbound traffic from the virtual network, using the `VirtualNetwork` tag in the rule) to overwrite the permit statement included in the default NSG rules.

Create rule 3 with lower priority (higher number) than the first rules.

You can use [service tags](#) to define network access controls on Network Security Groups or Azure Firewall.

For more information, see [application gateway infrastructure configuration](#).

## Cost optimization

For more information, see:

- [Load Balancing pricing ↗](#)
- [Virtual network Pricing ↗](#)
- [Application gateway pricing ↗](#)
- [Choose the right Azure Firewall SKU to meet your needs](#)
- [Traffic Manager pricing ↗](#)
- [Pricing calculator ↗](#)

## Performance efficiency

*Virtual machine scale sets* - Use Virtual Machine Scale Sets to automate the scalability of your virtual machines. Virtual machine scale sets are available on all Windows and Linux virtual machine sizes. You're only charged for the virtual machines deployed and the underlying infrastructure resources consumed. There are no incremental charges. The benefits of Virtual Machine Scale Sets are:

- Create and manage multiple virtual machines easily
- High availability and application resiliency
- Automated scaling as resource demand changes

For more information, see [Virtual Machine Scale Sets](#).

## Next steps

For more reference architectures using the same technologies, see:

- [Multi-region N-tier application](#)
- [AKS baseline for multi-region clusters](#)

# Multi-tier web application built for HA/DR

Azure   Azure Arc   SQL Server   Windows

This example scenario is applicable to any industry that needs to deploy resilient multitier applications built for high availability and disaster recovery. In this scenario, the application consists of three layers.

- Web tier: The top layer including the user interface. This layer parses user interactions and passes the actions to next layer for processing.
- Business tier: Processes the user interactions and makes logical decisions about the next steps. This layer connects the web tier and the data tier.
- Data tier: Stores the application data. Either a database, object storage, or file storage is typically used.

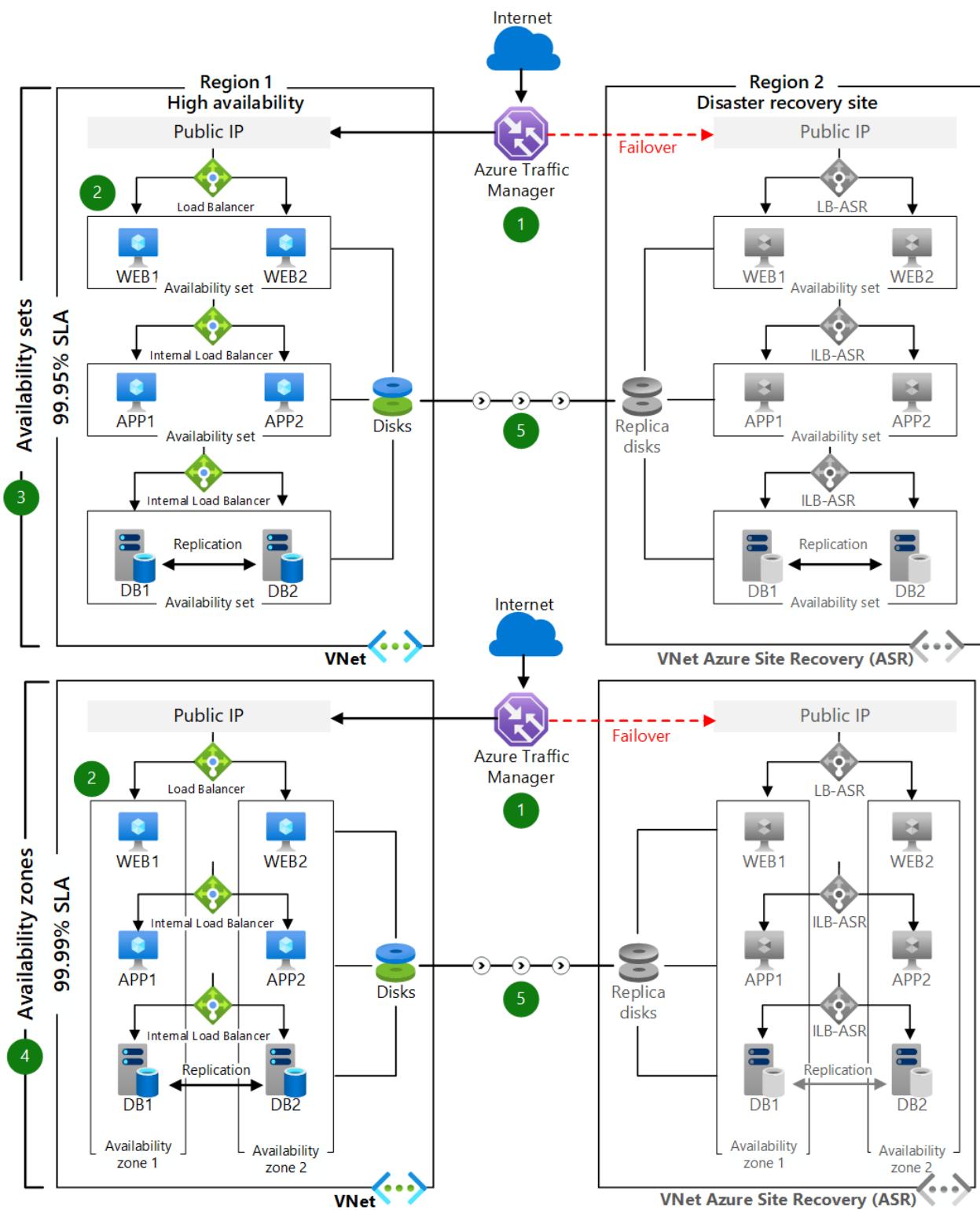
Common application scenarios include any mission-critical application running on Windows or Linux. This can be an off-the-shelf application such as SAP and SharePoint or a custom line-of-business application.

## Potential use cases

Other relevant use cases include:

- Deploying highly resilient applications such as SAP and SharePoint
- Designing a business continuity and disaster recovery plan for line-of-business applications
- Configure disaster recovery and perform related drills for compliance purposes

## Architecture



Download a [Visio file](#) of this architecture.

## Workflow

- Distribute the VMs in each tier across two availability zones in regions that support zones. In other regions, deploy the VMs in each tier within one availability set.

- The database tier can be configured to use Always On availability groups. With this SQL Server configuration, one primary read/write replica within an availability group is configured with up to eight secondary read-only replicas. If an issue occurs with the primary replica, the availability group fails over primary read/write activity to one of the secondary replicas, allowing the application to remain available. For more information, see [Overview of Always On availability groups for SQL Server](#).
- For disaster recovery scenarios, you can configure SQL Always On asynchronous native replication to the target region used for disaster recovery. You can also configure Azure Site Recovery replication to the target region if the data change rate is within supported limits of Azure Site Recovery.
- Users access the front-end ASP.NET web tier via the traffic manager endpoint.
- The traffic manager redirects traffic to the primary public IP endpoint in the primary source region.
- The public IP redirects the call to one of the web tier VM instances through a public load balancer. All web tier VM instances are in one subnet.
- From the web tier VM, each call is routed to one of the VM instances in the business tier through an internal load balancer for processing. All business tier VMs are in a separate subnet.
- The operation is processed in the business tier and the ASP.NET application connects to Microsoft SQL Server cluster in a back-end tier via an Azure internal load balancer. These back-end SQL Server instances are in a separate subnet.
- The traffic manager's secondary endpoint is configured as the public IP in the target region used for disaster recovery.
- In the event of a primary region disruption, you invoke Azure Site Recovery failover and the application becomes active in the target region.
- The traffic manager endpoint automatically redirects the client traffic to the public IP in the target region.

## Components

- **Availability sets** ensure that the VMs you deploy on Azure are distributed across multiple isolated hardware nodes in a cluster. If a hardware or software failure occurs within Azure, only a subset of your VMs are affected and your entire solution remains available and operational.
- **Availability zones** protect your applications and data from datacenter failures. Availability zones are separate physical locations within an Azure region. Each zone consists of one or more datacenters equipped with independent power, cooling, and networking.

- [Azure Site Recovery](#) allows you to replicate VMs to another Azure region for business continuity and disaster recovery needs. You can conduct periodic disaster recovery drills to ensure you meet the compliance needs. The VM will be replicated with the specified settings to the selected region so that you can recover your applications in the event of outages in the source region.
- [Azure Traffic Manager](#) is a DNS-based traffic load balancer that distributes traffic optimally to services across global Azure regions while providing high availability and responsiveness.
- [Azure Load Balancer](#) distributes inbound traffic according to defined rules and health probes. A load balancer provides low latency and high throughput, scaling up to millions of flows for all TCP and UDP applications. A public load balancer is used in this scenario to distribute incoming client traffic to the web tier. An internal load balancer is used in this scenario to distribute traffic from the business tier to the back-end SQL Server cluster.

## Alternatives

- Windows can be replaced by other operating systems because nothing in the infrastructure is dependent on the operating system.
- [SQL Server for Linux](#) can replace the back-end data store.
- The database can be replaced by any standard database application available.

## Scenario details

This scenario demonstrates a multitier application that uses ASP.NET and Microsoft SQL Server. In [Azure regions that support availability zones](#), you can deploy your virtual machines (VMs) in a source region across availability zones and replicate the VMs to the target region used for disaster recovery. In Azure regions that don't support availability zones, you can deploy your VMs within an [availability set](#) and replicate the VMs to the target region.

To route traffic between regions, you need a global load balancer. There are two main Azure offerings:

- Azure Front Door
- Azure Traffic Manager

When choosing a load balancer, consider your requirements and the feature set of the two offerings. How quickly do you want to fail over? Can you take on the overhead of TLS management? Are there any organizational cost constraints?

Front Door has Layer 7 capabilities: SSL offload, path-based routing, fast failover, caching, and others to improve performance and high-availability of your applications. You might experience faster packet travel times because the infrastructure is onboarded on Azure network sooner.

Because Front Door adds a new hop, there are added security operations. If the architecture complies with regulatory requirements, there might be restrictions about the additional traffic TLS termination point. The TLS cipher suites selected by Front Door must meet your organization's security bar. Also, Front Door expects the backend services to use [certificates used by Microsoft](#).

Another consideration is cost. The architecture should take advantage of the extensive feature set (not just failover) to justify the added cost.

Traffic Manager is a DNS-based load-balancing service. It balances and fails over only at the DNS level. For that reason, it can't fail over as quickly as Front Door, because of common challenges around DNS caching and systems not honoring DNS TTLs.

You can combine both load balancers, if needed. For example, you want the DNS-based failover but you want to add a POP experience in front of that traffic-managed infrastructure.

This architecture uses Traffic Manager because it's lightweight. The failover timing is sufficient for illustrative purposes.

## Considerations

### Scalability

You can add or remove VMs in each tier based on your scaling requirements. Because this scenario uses load balancers, you can add more VMs to a tier without affecting application uptime.

For other scalability topics, see the [performance efficiency checklist](#) in the Azure Architecture Center.

### Security

All the virtual network traffic into the front-end application tier is protected by network security groups. Rules limit the flow of traffic so that only the front-end application tier VM instances can access the back-end database tier. No outbound internet traffic is allowed from the business tier or database tier. To reduce the attack footprint, no direct

remote management ports are open. For more information, see [Azure network security groups](#).

For general guidance on designing secure scenarios, see the [Azure Security Documentation](#).

## Pricing

Configuring disaster recovery for Azure VMs using Azure Site Recovery will incur the following charges on an ongoing basis.

- Azure Site Recovery licensing cost per VM.
- Network egress costs to replicate data changes from the source VM disks to another Azure region. Azure Site Recovery uses built-in compression to reduce the data transfer requirements by approximately 50%.
- Storage costs on the recovery site. This is typically the same as the source region storage plus any additional storage needed to maintain the recovery points as snapshots for recovery.

We've provided a [sample cost calculator](#) for configuring disaster recovery for a three-tier application using six virtual machines. All of the services are pre-configured in the cost calculator. To see how the pricing would change for your particular use case, change the appropriate variables to estimate the cost.

## Contributors

*This article is maintained by Microsoft. It was originally written by the following contributors.*

Principal author:

- [Sujay Talasila](#) | Principal Product Lead

## Next steps

- [Deploy Traffic Manager in Azure](#)
- [Set up disaster recovery for Azure VMs](#)

## Related resources

For additional high availability and disaster recovery reference architectures, see:

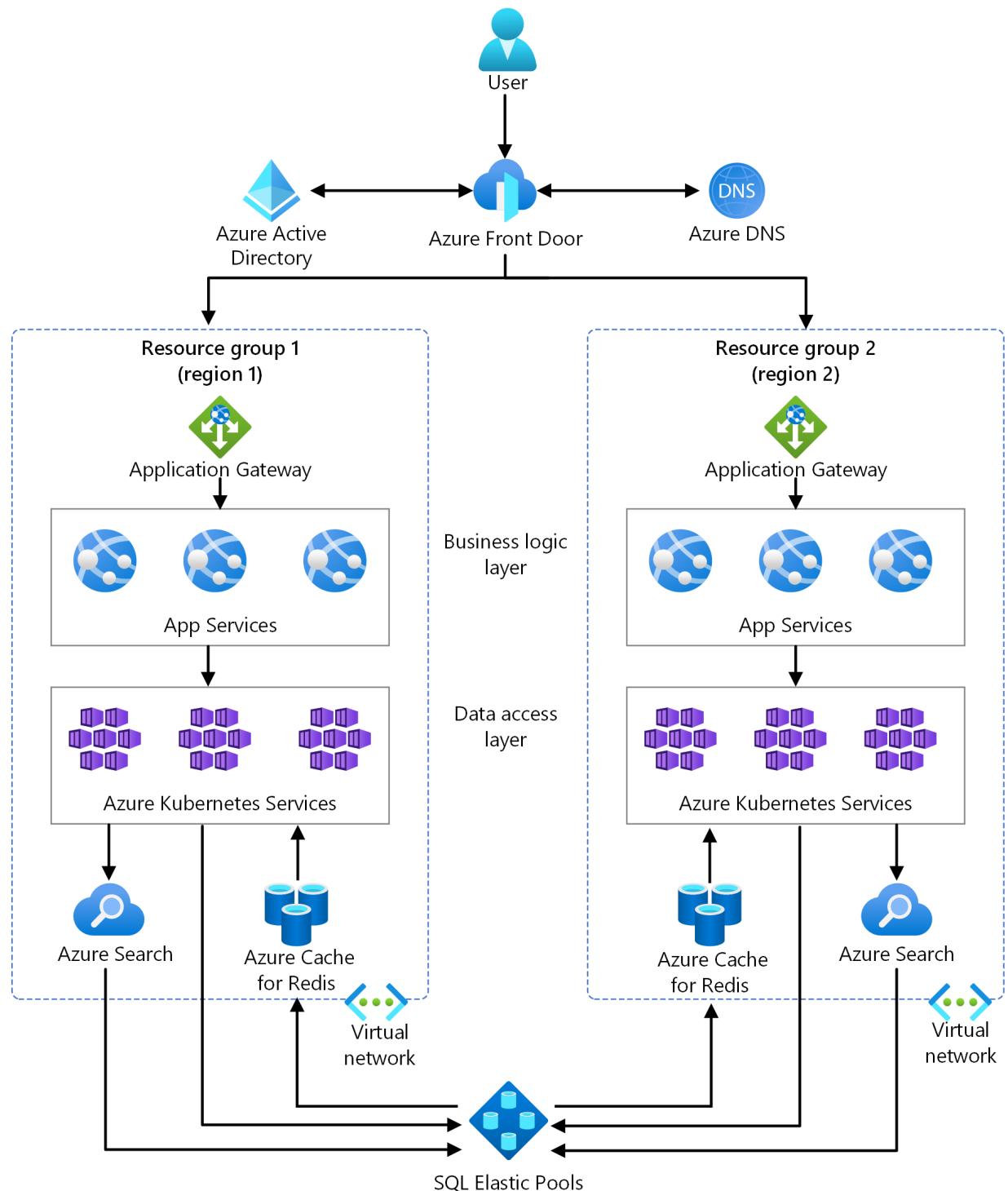
- Multi-region N-tier application
- Multi-region load balancing
- [Multi-region app with private database][Multi-region-app-with-private-database]
- Enterprise-scale disaster recovery

# Multitenant SaaS on Azure

Microsoft Entra ID   Azure App Service   Azure DNS   Azure Front Door   Azure Kubernetes Service (AKS)

When you identify a portion of your business's software solution that you can unbrand and market to other businesses, it adds an entire new revenue stream for a company. However, configuring the solution to account for the load that a slew of tenants brings is often a challenging obstacle to tackle. This solution tours a suite of Azure technologies that secure and balance the traffic.

## Architecture



Download a [Visio file](#) of this architecture.

## Workflow

A suite of Azure technologies secure and load balance the traffic.

1. Microsoft Azure Front Door handles a few initial tasks:

- Processing the initial request

- Load balancing across the regions
- SSL(HTTPS) termination and offloading
- Failing over if there's a regional outage

2. Azure DNS manages DNS records and ensures routing to the correct Azure Front Door endpoint.
3. The architecture uses Microsoft Entra ID as the identity provider for authentication.
4. Once routed to the appropriate region, Application Gateway routes and load balances, directing requests to the appropriate Azure App Service.
5. For this architecture, using App Service is the preferred service for:

- Any HTTP-based application.
- Serving web content.
- Exposing RESTful APIs.
- Implementing business logic behind the front-end application.

You can configure App Service to scale up and out automatically. It makes App Service a good fit for scaling a host of tenant HTTP-driven requests on demand.

6. The data-access layer services are also independently scaled based on load. Data services manage data models, connection clients, and drivers. The services also provide a consistent data interface for all higher-level services wishing to consume data in the application. You can deploy and scale these data services using the Azure Kubernetes Service (AKS). Each AKS cluster is responsible for a set of related features in the layer. AKS can implement a microservice architecture, which features a series of containers that each encapsulate specific functionality within the cluster. This allows for a high degree of abstraction and de-coupling within the code. It also allows for clusters to scale out individually to account for increased load from multiple tenants. Each cluster can scale up its resources if load increases on the cluster. The scale up doesn't affect the other clusters in the resource group as long as they aren't experiencing the same increase.
7. Store and manage relational data outside of the application framework. Doing so provides a single point of data entry for either region. You can achieve replication, availability, scalability, and security by leveraging the strength of Azure SQL Elastic Pools. Provision each tenant a database in a pool. Allocate the resources available

in the pool to databases on-demand as load and requests come in. This optimizes database resources available for tenants against your budget.

## Components

The primary components are the suggested components for the architecture in this solution. If any of the primary components don't fit your architecture, see the list of alternative components.

### Primary components

- [Azure Front Door](#): A regional load balancer that routes client traffic to the correct region. It can fail over to the second region if a regional failure happens, and it can secure the internet-facing entry point via [Azure Web Application Firewall](#).
- [Microsoft Entra ID](#): Acts as the identity provider for the entire application, enforcing authentication and end-to-end authorization of the request in the application.
- [Azure DNS](#): A hosting service in Azure for domain name resolution. In a multitenant solution, multiple clients access the solution via their own individual domains. Use Azure DNS to configure and resolve client requests to their correct application stack.
- [Application Gateway](#): Routes and load-balances traffic internally in the application to the various services that satisfy client business needs. While Azure Front Door balances load across high-level regions, it's Application Gateway that has awareness of the load on individual services within a group. Azure Front Door and Application Gateway combine to provide complex load-balancing at all levels in a multitenant solution. For more information on load-balancing options in Azure, visit this [overview on Azure load-balancing](#).
- [App Service](#): Azure's premier service for web applications and web-based APIs. Security integrates with services like Microsoft Entra ID and [Azure Key Vault](#). You can configure automatic scaling. Also, the amount of resources available to scale to is flexible between the various App Service plans on which the app can run. App Service can also leverage integrated DevOps capabilities for continuous integration and deployment to multiple environments. These and other supporting features of the Azure platform allow for developers to focus on the development of their applications.

- [Azure Kubernetes Service \(AKS\)](#): Orchestrates instances of container images deployed to a cluster. Managing multiple clients' data often involves implementing a suite of components to manage:
  - Data modeling
  - Data source connectivity
  - Extract, transform, load (ETL)
  - Import/export activities

Developing these many smaller components as container-based microservices creates an ideal scenario for the deployment to an AKS cluster. Tools for autoscaling, load balancing, and upgradeability are built into the framework. AKS integrates well with a continuous integration/continuous delivery (CI/CD) strategy using the available DevOps features and Azure Container Registry.

- [Azure SQL Elastic Pools](#): Provides a solution for managing a set of databases flexibly with a pool of resources. The service allocates resources on demand to the databases. It gives the developer of a multitenant SaaS architecture the power to deliver database resources to clients as they need it. The service also reduces the budget and overhead of maintaining multiple SQL Servers with large chunks of unused compute resources.
- [Azure Cognitive Search](#) (formerly known as Azure Search): A service that adds a powerful indexing and query engine to your application. It gives clients access to strong query functionality. They can also use Azure's AI capabilities to enrich and enhance the query functionality. Azure Cognitive Search can account for multitenancy using an index-per-tenant or service-per-tenant strategy.
- [Azure Cache for Redis](#): Applies a caching layer as a service to the solution, providing an in-memory managed cache to reduce latency and increase performance for the clients. High throughput allows for a high volume of requests to handle multiple tenants accessing the system. You can flexibly scale up the service as application loads increase. It also supports encryption at rest to protect and isolate cached tenant data.

## Alternatives components

- [Azure Virtual Machine Scale Sets](#): Allows for the deployment of services to a virtual machine (VM) environment that scales and grows automatically as needed. Virtual Machine Scale Sets integrate well with a [Load Balancer](#) or Application

Gateway to automatically rebalance load as the scale set grows. Virtual Machine Scale Sets provides the scalability this solution demands. In many cases though, it's unnecessary to manage the full VM environment, and we can defer that level of the stack to App Service or AKS.

- [Azure SQL Database](#): Implement as individual dedicated instances as a replacement for Elastic Pools. Using Azure SQL Database adds higher overhead in managing the instance directly and incurs more cost for allocated resources. That said, it's an acceptable alternative when the tenant requires a dedicated server. In particular, the client might require more control over the instance and dedicated available resources. Tenants that require a dedicated SQL Server can exist side by side with tenants on an Elastic Pool configuration. You can make a tier of SQL databases one of the pricing options available to tenants when purchasing licenses for the SaaS.
- [SQL Server on Virtual Machines](#): Another option for the deployment of SQL databases. The tenant might have pre-existing IT infrastructure and existing SQL Servers on premises. In that case, the tenant might want to use their current licenses, either as a full migration or in a hybrid scenario. The decoupled nature of the SaaS allows for the data layer of the application to target any SQL Database via configuration.

## Scenario details

When you identify a portion of your business's software solution that you can unbrand and market to other businesses, it adds an entire new revenue stream for a company. However, configuring the solution to account for the load that a slew of tenants brings is often a challenging obstacle to tackle.

Azure offers a range of services for managing a software solution that:

- Flexibly maintains databases for all clients.
- Scales the business and logic tier of the solution to prevent bottlenecks at the compute layer.
- Integrates availability and regional failover.
- Provides end-to-end security at all levels of the solution.

## Potential use cases

These use cases have design patterns that can benefit from a multitenant SaaS solution hosted on Azure:

- Develop a customer relationship management (CRM) solution that clients can market and sell to customers.
- Implement a content management system (CMS) system and deliver it to multiple users using this architecture.

## Considerations

These considerations implement the pillars of the Azure Well-Architected Framework, which is a set of guiding tenets that can be used to improve the quality of a workload. For more information, see [Microsoft Azure Well-Architected Framework](#).

## Multitenancy

A multitenant solution is the key consideration in this solution. The solution handles a number of clients simultaneously. It also allocates enough resources to process all client requests effectively. While processing requests, the solution secures traffic from global endpoints and isolates client data to prevent breaches and cross-contamination. Deploy clients to a pair of regional resource groups based on their primary location. Doing so optimizes regional availability.

You can deploy many clients to a single compute group because the system isolates requests based on authentication and client keys, which differentiates requests based on these unique identifiers. The system can encrypt all client requests separately by their keys so that no client can decrypt any other client's data. Managing multiple clients on a single compute stack gives you the ability to optimize resource allocation to provide clients the responsiveness they need at cost.

You manage client databases in a similar way outside of the compute stack, because a client request could arrive from either of the regional stacks. Many client databases can exist on the same Elastic Pool, isolated and secured by transparent data encryption (TDE). You can configure each database to encrypt data using a client-managed key and decrypt the data just in time (JIT). Decrypting JIT protects client data from both the developer and other clients. The system leverages the Elastic Pool to provide resources on demand to the clients assigned to it while keeping costs low for you. You can assign replication policies to each Elastic Pool to provide backup and failover for client data. Bring more Elastic Pools online as you onboard more clients into the system.

For more information about multitenant solutions, see [Architect multitenant solutions on Azure](#).

## Reliability

Reliability ensures your application can meet the commitments you make to your customers. For more information, see [Overview of the reliability pillar](#).

## Scalability and Availability

This solution is designed to account for a large number of tenants using the SaaS. It takes advantage of the large number of scalable components and services to grow based on load. This architecture isn't designed for solutions that service a few tenants, or a small load of requests and data. It could stress the budget of a solution targeting a single client or smaller load. It's also unnecessary to have the multiregion overhead where high global availability isn't a requirement, because it adds unnecessary complexity and cost.

## Security

Security provides assurances against deliberate attacks and the abuse of your valuable data and systems. For more information, see [Overview of the security pillar](#).

The system addresses security from end-to-end at each level of the application:

- Azure Front Door provides built-in HTTPS support for its domains. This means the system can encrypt all traffic to the SaaS application. Azure Front Door also implements Azure Web Application Firewall, protecting the SaaS stack from attacks at the edge, before the system routes requests to the application.
- Each application stack in each region lies within an Azure Virtual Network. The system restricts traffic into the virtual network accepting requests from Azure Front Door, protecting all application services from external traffic. Once inside the secure firewall, Application Gateway can terminate SSL and provide performant load balancing and routing within the application.
- You can securely manage all credentials, secrets, and connection strings by using Azure Key Vault. By managing this sensitive data as secrets, developers can inject credentials into the application at the time of deployment. Doing so makes sure that the code isn't polluted with sensitive information. Using secrets protects client data by ensuring that a breach in code or man-in-the-middle attack wouldn't gain access to tenant databases.

- In this scenario, the data of multiple tenants might exist side by side on the same database server, if not the same database. Using TDE and JIT decryption protects data on the database. The system encrypts all data on the database at rest, and only decrypts it when requested by the tenant. Clients can provide their own keys, and you can store all client keys in Azure Key Vault to manage encryption for multiple tenants. It protects client data end to end, prevents the developer from having access to client data, isolates data between tenants, and helps to meet compliance requirements for security and data.

## Cost optimization

Cost optimization is about looking at ways to reduce unnecessary expenses and improve operational efficiencies. For more information, see [Overview of the cost optimization pillar](#).

Azure App Service provides many pricing tiers based on the expected compute resources required. For a multitenant SaaS, high availability and scale-out capabilities are key components in choosing the service plan. If you expect to host many tenants, choosing a premium or isolated tier might be necessary to provide the compute resources necessary to account for the high traffic. The standard, premium, and isolated tiers are all dedicated VM instances. You can calculate cost per unit of time by how many VMs of said tier you've specified. For more information, visit the [overview of App Service pricing plans](#).

Azure Kubernetes Service provides a cost-effective container service. Charges for AKS nodes only occur on usage, so you're only charged for:

- The VMs
- Consumed storage and network resources
- Scaling cost directly related to usage

Using AKS as the data-tier service is ideal if you're looking to reduce costs. For an estimate on pricing out a layer of AKS instances, visit the [Kubernetes service calculator](#).

By design, the Azure SQL Elastic Pool pricing is highly cost-effective in a multitenant scenario. Tenant databases in an Elastic Pool will share the available resources. As demand shifts between tenants over time, resources will shift as well. Azure SQL Elastic Pool provides the maximum available resources to demanded databases without the need for resource overhead on all databases. The service keeps cost low for the developer of the SaaS and the tenants. Use the [Azure SQL Database pricing calculator](#)

to price out and determine the tier and amount of resources needed to serve your tenants and their data.

- Using a virtual core (vCore) pricing model provides greater flexibility in scaling to meet required resources. Also, you can take advantage of the Azure Hybrid Benefit. Existing SQL Server licenses provide a discount to vCore SQL resources in the cloud. Therefore, in an instance when on-premises servers are already part of the developer infrastructure, you can manage cost even more by using these discounts. You can estimate your potential savings by using the [Azure Hybrid Benefit savings calculator](#).
- You can also save cost on SQL Server resources by purchasing [Azure SQL Database reserved capacity](#). Purchasing reserved capacity marks a commitment of long-term SQL Database usage. The term is usually between one to three years. In return, you get discounts on the compute costs of the resources in reservation. For instance, you could reserve 32 general purpose vCores for a year, which reduces the cost of those 32 vCores for that year. Having multiple tenants purchasing licenses for a SaaS is a strong indicator that making use of Reserved capacity fits the solution, and an ideal cost saver in this workload.

You can find the pricing structure for Azure Cache for Redis on the [Azure Cache for Redis pricing](#) page. Adjust the cache tier at any time between a Basic, Standard, and Premium tier based on need. You'll see higher pricing on the larger cache limits and additional features such as replication and disaster recovery. Azure Cache for Redis also offers reserved capacity pricing for long-term usage commitments.

Azure Front Door's pricing depends on the amount of data transfer in and out of the service. For outbound data, the pricing is different based on zones. Different regions will incur different costs. If you come across a price differential, estimate the cost separately. The price includes some routing and domain capacity, but the system incurs costs past the initial limits. Azure Web Application Firewall incurs a small additional charge per policy or rule applied. You can find the pricing details for Azure Front Door on the [Azure Front Door pricing](#) page.

The [pricing for Azure Cognitive Search](#) is a fully tiered system. A free tier is available for development and testing. After that, each tier incurs a per-hour cost for each Cognitive Search instance allocated. As the tiers increase, the total storage, number of indexes, and scale-out limits also increase. Azure Cognitive Search provides image extraction as a service at the same rate to all paid tiers.

## Next steps

- Application and service principal objects in Microsoft Entra ID is about leveraging Microsoft Entra ID to implement multitenant apps.
- Multitenant SaaS database tenancy patterns covers implementing multitenancy patterns in SQL.
- Tenancy in Microsoft Entra ID is also about leveraging Microsoft Entra ID to implement multitenant apps.

## Related resources

- Run a web application in multiple Azure regions for high availability is a reference for the multiregion requirement of the solution.
- Multitier web application built for high availability and disaster recovery on Azure is a similar example workload scenario. It describes many of the considerations for a large-scale application on Azure.

# Network-hardened web application with private connectivity to PaaS datastores

Azure App Service

Azure Front Door

Azure Private Link

Azure SQL Database

Azure Firewall

This article describes how to set up an [Azure App Service web app](#) in a network environment that enforces strict policies for inbound and outbound network flows. In such cases, the web app can't be directly exposed to the internet. Instead, all traffic needs to go through an [Azure Firewall](#) or a third-party network virtual appliance (NVA).

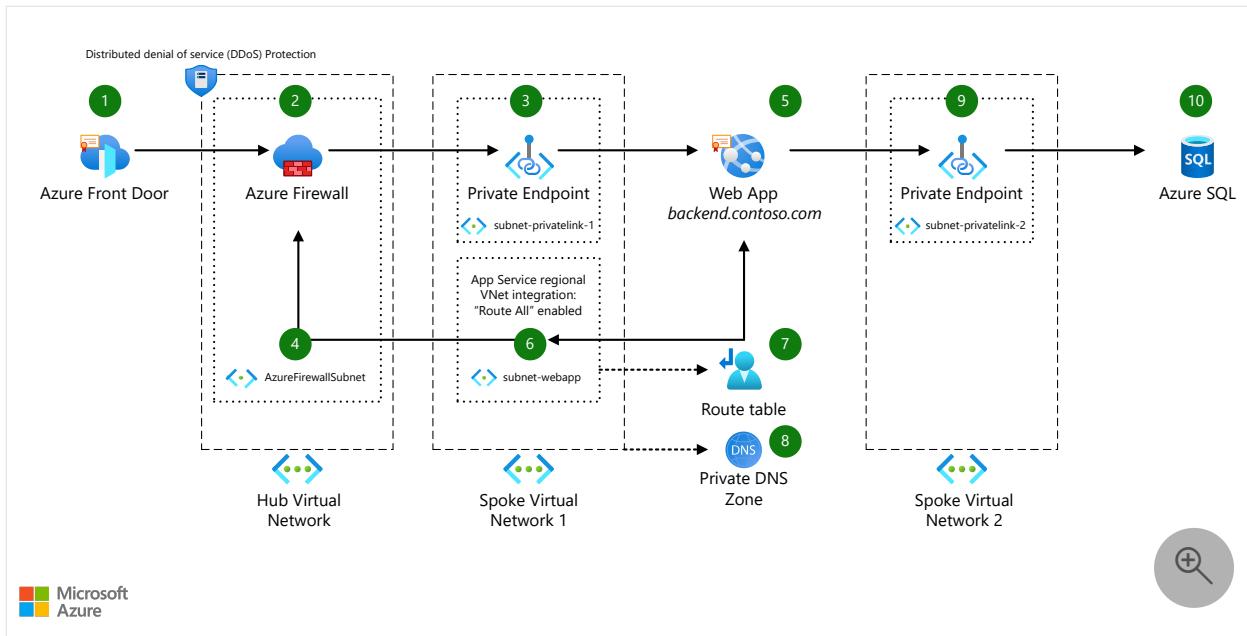
The example shows a scenario in which a web app is protected with [Azure Front Door](#) and an Azure Firewall. It connects with improved security to an [Azure SQL Database](#).

## Potential use cases

These use cases have similar design patterns:

- Connect from a web app or an [Azure Function](#) to any platform as a service (PaaS) offering that supports [Azure Private Link-based private endpoints](#) for inbound network flows. Examples include Azure Storage, Azure Cosmos DB, and other web apps.
- Connect from a web app or an Azure Function to a [virtual machine \(VM\)](#) in the cloud or on premises by using virtual private networks (VPNs) or [Azure ExpressRoute](#).

## Architecture



Download a [Visio file](#) of this architecture.

## Dataflow

Cross-reference the following 10 steps with the annotated architecture diagram shown previously while we describe the solution's dataflow:

1. An Azure Front Door instance provides [Azure Web Application Firewall \(WAF\)](#) features and terminates TLS/SSL connections from clients. [End-to-end TLS](#) ensures the network traffic is re-encrypted before it's forwarded to Azure Firewall.
2. A custom fully qualified domain name (FQDN) is chosen to represent the back-end web app and is mapped through CNAME or A DNS resource records to the public IP address of an Azure firewall or third-party NVA.
3. A private endpoint for the web app is created in a virtual network subnet (`subnet-privatelink-1` in the example).
4. The Azure Firewall or third-party NVA is deployed to its own reserved subnet in a virtual network (*Hub Virtual Network* in the example). The appliance is configured to perform destination network address translation (DNAT) of incoming requests to the private IP address or the private endpoint associated with the web app. Azure Firewall can also handle source NAT (SNAT) for Internet-outbound egress traffic from your Azure VMs.
5. The web app is assigned the custom FQDN through the [domain verification ID property of the web app](#). This allows the custom FQDN already mapped to the public IP address of the Azure Firewall or third-party NVA to be reused with the

web app without altering Domain Name System (DNS) name resolution and network flows.

6. The web app connects to a virtual network subnet (*subnet-webapp* in the example) through regional VNet integration. The **ROUTE ALL** setting is enabled, which forces all outbound traffic from the web app into the virtual network and allows the web app to inherit the virtual network's DNS resolution configuration, including custom DNS servers and integration with [Azure Private DNS zones](#) used for private endpoint name resolution. Additional details on how to configure [Azure App Service virtual network integration routing](#) can be found here.
7. A custom [route table](#) that's attached to the web app subnet (*subnet-webapp* in the example) forces all outbound traffic that comes from the web app to go to the Azure Firewall or third-party NVA.

 **Note**

**Azure Route Server** is an alternative to manually maintained custom route tables that uses Border Gateway Protocol (BGP) to automate route propagation to your Azure virtual network subnets.

8. One or more private DNS zones link to the virtual network that contains the web app (*Spoke Virtual Network 1* in the example) to allow DNS resolution of PaaS resources deployed with private endpoints.
9. A private endpoint for an Azure SQL Database virtual server is created in a virtual network subnet (*subnet-privatelink-2* in the example). A corresponding DNS record is created on the matching Azure Private DNS zone.
10. The web app can now be accessed only through Azure Front Door and Azure Firewall. It can also establish a connection to the Azure SQL Database virtual server through the private endpoint, securing the communication over private IP addresses only.

## Components

- [Azure Virtual Networks](#) form the underlying network traffic segmentation vehicle for this solution. Consider implementing [Azure Virtual Network Manager \(AVNM\)](#) to apply consistent administration and network security group policies to multiple VNets simultaneously.
- [Azure Front Door](#) provides Azure WAF features and terminates TLS/SSL connections from clients.

- [Azure Firewall](#) provides security to the web app for both ingress and egress.
- [Azure App Service](#) allows you to create web apps and deploy them in a cloud infrastructure.
- [Azure Private Link private endpoints](#) allow you to connect privately and with improved security to Azure services.
- [Azure SQL](#) connects to the web app via a private endpoint.

## Alternatives

- You can deploy the web app to an internal, single-tenant [App Service Environment](#) to provide isolation from the public Internet. This example uses an Azure App Service web app to reduce operating costs.
- You can replace Azure Front Door with an [Azure Application Gateway](#) if you also need to deploy the WAF component of the solution behind a firewall or within a virtual network.

## Considerations

These considerations implement the pillars of the Azure Well-Architected Framework, a set of guiding tenets that can be used to improve the quality of a workload. For more information, see [Microsoft Azure Well-Architected Framework](#).

## Reliability

Reliability ensures your application can meet the commitments you make to your customers. For more information, see [Overview of the reliability pillar](#).

Azure Front Door is a global service with built-in availability and redundancy and a high service level agreement ([SLA](#)).

Azure Firewall features [built-in availability](#) and a [high SLA](#). You can deploy it to span multiple [availability zones](#) to [increase the SLA](#). If you use a third-party or custom NVA, you can achieve the same SLA targets by configuring your deployment to use availability sets or availability zones.

Azure web apps support built-in availability. You can deploy them across [multiple availability zones](#).

You can further increase the availability of the solution by spreading it across multiple [Azure regions](#). You can accomplish this by deploying new instances of all components (except Azure Front Door) to other Azure regions and then configuring the original

Azure Front Door instance with [multiple back-end targets](#). If you use Azure SQL as your data store, you can then [join multiple servers to an auto-failover group to enable transparent and coordinated failover of multiple databases](#).

See these reference architectures to learn about deploying highly available web applications in Azure and setting up multi-region SQL Server instances to work with private endpoints:

- [Highly available multi-region web application](#)

## Security

Security provides assurances against deliberate attacks and the abuse of your valuable data and systems. For more information, see [Overview of the security pillar](#).

Specifically, this solution deploys an Azure Front Door instance, which terminates TLS/SSL connections from clients and provides a rich set of WAF configurations. We recommend that you further lock down your applications to accept traffic coming only from your Azure Front Door instance. You can do this in several ways, depending on the NVA you're using and your application configuration.

Some security options to consider integrating into your solution include:

- Placing Azure Front Door Premium on a private endpoint. An [Azure Front Door Premium private endpoint](#) secures your origin to a virtual network, ensuring consumers interact with your solution using non-Internet routable private IP addresses only.
- Associating [network security groups \(NSGs\)](#) to each VNet subnet. NSGs protect ingress and egress network traffic at Open Systems Interconnection (OSI) Layer 4 (Transport Layer). [Private Endpoint support for NSGs \(preview\)](#) ↗ enables you to implement advanced security controls on VNet egress traffic to a private endpoint.
- Configuring your Azure Firewall or NVA to accept traffic only from the [AzureFrontDoor.Backend Azure IP ranges](#) ↗.
- Configuring your NVA to integrate with [Azure service tags](#).
- Configuring your application to accept traffic only from your Azure Front Door instance by validating request headers.
- Defending against threats with Security Information and Event Management (SIEM) plus eXtended Detection and Response (XDR). Solutions within this category include [Microsoft Sentinel](#), [Microsoft Defender for Identity](#), and [Microsoft Defender for Cloud](#).
- Enabling enterprise-scale Distributed Denial of Service (DDoS) protection with [Azure DDoS Network Protection](#). This solution protects your Front Door and Azure

Firewall public IP addresses against abuse and provides customers with deep insights into Microsoft's automatic mitigation processes.

- Considering [Microsoft Purview](#) to establish unified data governance not only for your Azure SQL data, but potentially all data in your hybrid cloud, multicloud enterprise

For more information, see [How do I lock down the access to my backend to only Azure Front Door?](#).

The solution also uses an Azure SQL Database virtual server that accepts traffic only through a private endpoint, locking down traffic that comes from external sources. The web app used in the solution is configured to ensure proper DNS resolution of private endpoints and allow secure communication with the SQL Server instance.

When you deploy resources that use private endpoints in your environments, it's important to configure your DNS infrastructure properly. For more information, see [Azure private endpoint DNS configuration](#).

## Cost optimization

Cost optimization is about looking at ways to reduce unnecessary expenses and improve operational efficiencies. For more information, see [Overview of the cost optimization pillar](#).

The example scenario features a deployment within a hardened network environment. So an Azure Firewall or third-party NVA most likely already exists in the target infrastructure.

The main consideration for the remaining infrastructure is the stock keeping unit (SKU) of the App Service plan that hosts the web app. Private endpoints for web apps are [available only in the Premium App Service plan SKUs](#).

Use the [Azure pricing calculator](#) to estimate your costs. Here are two possible pricing estimates:

- [Infrastructure, including Azure Firewall](#)
- [Infrastructure, excluding Azure Firewall](#)

## Operational excellence

Operational excellence covers the operations processes that deploy an application and keep it running in production. For more information, see [Overview of the operational excellence pillar](#).

You can use [Azure Monitor](#) to monitor all components of this solution. You can use [Log Analytics](#) to monitor logs related to the WAF and network inspection rules of Azure Front Door and Azure Firewall. You can use [Application Insights](#) to monitor performance and availability and gain insights into your use of web applications.

## Performance efficiency

Performance efficiency is the ability of your workload to scale to meet the demands placed on it by users in an efficient manner. For more information, see [Performance efficiency pillar overview](#).

All components of the solution either provide transparent built-in scalability or expose a rich set of features, like [Azure web app autoscale](#), for scaling the number of available instances.

## Deploy this scenario

### Prerequisites

- An Azure account. If you don't have an Azure subscription, [create a free account](#) before you start.
- A publicly routable domain. Additionally, you must have permissions to create two DNS records in your public DNS zone.
- A valid TLS/SSL certificate to use for your web app. For instructions on how to do this, see [Add a TLS/SSL certificate in Azure App Service](#).

### Walkthrough

The solution is made up of several [Bicep](#) files that deploy the required infrastructure. You can deploy the solution by following the instructions in the [Hardened Web App GitHub repository](#).

## Contributors

*This article is maintained by Microsoft. It was originally written by the following contributors:*

Principal authors:

- [Davide Maccarrone](#) | Senior Consultant

- [Tim Warner](#) | Senior Content Developer
- [Mikey Lombardi](#) | Senior Content Developer

## Next steps

- [Azure Web Apps deployment best practices](#)
- [Azure private endpoint DNS configuration](#)
- [Azure Web Application Firewall on Azure Front Door](#)

## Related resources

- [Azure Architecture Center](#)
- [Azure Well-Architected Framework](#)
- [Azure Cloud Adoption Framework](#)
- [Azure Firewall architecture overview](#)

# Network topology and connectivity for Azure VMware Solution

Article • 04/24/2023

When using a VMware software-defined datacenter (SDDC) with an Azure cloud ecosystem, you have a unique set of design considerations to follow for both cloud-native and hybrid scenarios. This article provides key considerations and best practices for networking and connectivity to, from, and within Azure and [Azure VMware Solution](#) deployments.

The article builds on several Cloud Adoption Framework [enterprise-scale landing zones](#) architectural principles and recommendations for managing network topology and connectivity at scale. You can use this Azure landing zone design area guidance for mission-critical Azure VMware Solution platforms. Design areas include:

- **Hybrid integration** for connectivity between on-premises, multicloud, edge, and global users. For more information, see [Enterprise-scale support for hybrid and multicloud](#).
- **Performance and reliability at scale** for workload scalability and consistent, low-latency experience. A subsequent article covers [Dual region deployments](#).
- **Zero-trust-based network security** for network perimeter and traffic flow security. For more information, see [Network security strategies on Azure](#).
- **Extensibility** for easy expansion of network footprints without any need for design reworks.

## General design considerations and recommendations

The following sections provide general design considerations and recommendations for Azure VMware Solution network topology and connectivity.

### Hub-spoke vs. Virtual WAN network topology

If you don't have an ExpressRoute connection from on-premises to Azure and you're instead using S2S VPN, you can use Virtual WAN to transit [connectivity between your on-premises VPN and the Azure VMware Solution ExpressRoute](#). If you're using a hub-spoke topology, you need Azure Route Server. For more information, see [Azure Route Server support for ExpressRoute and Azure VPN](#).

## Private clouds and clusters

- All clusters can communicate within an Azure VMware Solution private cloud because they all share the same /22 address space.
- All clusters share the same connectivity settings, including internet, ExpressRoute, HCX, public IP, and ExpressRoute Global Reach. Application workloads can also share some basic networking settings like network segments, dynamic host configuration protocol (DHCP), and Domain Name System (DNS) settings.
- Design private clouds and clusters in advance before your deployment. The number of private clouds you require directly affects your networking requirements. Each private cloud requires its own [/22 address space for private cloud management](#) and [IP address segment for VM workloads](#). Consider defining those address spaces in advance.
- Discuss with your VMware and networking teams how to segment and distribute your private clouds, clusters, and network segments for workloads. Plan well and avoid wasting IP addresses.

For more information about managing IP addresses for private clouds, see [Define the IP address segment for private cloud management](#).

For more information about managing IP addresses for VM workloads, see [Define the IP address segment for VM workloads](#).

## DNS and DHCP

For DHCP, use the DHCP service built into NSX-T Data Center, or use a local DHCP server in a private cloud. Don't route broadcast DHCP traffic over the WAN back to on-premises networks.

For DNS, depending on the scenario you adopt and your requirements, you have multiple options:

- For an Azure VMware Solution environment only, you can deploy a new DNS infrastructure in your Azure VMware Solution private cloud.
- For Azure VMware Solution connected to an on-premises environment, you can use existing DNS infrastructure. If necessary, deploy DNS forwarders to extend into Azure Virtual Network or, preferably, into Azure VMware Solution. For more information, see [Add a DNS forwarder service](#).
- For Azure VMware Solution connected to both on-premises and Azure environments and services, you can use existing DNS servers or DNS forwarders in

your hub virtual network if available. You can also extend existing on-premises DNS infrastructure to the Azure hub virtual network. For details, see the [enterprise-scale landing zones diagram](#).

For more information, see the following articles:

- [DHCP and DNS resolution considerations](#)
- [Configure DHCP for Azure VMware Solution](#)
- [Configure DHCP on L2 stretched VMware HCX networks](#)
- [Configure a DNS forwarder in the Azure portal](#)

## Internet

Outbound options for enabling internet and filtering and inspecting traffic include:

- Azure Virtual Network, NVA, and Azure Route Server using Azure internet access.
- On-premises default route using on-premises internet access.
- Virtual WAN secured hub with Azure Firewall or NVA, using Azure internet access.

Inbound options for delivering content and applications include:

- Azure Application Gateway with L7, Secure Sockets Layer (SSL) termination, and Web Application Firewall.
- DNAT and load balancer from on-premises.
- Azure Virtual Network, NVA, and Azure Route Server in various scenarios.
- Virtual WAN secured hub with Azure Firewall, with L4 and DNAT.
- Virtual WAN secured hub with NVA in various scenarios.

## ExpressRoute

The Azure VMware Solution out-of-the-box private cloud deployment automatically one free 10 Gbps ExpressRoute circuit. This circuit connects Azure VMware Solution to the D-MSEE.

Consider deploying Azure VMware Solution in [Azure paired regions](#) near your datacenters.

## Global Reach

- Global Reach is a required ExpressRoute add-on for Azure VMware Solution to communicate with on-premises datacenters, Azure Virtual Network, and Virtual WAN. The alternative is to design your network connectivity with Azure Route Server.

- You can peer the Azure VMware Solution ExpressRoute circuit with other ExpressRoute circuits using Global Reach at no charge.
- You can use Global Reach for peering ExpressRoute circuits through an ISP and for ExpressRoute Direct circuits.
- Global Reach isn't supported for ExpressRoute Local circuits. For ExpressRoute Local, transit from Azure VMware Solution to on-premises datacenters via third-party NVAs in an Azure virtual network.
- Global Reach isn't available in all locations.

## Bandwidth

Choose an appropriate [virtual network gateway SKU](#) for optimal bandwidth between Azure VMware Solution and Azure Virtual Network. Azure VMware Solution supports a [maximum of four ExpressRoute circuits](#) to an ExpressRoute gateway in one region.

## Network security

Network security involves traffic inspection and port mirroring.

*East-West traffic inspection* within an SDDC uses NSX-T Data Center or NVAs to inspect traffic to Azure Virtual Network across regions.

*North-South traffic inspection* inspects bidirectional traffic flow between Azure VMware Solution and datacenters. North-south traffic inspection can use:

- A third-party firewall NVA and Azure Route Server over Azure internet.
- An on-premises default route over on-premises internet.
- Azure Firewall and Virtual WAN over Azure internet
- NSX-T Data Center within the SDDC over Azure VMware Solution internet.
- A third-party firewall NVA in Azure VMware Solution within the SDDC over Azure VMware Solution internet

## Ports and protocol requirements

Configure all necessary ports for an on-premises firewall to ensure proper access to all Azure VMware Solution private cloud components. For more information, see [Required network ports](#).

## Azure VMware Solution management access

- Consider using an Azure Bastion host in Azure Virtual Network to access the Azure VMware Solution environment during deployment.
- Once you establish routing to your on-premises environment, Azure VMware Solution management network doesn't honor the `0.0.0.0/0` routes from on-premises networks, so you need to advertise more specific routes for your on-premises networks.

## Business continuity, disaster recovery (BCDR), and migrations

- In VMware HCX migrations, the default gateway remains on-premises. For more information, see [Deploy and configure VMware HCX](#).
- VMware HCX migrations can use HCX L2 extension. Migrations that require Layer 2 extension also require ExpressRoute. S2S VPN is supported as long as the minimum [network underlay minimum requirements](#) are met. Maximum transmission unit (MTU) size should be 1350 to accommodate the overhead of HCX. For more information about Layer 2 extension design, see [Layer 2 bridging in manager mode \(VMware.com\)](#).

## Next steps

- For more information about Azure VMware Solution in hub-and-spoke networks, see [Integrate Azure VMware Solution in a hub and spoke architecture](#).
- For more information on VMware NSX-T Data Center network segments, see [Configure NSX-T Data Center network components using Azure VMware Solution](#).
- To learn Cloud Adoption Framework enterprise-scale landing zone architectural principles, various design considerations, and best practices for Azure VMware Solution, see the next article in this series:

[Single Region Hub & Spoke Topologies](#)

# Private Link and DNS integration at scale

Article • 10/25/2023

This article describes how to integrate Azure Private Link for PaaS services with Azure Private DNS zones in hub and spoke network architectures.

## Introduction

Many customers build their network infrastructure in Azure using the hub and spoke network architecture, where:

- Networking shared services (such as network virtual appliances, ExpressRoute/VPN gateways, or DNS servers) deploy in the **hub** virtual network (VNet).
- **Spoke** VNets consume the shared services by way of VNet peering.

In hub and spoke network architectures, application owners are typically provided with an Azure subscription, which includes a VNet (*a spoke*) connected to the *hub* VNet. In this architecture, they can deploy their virtual machines and have private connectivity to other VNets or to on-premises networks by way of ExpressRoute or VPN.

A central network virtual appliance (NVA), such as Azure Firewall, provides Internet-outbound connectivity.

Many application teams build their solutions using a combination of Azure IaaS and PaaS resources. Some Azure PaaS services (such as SQL Managed Instance) can be deployed in customer VNets. As a result, traffic stays private within the Azure network and is fully routable from on-premises.

But some Azure PaaS services (such as Azure Storage or Azure Cosmos DB) can't be deployed in a customer's VNets and are accessible over their public endpoint. In some cases, this configuration causes a contention with a customer's security policies. Corporate traffic might not allow the deployment or accessing of corporate resources (such as a SQL database) over public endpoints.

Azure Private Link supports access to a [list of Azure services](#) over private endpoints, but it requires that you register those private endpoint records in a corresponding [private DNS zone](#).

This article describes how application teams can deploy Azure PaaS services in their subscriptions that are only accessible over private endpoints.

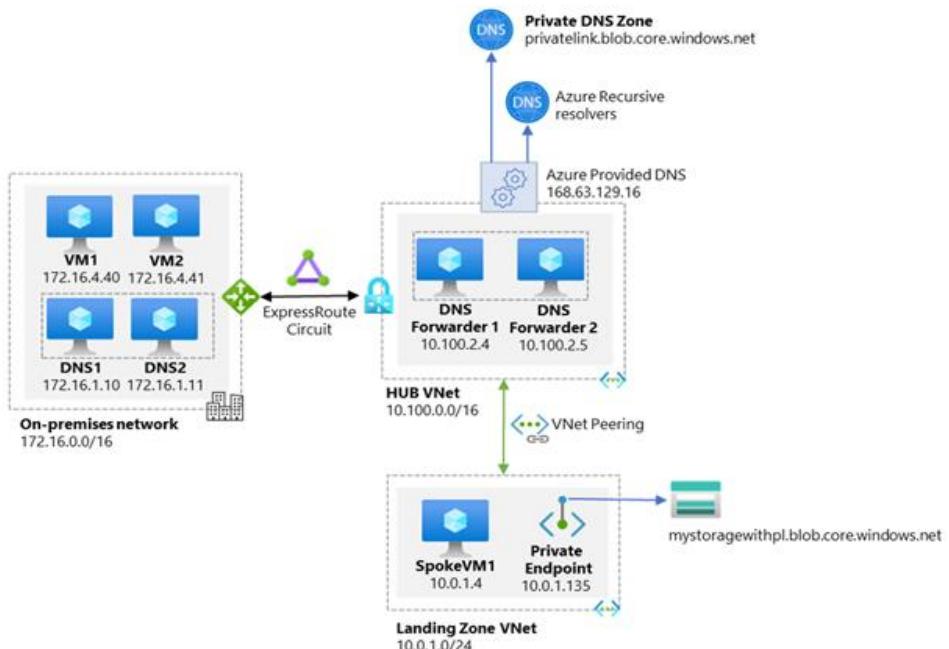
This article also describes how application teams can ensure that services automatically integrate with private DNS zones. They do the automation through Azure Private DNS, which removes the need to manually create or delete records in DNS.

## Private Link and DNS integration in hub and spoke network architectures

Private DNS zones are typically hosted centrally in the same Azure subscription where the hub VNet deploys. This central hosting practice is driven by [cross-premises DNS name resolution](#) and other needs for central DNS resolution such as Active Directory. In most cases, only networking and identity administrators have permissions to manage DNS records in the zones.

Application teams have permissions to create Azure resource in their own subscription. They don't have any permissions in the central networking connectivity subscription, which includes managing DNS records in the private DNS zones. This access limitation means they don't have the ability to [create the DNS records required](#) when deploying Azure PaaS services with private endpoints.

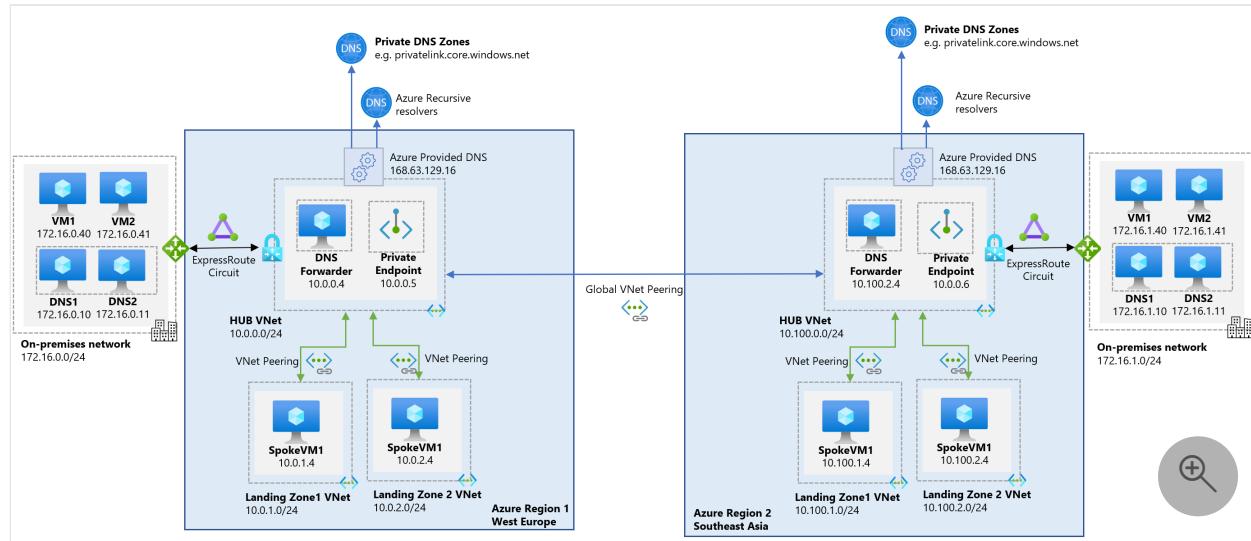
The following diagram shows a typical high-level architecture for enterprise environments with central DNS resolution and where name resolution for Private Link resources is done via Azure Private DNS:



From the previous diagram, it's important to highlight that:

- On-premises DNS servers have conditional forwarders configured for each private endpoint public DNS zone, pointing to the DNS servers `10.100.2.4` and `10.100.2.5` hosted in the hub VNet.
- The DNS servers `10.100.2.4` and `10.100.2.5` hosted in the hub VNet use the Azure-provided DNS resolver (`168.63.129.16`) as a forwarder.
- The hub VNet must be linked to the Private DNS zone names for Azure services (such as `privatelink.blob.core.windows.net`, as shown in the diagram).
- All Azure VNets use the DNS servers hosted in the hub VNet (`10.100.2.4` and `10.100.2.5`) as the primary and secondary DNS servers.
- If the DNS servers `10.100.2.4` and `10.100.2.5` aren't authoritative for customer's corporate domains (for example, Active Directory domain names), they should have conditional forwarders for the customer's corporate domains, pointing to the on-premises DNS Servers (`172.16.1.10` and `172.16.1.11`) or DNS servers deployed in Azure that are authoritative for such zones.

While the previous diagram depicts a single hub and spoke architecture, customers might need to extend their Azure footprint across multiple regions to address resiliency, proximity or data residency requirements, several scenarios have emerged where the same Private-Link-enabled PaaS instance must be accessed through multiple Private Endpoints (PE's).



The following diagram shows a typical high-level architecture for enterprise environments with central DNS resolution deployed in the hub (one per region) where name resolution for Private Link resources is done via Azure Private DNS.

It is recommended to deploy multiple regional private endpoints associated to the PaaS instance, one in each region where clients exist, enable per-region Private Link and Private DNS Zones. When working with PaaS services with built-in DR capabilities (geo-

redundant storage accounts, SQL DB failover groups, etc.), multiple region Private Endpoints are mandatory.

This scenario requires manual maintenance/updates of the Private Link DNS record set in every region as there is currently no automated lifecycle management for these.

For other use cases, a single global Private Endpoint can be deployed, making accessible to all clients by adding routing from the relevant regions to the single Private Endpoint in a single region.

To enable resolution, and therefore connectivity, from on premise networks to the `privatelink` private DNS zone and private endpoints, the appropriate DNS configuration (conditional forwarders etc.) need to be provisioned in the DNS infrastructure.

There are two conditions that must be true for application teams to create any required Azure PaaS resources in their subscription:

- Central networking and/or central platform team must ensure that application teams can only deploy and access Azure PaaS services by way of private endpoints.
- Central networking and/or central platform teams must ensure that when they create private endpoints, they set up how to handle the corresponding records. Set up the corresponding records such that they're automatically created in the centralized private DNS zone that matches the service being created.
- DNS records must follow the lifecycle of the private endpoint, in that, it's automatically removed when the private endpoint is deleted.

#### ⓘ Note

if FQDNs in network rules based on DNS resolution is needed to be used in Azure Firewall and Firewall policy (This capability allows you to filter outbound traffic with any TCP/UDP protocol -including NTP, SSH, RDP, and more-). You must enable Azure Firewall DNS Proxy to use FQDNs in your network rules, then those spoke VNets are forced to change their DNS setting from custom DNS server to Azure Firewall DNS Proxy. Changing the DNS settings of a spoke VNet requires reboot of all VMs inside that VNet.

The following sections describe how application teams enable these conditions by using [Azure Policy](#). The example uses Azure Storage as the Azure service that application teams need to deploy. But the same principle applies to most [Azure services that support Private Link](#).

# Configuration required by the platform team

The platform team configuration requirements include creating the private DNS zones, setting up policy definitions, deploying policies, and setting up the policy assignments.

## Create private DNS zones

Create private DNS zones in the central connectivity subscription for the supported Private Link services. For more information, see [Azure Private Endpoint DNS configuration](#).

In this case, **Storage account with blob** is the example. It translates to creating a `privatelink.blob.core.windows.net` private DNS zone in the connectivity subscription.

Name ↑↓	Type ↑↓	Location ↑↓
privatelink.blob.core.windows.net	Private DNS zone	Global

## Policy definitions

In addition to the private DNS zones, you also need to [create a set of custom Azure Policy definitions](#). These definitions enforce the use of private endpoints and automate creating the DNS record in the DNS zone that you create:

1. Deny public endpoint for PaaS services policy.

This policy prevents users from creating Azure PaaS services with public endpoints and gives them an error message if they don't select the private endpoint when creating the resource.

**Create storage account** ...

Basics Networking Data protection Advanced Tags Review + create

**Network connectivity**

You can connect to your storage account either publicly, via public IP addresses or service endpoints, or privately, using a private endpoint.

Connectivity method \*

Public endpoint (all networks)  
 Public endpoint (selected networks)  
 Private endpoint

ⓘ All networks will be able to access this storage account.  
Learn more about connectivity methods ↗

# Create storage account

 Validation failed. Click here to view details. →

Basics Networking Data protection Advanced Tags **Review + create**

**Summary** Raw Error

## ERROR DETAILS



Resource 'mystorageaccountwithpl' was disallowed by policy. (Code: RequestDisallowedByPolicy)

Initiative: Deny-Public-Endpoints-for-PaaS-Services  
Policy: Deny-PublicEndpoint-Storage

The exact policy rule might differ between PaaS services. For Azure Storage accounts, look at the **networkAcls.defaultAction** property that defines whether requests from public networks are allowed or not. In this case, set a condition to deny creating the **Microsoft.Storage/storageAccounts** resource type if the property **networkAcls.defaultAction** isn't `Deny`. The following policy definition shows the behavior:

JSON

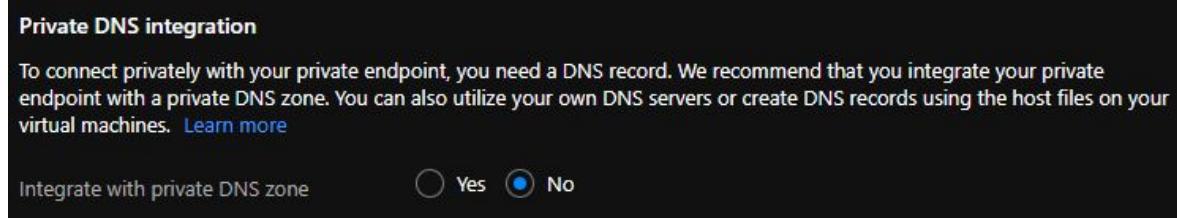
```
{  
  "mode": "All",  
  "policyRule": {  
    "if": {  
      "allOf": [  
        {  
          "field": "type",  
          "equals": "Microsoft.Storage/storageAccounts"  
        },  
        {  
          "field":  
            "Microsoft.Storage/storageAccounts/networkAcls.defaultAction",  
          "notEquals": "Deny"  
        }  
      ]  
    },  
    "then": {  
      "effect": "Deny"  
    }  
  }  
}
```

```
}
```

- Deny the ability to create a private DNS zone with the `privatelink` prefix policy.

Use a centralized DNS architecture with a conditional forwarder and private DNS zones hosted in the subscriptions managed by the platform team. It's necessary to prevent the application teams owners from creating their own Private Link private DNS zones and linking services into their subscriptions.

Ensure that when your application team creates a private endpoint, the option to `Integrate with private DNS zone` is set to `No` in the Azure portal.



If you select `Yes`, Azure Policy prevents you from creating the private endpoint. In the policy definition, it denies the ability to create the `Microsoft.Network/privateDnsZones` resource type if the zone has the `privatelink` prefix. The following policy definition shows the `privatelink` prefix:

JSON

```
{
  "description": "This policy restricts creation of private DNS zones with the `privatelink` prefix",
  "displayName": "Deny-PrivateDNSZone-PrivateLink",
  "mode": "All",
  "parameters": null,
  "policyRule": {
    "if": {
      "allOf": [
        {
          "field": "type",
          "equals": "Microsoft.Network/privateDnsZones"
        },
        {
          "field": "name",
          "contains": "privatelink."
        }
      ]
    },
    "then": {
      "effect": "Deny"
    }
  }
}
```

```
}
```

3. `DeployIfNotExists` policy to automatically create the required DNS record in the central private DNS zone.

The following policy examples show two approaches for identifying which `privateDNSZoneGroup` is created on a Private Endpoint.

The [first policy](#) relies on the `groupId` while the [second policy](#) uses both `privateLinkServiceId` and `groupID`. Use the [second policy](#) when `groupId` will clash (collide) with another resource.

For example, the `groupId` SQL is used for both Cosmos DB and Synapse Analytics. If both resource types deploy and the [first policy](#) has been assigned to create the `privateDNSZoneGroup` on the Private Endpoint entry, it's created and mapped to the incorrect Private DNS Zone, of either Cosmos DB or Synapse Analytics. It then might toggle between each of the zones due to the clashing `groupId` that the first policy looks for in its policy rule.

For a list of Private-link resources `groupId`, see the subresources column in [What is a private endpoint?](#).

### Tip

Azure Policy built-in definitions are constantly being added, deleted, and updated. It's highly recommended to use built-in policies versus managing your own policies (where available). Use the [AzPolicyAdvertiser](#) to find existing built-in policies that have the following name of 'xxx ... to use private DNS zones'. In addition, Azure Landing Zones (ALZ) has a policy initiative, [Configure Azure PaaS services to use private DNS zones](#), that contains built-in policies as well and periodically updated. If a built-in policy isn't available for your situation, consider creating an issue on the `azure-policy` feedback site [Azure Governance · Community](#) following the [New built-in Policy Proposals process on the Azure Policy GitHub repo.](#)

## First `DeployIfNotExists` Policy - Matching on `groupId` only

This policy triggers if you create a private endpoint resource with a service-specific `groupId`. The `groupId` is the ID of the group obtained from the remote resource (service) that this private endpoint should connect to. It then triggers a deployment of a

`privateDNSZoneGroup` within the private endpoint, which associates the private endpoint with the private DNS zone. In the example, the `groupId` for Azure Storage blobs is `blob`. For more information on the `groupId` for other Azure services, see [Azure Private Endpoint DNS configuration](#), under the **Subresource** column. When the policy finds the `groupId` in the private endpoint, it deploys a `privateDNSZoneGroup` within the private endpoint, and links it to the private DNS zone resource ID that's specified as the parameter. In the example, the private DNS zone resource ID is:

```
/subscriptions/<subscription-id>/resourceGroups/<resourceGroupName>/providers/Microsoft.Network/privateDnsZones/privatelink.blob.core.windows.net
```

The following code sample shows the policy definition:

JSON

```
{
  "mode": "Indexed",
  "policyRule": {
    "if": {
      "allOf": [
        {
          "field": "type",
          "equals": "Microsoft.Network/privateEndpoints"
        },
        {
          "count": {
            "field":
              "Microsoft.Network/privateEndpoints/privateLinkServiceConnections[*].groupId
              s[*]",
            "where": {
              "field":
                "Microsoft.Network/privateEndpoints/privateLinkServiceConnections[*].groupId
                s[*]",
              "equals": "blob"
            }
          }
        },
        {
          "greaterOrEquals": 1
        }
      ]
    },
    "then": {
      "effect": "deployIfNotExists",
      "details": {
        "type": "Microsoft.Network/privateEndpoints/privateDnsZoneGroups",
        "roleDefinitionIds": [
          "/providers/Microsoft.Authorization/roleDefinitions/4d97b98b-1d4f-
          4787-a291-c67834d212e7"
        ],
        "deployment": {
          "name": "DeployPrivateDnsZoneGroup"
        }
      }
    }
  }
}
```

```
"properties": {
    "mode": "incremental",
    "template": {
        "$schema": "https://schema.management.azure.com/schemas/2019-04-01/deploymentTemplate.json#",
        "contentVersion": "1.0.0.0",
        "parameters": {
            "privateDnsZoneId": {
                "type": "string"
            },
            "privateEndpointName": {
                "type": "string"
            },
            "location": {
                "type": "string"
            }
        },
        "resources": [
            {
                "name": "[concat(parameters('privateEndpointName'), '/deployedByPolicy')]",
                "type":
"Microsoft.Network/privateEndpoints/privateDnsZoneGroups",
                "apiVersion": "2020-03-01",
                "location": "[parameters('location')]",
                "properties": {
                    "privateDnsZoneConfigs": [
                        {
                            "name": "storageBlob-privateDnsZone",
                            "properties": {
                                "privateDnsZoneId": "
[parameters('privateDnsZoneId')]"
                            }
                        }
                    ]
                }
            },
            "parameters": {
                "privateDnsZoneId": {
                    "value": "[parameters('privateDnsZoneId')]"
                },
                "privateEndpointName": {
                    "value": "[field('name')]"
                },
                "location": {
                    "value": "[field('location')]"
                }
            }
        ]
    }
},
```

```

"parameters": {
    "privateDnsZoneId": {
        "type": "String",
        "metadata": {
            "displayName": "privateDnsZoneId",
            "strongType": "Microsoft.Network/privateDnsZones"
        }
    }
}

```

## Second `DeployIfNotExists` Policy - Matching on `groupId` & `privateLinkServiceId`

This policy triggers if you create a private endpoint resource with a service-specific `groupId` and `privateLinkServiceId`. The `groupId` is the ID of the group obtained from the remote resource (service) that this private endpoint should connect to. The `privateLinkServiceId` is the resource ID of the remote resource (service) this private endpoint should connect to. Then, trigger a deployment of a `privateDNSZoneGroup` within the private endpoint, which associates the private endpoint with the private DNS zone.

In the example, the `groupId` for Azure Cosmos DB (SQL) is `SQL` and the `privateLinkServiceId` must contain `Microsoft.DocumentDb/databaseAccounts`. For more information on the `groupId` and `privateLinkServiceId` for other Azure services, see [Azure Private Endpoint DNS configuration](#), under the **Subresource** column. When the policy finds `groupId` and `privateLinkServiceId` in the private endpoint, it deploys a `privateDNSZoneGroup` within the private endpoint. And it's linked to the private DNS zone resource ID that's specified as the parameter. The following policy definition shows the private DNS zone resource ID:

```

/subscriptions/<subscription-
id>/resourceGroups/<resourceGroupName>/providers/Microsoft.Network/privateDnsZones/
privatelink.documents.azure.com

```

The following code sample shows the policy definition:

JSON

```
{
    "mode": "Indexed",
    "policyRule": {
        "if": {
            "allOf": [
                {

```

```
        "field": "type",
        "equals": "Microsoft.Network/privateEndpoints"
    },
{
    "count": {
        "field":
"Microsoft.Network/privateEndpoints/privateLinkServiceConnections[*]",
        "where": {
            "allOf": [
                {
                    "field":
"Microsoft.Network/privateEndpoints/privateLinkServiceConnections[*].private
LinkServiceId",
                    "contains": "Microsoft.DocumentDb/databaseAccounts"
                },
                {
                    "field":
"Microsoft.Network/privateEndpoints/privateLinkServiceConnections[*].groupId
s[*]",
                    "equals": "[parameters('privateEndpointGroupId')]"
                }
            ]
        }
    },
    "greaterOrEquals": 1
}
],
{
    "then": {
        "effect": "[parameters('effect')]",
        "details": {
            "type": "Microsoft.Network/privateEndpoints/privateDnsZoneGroups",
            "roleDefinitionIds": [
                "/providers/Microsoft.Authorization/roleDefinitions/4d97b98b-1d4f-
4787-a291-c67834d212e7"
            ],
            "deployment": {
                "properties": {
                    "mode": "incremental",
                    "template": {
                        "$schema": "https://schema.management.azure.com/schemas/2019-
04-01/deploymentTemplate.json#",
                        "contentVersion": "1.0.0.0",
                        "parameters": {
                            "privateDnsZoneId": {
                                "type": "string"
                            },
                            "privateEndpointName": {
                                "type": "string"
                            },
                            "location": {
                                "type": "string"
                            }
                        },
                        "resources": [

```

```

        {
            "name": "[concat(parameters('privateEndpointName'),
'/_deployedByPolicy')]",
            "type": "Microsoft.Network/privateEndpoints/privateDnsZoneGroups",
            "apiVersion": "2020-03-01",
            "location": "[parameters('location')]",
            "properties": {
                "privateDnsZoneConfigs": [
                    {
                        "name": "cosmosDB-privateDnsZone",
                        "properties": {
                            "privateDnsZoneId": "[parameters('privateDnsZoneId')]"
                        }
                    }
                ]
            }
        },
        "parameters": {
            "privateDnsZoneId": {
                "value": "[parameters('privateDnsZoneId')]"
            },
            "privateEndpointName": {
                "value": "[field('name')]"
            },
            "location": {
                "value": "[field('location')]"
            }
        }
    },
    "parameters": {
        "privateDnsZoneId": {
            "type": "String",
            "metadata": {
                "displayName": "Private Dns Zone Id",
                "description": "The private DNS zone to deploy in a new private DNS zone group and link to the private endpoint",
                "strongType": "Microsoft.Network/privateDnsZones"
            }
        },
        "privateEndpointGroupId": {
            "type": "String",
            "metadata": {
                "displayName": "Private Endpoint Group Id",
                "description": "A group Id for the private endpoint"
            }
        },
        "effect": {

```

```
        "type": "String",
        "metadata": {
            "displayName": "Effect",
            "description": "Enable or disable the execution of the policy"
        },
        "allowedValues": [
            "DeployIfNotExists",
            "Disabled"
        ],
        "defaultValue": "DeployIfNotExists"
    }
}
}
```

## Policy assignments

After policy definitions are deployed, [assign the policies](#) at the desired scope in your management group hierarchy. Ensure that the policy assignments target the Azure subscriptions the application teams use to deploy PaaS services with private endpoint access exclusively.

### ⓘ Important

In addition to [assigning the roleDefinition](#) defined in the policy, remember to assign the [Private DNS Zone Contributor role](#) role in the subscription and resource group where the private DNS zones are hosted to the [managed identity created by the DeployIfNotExists policy assignment](#) that will be responsible to create and manage the private endpoint DNS record in the private DNS zone. This is because the private endpoint is located in the application owner Azure subscription, while the private DNS zone is located in a different subscription (such as central connectivity subscription).

After the platform team finishes the configuration:

- The applications teams' Azure subscriptions are ready for the team to then create Azure PaaS services with private endpoint access exclusively.
- The team must ensure the DNS records for private endpoints are automatically registered (and removed once a private endpoint is deleted) from the corresponding private DNS zones.

## Application owner experience

After the platform team deploys the platform infrastructure components (private DNS zones and policies), the application owner has the following experience when they try to deploy an Azure PaaS service into the Azure subscription. This experience is the same whether they do their activities through the Azure portal or other clients, such as PowerShell or CLI, since Azure policies govern their subscriptions.

1. Create a storage account through the Azure portal. In the **Basics** tab, choose the settings you want, provide a name for your storage account, and select **Next**.

The screenshot shows the 'Create storage account' wizard in the Azure portal, specifically the 'Basics' tab. The tab navigation bar includes 'Basics', 'Networking', 'Data protection', 'Advanced', 'Tags', and 'Review + create'. Below the tabs, a descriptive text explains Azure Storage's features and cost. A link to 'Learn more about Azure storage accounts' is provided. The 'Project details' section asks to select a subscription and resource group. The 'Subscription' dropdown is set to 'Azure Internal Subscription'. The 'Resource group' dropdown is set to 'storage' with a 'Create new' option below it. The 'Instance details' section asks for the storage account name and location. The 'Storage account name' input field contains 'mystoragewithpl' with a green checkmark icon. The 'Location' dropdown is set to '(Europe) North Europe'. The entire form is contained within a dark-themed card.

2. In the networking tab, select **Private endpoint**. If you select an option other than **Private endpoint**, the Azure portal won't allow you to create the storage account in the **Review + create** section of the deployment wizard. The policy prevents you from creating this service if the public endpoint is enabled.

## Create storage account

Basics Networking Data protection Advanced Tags Review + create

### Network connectivity

You can connect to your storage account either publicly, via public IP addresses or service endpoints, or privately, using a private endpoint.

Connectivity method \*

- Public endpoint (all networks)
- Public endpoint (selected networks)
- Private endpoint

### Private endpoint

Create a private endpoint to allow a private connection to this resource. Additional private endpoint connections can be created. [Learn more about private endpoints](#)

Name	Subscription	Resource group	Region	Target sub-resource
Click on add to create a private endpoint				

[Add](#)

### Network routing

Determine how to route your traffic as it travels from the source to its Azure endpoint. Microsoft network routing is recommended.

Routing preference \*

- Microsoft network routing (default)
- Internet routing

[Review + create](#)

[< Previous](#)

[Next : Data protection >](#)

3. It's possible to create the private endpoint now or after you create the storage account. This example shows creating the private endpoint after the storage account is created. Select **Review + create** to complete the step.

4. After you create the storage account, make a private endpoint through the Azure portal.

## Create a private endpoint

1 Basics

2 Resource

3 Configuration

4 Tags

5 Review + create

Use private endpoints to privately connect to a service or resource. Your private endpoint must be in the same region as your virtual network, but can be in a different region from the private link resource that you are connecting to. [Learn more](#)

### Project details

Subscription \* ⓘ

Azure Internal Subscription

Resource group \* ⓘ

storage

[Create new](#)

### Instance details

Name \*

mystoragewithpl-pe

Region \*

(Europe) North Europe

5. In the **Resource** section, locate the storage account you created in the previous step. Under target subresource, select **Blob**, and then select **Next**.

## Create a private endpoint

✓ Basics    2 Resource    3 Configuration    4 Tags    5 Review + create

Private Link offers options to create private endpoints for different Azure resources, like your private link service, a SQL server, or an Azure storage account. Select which resource you would like to connect to using this private endpoint. [Learn more](#)

Connection method ⓘ  Connect to an Azure resource in my directory.  Connect to an Azure resource by resource ID or alias.

Subscription \* ⓘ

Resource type \* ⓘ

Resource \* ⓘ

Target sub-resource \* ⓘ

6. In the **Configuration** section, after selecting your VNet and subnet, be sure that **Integrate with private DNS zone** is set to **No**. Otherwise, the Azure portal prevents you from creating the private endpoint. Azure Policy won't allow you to create a private DNS zone with the `privatelink` prefix.

✓ Basics    ✓ Resource    3 Configuration    4 Tags    5 Review + create

**Networking**

To deploy the private endpoint, select a virtual network subnet. [Learn more](#)

Virtual network \* ⓘ

Subnet \* ⓘ

ⓘ If you have a network security group (NSG) enabled for the subnet above, it will be disabled for private endpoints on this subnet only. Other resources on the subnet will still have NSG enforcement.

**Private DNS integration**

To connect privately with your private endpoint, you need a DNS record. We recommend that you integrate your private endpoint with a private DNS zone. You can also utilize your own DNS servers or create DNS records using the host files on your virtual machines. [Learn more](#)

Integrate with private DNS zone  Yes  No

7. Select **Review + create**, and then select **Create** to deploy the private endpoint.
8. After a few minutes, the `DeployIfNotExists` policy triggers. The subsequent `dnsZoneGroup` deployment then adds the required DNS records for the private endpoint in the centrally managed DNS zone.
9. After you create the private endpoint, select it, and review its FQDN and private IP:

mystoragewithpl-pe | DNS configuration

Private endpoint | Directory: Microsoft

Search (Ctrl+ /) <> Add configuration Refresh

Overview Activity log Access control (IAM) Tags Diagnose and solve problems

Settings

DNS configuration Properties Locks

Private DNS integration To connect privately with your private endpoint, you need a DNS record. We recommend that you integrate your private endpoint using a private DNS zone. You can also utilize your own DNS servers. [Learn more](#)

Custom DNS records To be configured correctly, the following FQDNs are required in your private DNS setup. [Learn more](#)

FQDN	IP addresses
mystoragewithpl.blob.core.windows.net	10.1.0.132

10. Check the activity log for the resource group where the private endpoint was created. Or you can check the activity log of the private endpoint itself. You'll notice that after a few minutes, a `DeployIfNotExist` policy action runs and that configures the DNS zone group on the private endpoint:

mystoragewithpl-pe | Activity log

Private endpoint | Directory: Microsoft

Search (Ctrl+ /) <> Activity Edit columns Refresh Diagnostics settings Download as CSV Logs Pin current filters Reset filters

Overview Activity log Access control (IAM) Tags Diagnose and solve problems

Settings

DNS configuration Properties Locks

Monitoring

Alerts Metrics

Automation

Tasks (preview) Export template

Management Group : None Subscription : Azure Internal Subscription Event severity : All Timespan : Last 6 hours Resource

+ Add Filter

11 items.

Operation name	Status	Time	Time stamp	Subscription
DeployIfNotExists' Policy action.	Succeeded	22 minutes ...	Tue Mar 09 ...	Azure Internal Subscription
DeployIfNotExists	Accepted	22 minutes ...	Tue Mar 09 ...	Azure Internal Subscription
'DeployIfNotExists' Policy action.	Accepted	22 minutes ...	Tue Mar 09 ...	Azure Internal Subscription
Put Private DNS Zone Group	Started	22 minutes ...	Tue Mar 09 ...	Azure Internal Subscription
Put Private DNS Zone Group	Started	22 minutes ...	Tue Mar 09 ...	Azure Internal Subscription
Put Private DNS Zone Group	Accepted	22 minutes ...	Tue Mar 09 ...	Azure Internal Subscription
Put Private DNS Zone Group	Accepted	22 minutes ...	Tue Mar 09 ...	Azure Internal Subscription
Write PrivateDnsZoneGroups	Succeeded	22 minutes ...	Tue Mar 09 ...	Azure Internal Subscription
Put Private DNS Zone Group	Succeeded	22 minutes ...	Tue Mar 09 ...	Azure Internal Subscription
DeployIfNotExists	Succeeded	22 minutes ...	Tue Mar 09 ...	Azure Internal Subscription

11. If the central networking team goes to the `privatelink.blob.core.windows.net` private DNS zone, they'll confirm that the DNS record is there for the private endpoint you created, and both the name and IP address match the values within the private endpoint.

The screenshot shows the Azure portal interface for managing a private DNS zone. The left sidebar has sections for Overview, Activity log, Access control (IAM), Tags, Diagnose and solve problems, Settings (Virtual network links, Properties, Locks, Monitoring, Alerts, Metrics), and a search bar for record sets. The main area displays the 'Essentials' section with fields for Resource group, Subscription, Subscription ID, and Tags. Below this is a table of record sets:

Name	Type	TTL	Value
@	SOA	3600	Email: azureprivatedns-host.microsoft.com Host: azureprivatedns.net Refresh: 3600 Retry: 300 Expire: 2419200 Minimum TTL: 10 Serial number: 1
mystoragewithpl	A	10	10.1.0.132

At this point, application teams can use the storage account through a private endpoint from any VNet in the hub and spoke network environment and from on-premises. The DNS record has been automatically recorded in the private DNS zone.

If an application owner deletes the private endpoint, the corresponding records in the private DNS zone are automatically removed.

## Next steps

Review [DNS for on-premises and Azure resources](#). Review [Plan for virtual machine remote access](#).

### Important

This article outlines DNS and Private link integration at scale using DINE (DeployIfNotExists) policies assigned to the Management Group. Which means there is no need to handle the DNS integration in code when creating Private Endpoints with this approach, as it is handled by the policies. It is also unlikely that the application teams have RBAC access to the centralized Private DNS Zones also.

Below are helpful links to review when creating Private Endpoint with Bicep and HashiCorp Terraform.

For Private Endpoint creation with Infrastructure-as-Code:

- [Quickstart Create a private endpoint using Bicep](#).
- Create a private endpoint using HashiCorp Terraform [azurerm\\_private\\_endpoint](#) in Terrafrom Registry.

You can still create private endpoints in your Infrastructure-as-Code tooling however, if using the DINE policy approach as outlined in this article you should leave the DNS integration side out of your code and let the DINE policies that have the required RBAC to the Private DNS Zones handle this instead.

# SD-WAN connectivity architecture with Azure Virtual WAN

Article • 08/24/2023

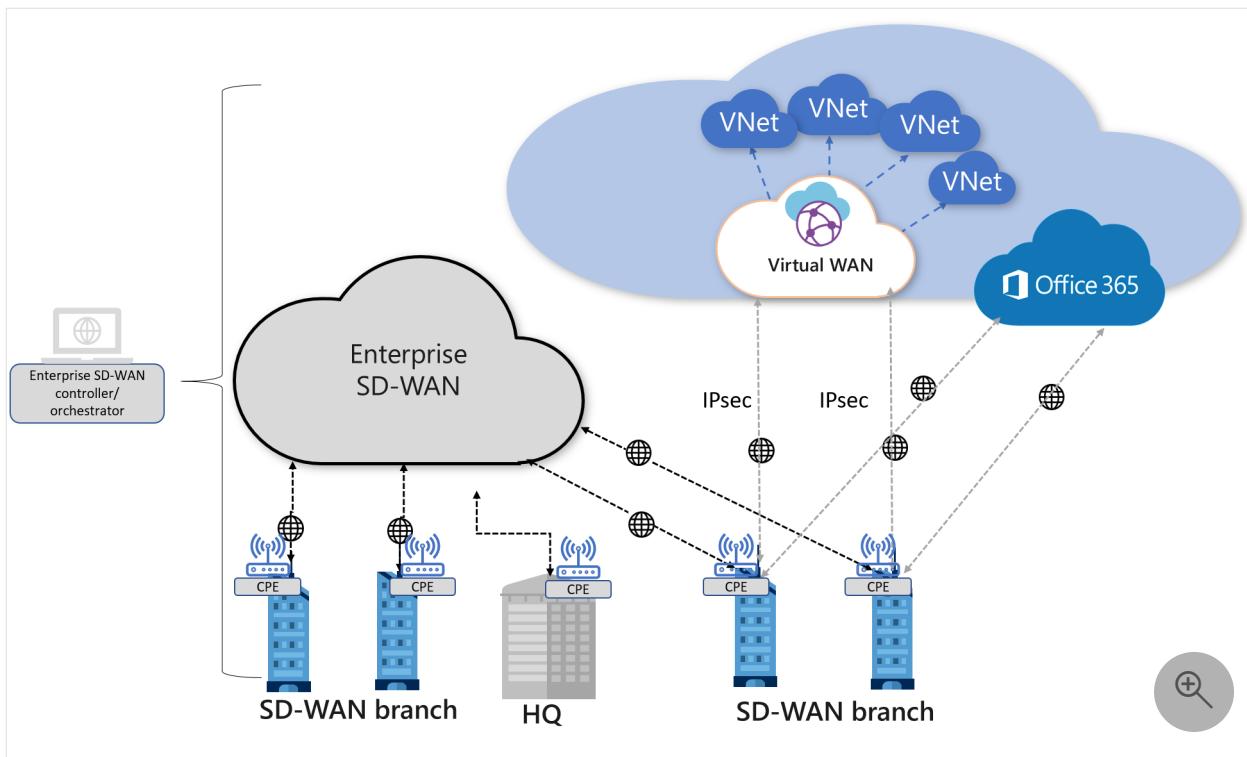
Azure Virtual WAN is a networking service that brings together many cloud connectivity and security services with a single operational interface. These services include branch (via Site-to-site VPN), remote user (Point-to-site VPN), private (ExpressRoute) connectivity, intra-cloud transitive connectivity for VNets, VPN and ExpressRoute interconnectivity, routing, Azure Firewall, and encryption for private connectivity.

Although Azure Virtual WAN is a cloud-based SD-WAN that provides a rich suite of Azure first-party connectivity, routing, and security services, Azure Virtual WAN also is designed to enable seamless interconnection with premises-based SD-WAN and SASE technologies and services. Many such services are offered by our [Virtual WAN](#) ecosystem and Azure Networking Managed Services partners ([MSPs](#)). Enterprises that are transforming their private WAN to SD-WAN have options when interconnecting their private SD-WAN with Azure Virtual WAN. Enterprises can choose from these options:

- Direct Interconnect model
- Direct Interconnect model with NVA-in-VWAN-hub
- Indirect Interconnect model
- Managed Hybrid WAN model using their favorite managed service provider [MSP](#)

In all of these cases, the interconnection of Virtual WAN with SD-WAN is similar from the connectivity side, but may vary on the orchestration and operational side.

## Direct Interconnect model



In this architecture model, the SD-WAN branch customer-premises equipment (CPE) is directly connected to Virtual WAN hubs via IPsec connections. The branch CPE may also be connected to other branches via the private SD-WAN, or use Virtual WAN for branch to branch connectivity. Branches that need to access their workloads in Azure will be able to directly and securely access Azure via the IPsec tunnel(s) that are terminated in the Virtual WAN hub(s).

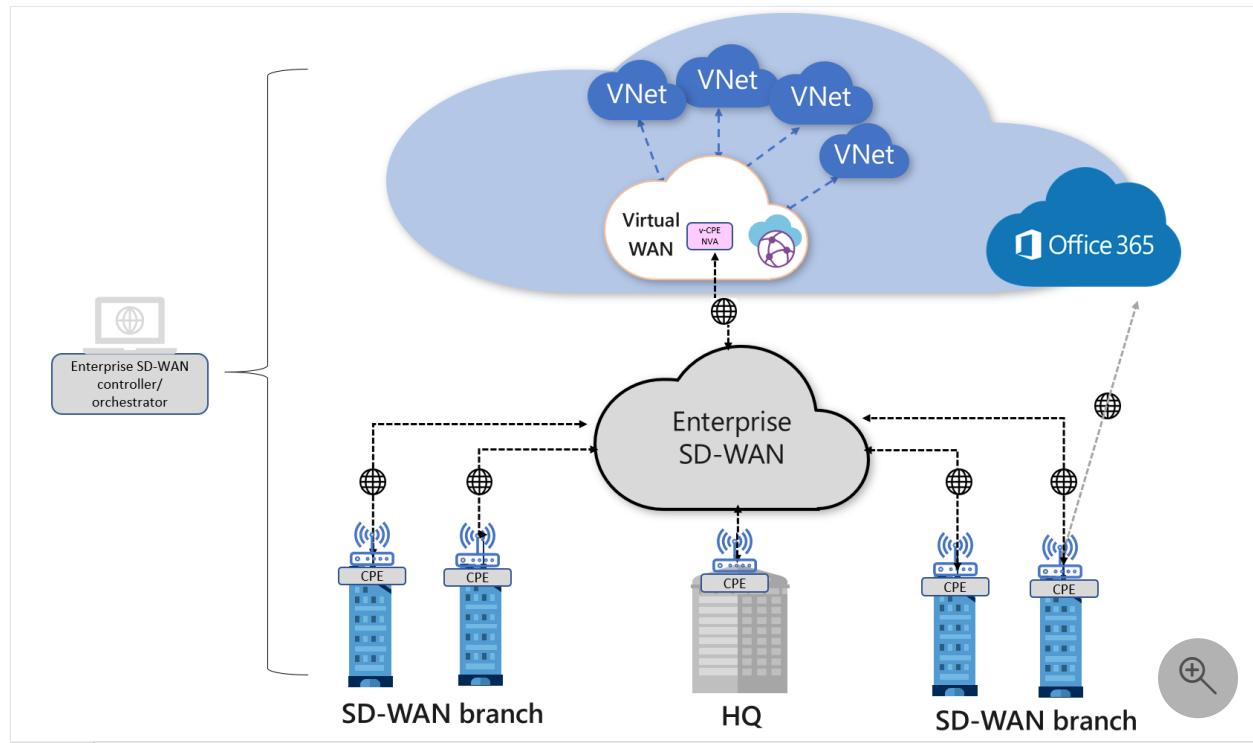
SD-WAN CPE partners can enable automation in order to automate the normally tedious and error-prone IPsec connectivity from their respective CPE devices. Automation allows the SD-WAN controller to talk to Azure via the Virtual WAN API to configure the Virtual WAN sites, and push necessary IPsec tunnel configuration to the branch CPEs. See [Automation guidelines](#) for the description of Virtual WAN interconnection automation by various SD-WAN partners.

The SD-WAN CPE continues to be the place where traffic optimization and path selection is implemented and enforced.

In this model, some vendor proprietary traffic optimization based on real-time traffic characteristics may not be supported because the connectivity to Virtual WAN is over IPsec and the IPsec VPN is terminated on the Virtual WAN VPN gateway. For example, dynamic path selection at the branch CPE is feasible due to the branch device exchanging various network packet information with another SD-WAN node, hence identifying the best link to use for various prioritized traffic dynamically at the branch. This feature may be useful in areas where last mile optimization (branch to the closest Microsoft POP) is required.

With Virtual WAN, users can get Azure Path Selection, which is policy-based path selection across multiple ISP links from the branch CPE to Virtual WAN VPN gateways. Virtual WAN allows for the setup of multiple links (paths) from the same SD-WAN branch CPE; each link represents a dual tunnel connection from a unique public IP of the SD-WAN CPE to two different instances of Azure Virtual WAN VPN gateway. SD-WAN vendors can implement the most optimal path to Azure, based on traffic policies set by their policy engine on the CPE links. On the Azure end, all connections coming in are treated equally.

## Direct Interconnect model with NVA-in-VWAN-hub



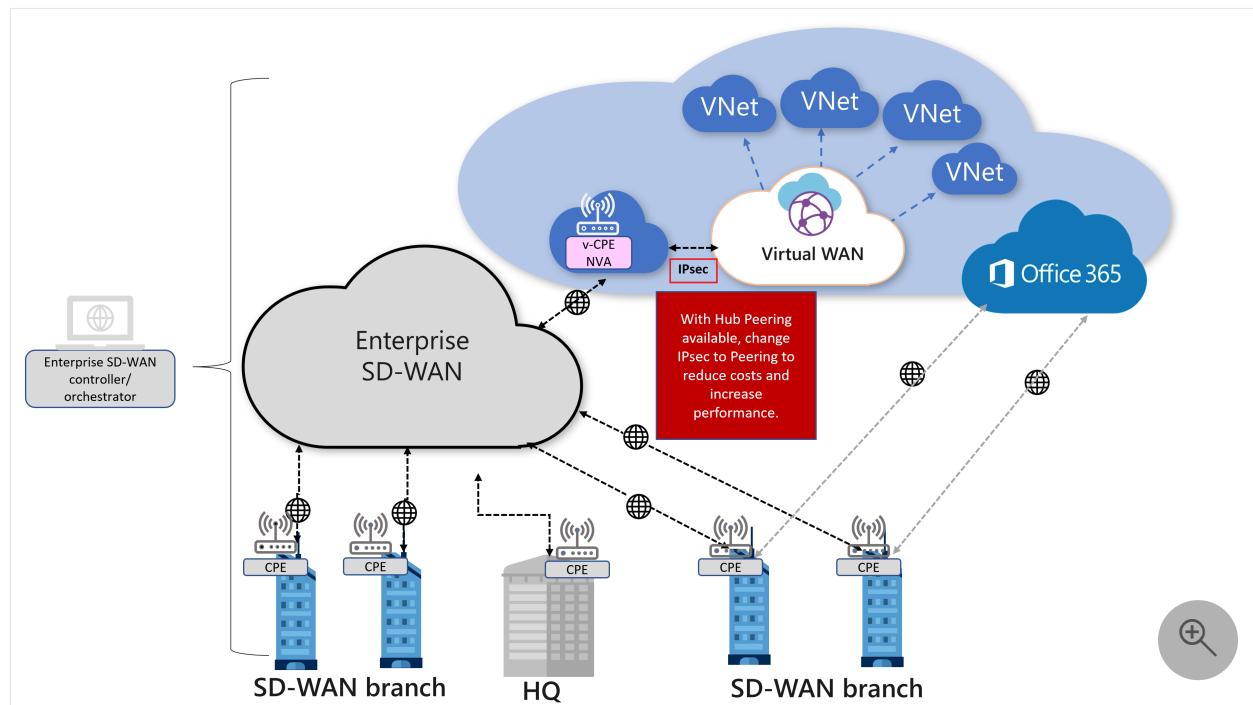
This architecture model supports the deployment of a third-party [Network Virtual Appliance \(NVA\)](#) directly into the virtual hub. This allows customers who want to connect their branch CPE to the same brand NVA in the virtual hub so that they can take advantage of proprietary end-to-end SD-WAN capabilities when connecting to Azure workloads.

Several Virtual WAN Partners have worked to provide an experience that configures the NVA automatically as part of the deployment process. Once the NVA has been provisioned into the virtual hub, any additional configuration that may be required for the NVA must be done via the NVA partners portal or management application. Direct access to the NVA isn't available. The NVAs that are available to be deployed directly into the Azure Virtual WAN hub are engineered specifically to be used in the virtual hub.

For partners that support NVA in VWAN hub and their deployment guides, please see the [Virtual WAN Partners](#) article.

The SD-WAN CPE continues to be the place where traffic optimization and path selection is implemented and enforced. In this model, vendor proprietary traffic optimization based on real-time traffic characteristics is supported because the connectivity to Virtual WAN is via the SD-WAN NVA in the hub.

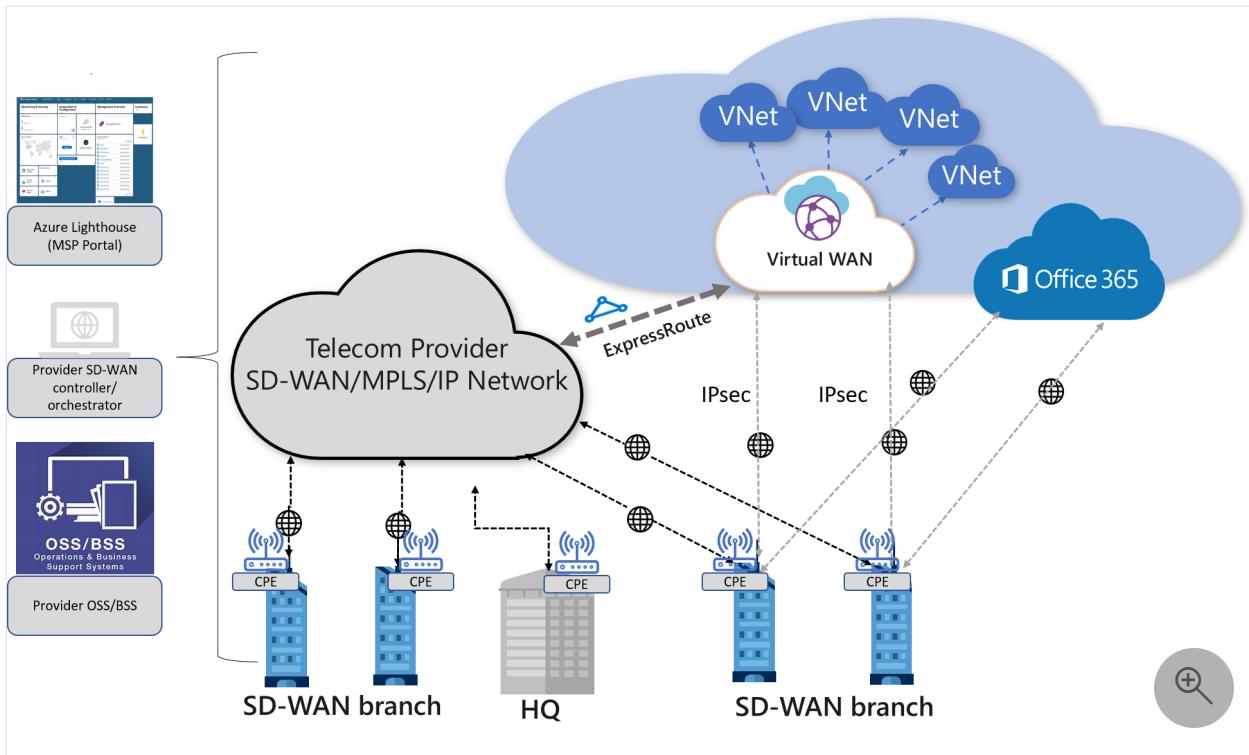
## Indirect Interconnect model



In this architecture model, SD-WAN branch CPEs are indirectly connected to Virtual WAN hubs. As the figure shows, an SD-WAN virtual CPE is deployed in an enterprise VNet. This virtual CPE is, in turn, connected to the Virtual WAN hub(s) using IPsec. The virtual CPE serves as an SD-WAN gateway into Azure. Branches that need to access their workloads in Azure will be able to access them via the v-CPE gateway.

Since the connectivity to Azure is via the v-CPE gateway (NVA), all traffic to and from Azure workload VNets to other SD-WAN branches go via the NVA. In this model, the user is responsible for managing and operating the SD-WAN NVA including high availability, scalability, and routing.

## Managed Hybrid WAN model



In this architecture model, enterprises can leverage a managed SD-WAN service offered by a Managed Service Provider (MSP) partner. This model is similar to the direct or indirect models described above. However, in this model, the SD-WAN design, orchestration, and operations are delivered by the SD-WAN Provider.

Azure Networking MSP partners can use [Azure Lighthouse](#) to implement the SD-WAN and Virtual WAN service in the enterprise customer's Azure subscription, as well as operate the end-to-end hybrid WAN on behalf of the customer. These MSPs may also be able to implement Azure ExpressRoute into the Virtual WAN and operate it as an end-to-end managed service.

## Additional information

- [Azure Virtual WAN FAQ](#)
- [Solving Remote Connectivity](#)

# Traditional Azure networking topology

Article • 03/27/2023

## ⓘ Important

Try the new [Topology \(Preview\)](#) experience, which offers a visualization of Azure resources for ease of inventory management and monitoring network at scale. Use the Topology preview to visualize resources and their dependencies across subscriptions, regions and locations. [Select this link](#) to navigate to the experience.

Explore key design considerations and recommendations surrounding network topologies in Microsoft Azure.

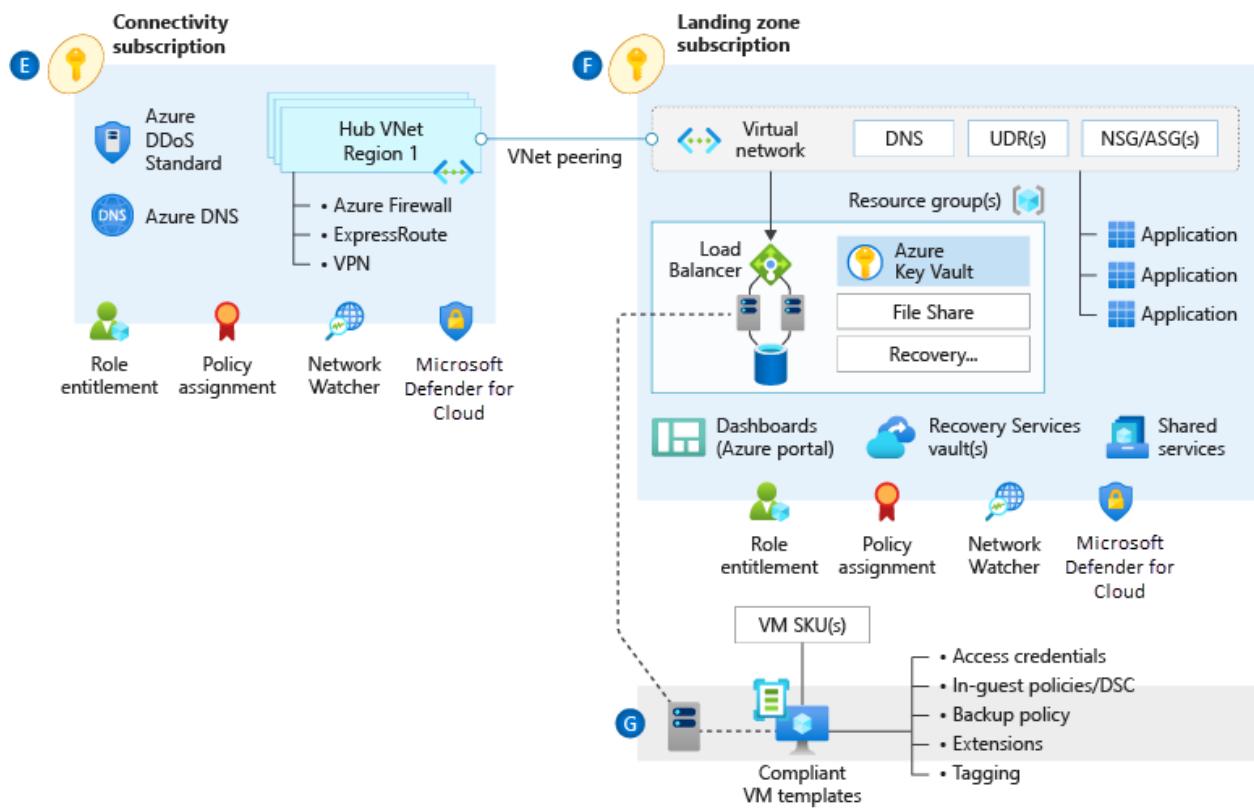


Figure 1: A traditional Azure network topology.

## Design considerations:

- Various network topologies can connect multiple landing zone virtual networks. Examples of network topologies include one large flat virtual network, multiple virtual networks connected with multiple Azure ExpressRoute circuits or connections, hub-and-spoke, full mesh, and hybrid.

- Virtual networks can't traverse subscription boundaries. However, you can achieve connectivity between virtual networks across different subscriptions by using virtual network peering, an ExpressRoute circuit, or VPN gateways.
- Virtual network peering is the preferred method to connect virtual networks in Azure. You can use virtual network peering to connect virtual networks in the same region, across different Azure regions, and across different Azure Active Directory (Azure AD) tenants.
- Virtual network peering and global virtual network peering aren't transitive. To enable a transit network, you need user-defined routes (UDRs) and network virtual appliances (NVAs). For more information, see [Hub-spoke network topology in Azure](#).
- You can share an Azure DDoS Protection plan across all virtual networks in a single Azure AD tenant to protect resources with public IP addresses. For more information, see [Azure DDoS Protection](#).
  - Azure DDoS Protection plans cover only resources with public IP addresses.
  - The cost of an Azure DDoS Protection plan includes 100 public IP addresses across all protected virtual networks associated with the DDoS Protection plan. Protection for more resources is available at a separate cost. For more information on Azure DDoS Protection plan pricing, see the [Azure DDoS Protection pricing page](#) or the [FAQ](#).
  - Review the [supported resources of Azure DDoS Protection plans](#).
- You can use ExpressRoute circuits to establish connectivity across virtual networks within the same geopolitical region or by using the premium add-on for connectivity across geopolitical regions. Keep these points in mind:
  - Network-to-network traffic might experience more latency, because traffic must hairpin at the Microsoft Enterprise Edge (MSEE) routers.
  - The ExpressRoute gateway SKU constrains bandwidth.
  - Deploy and manage UDRs if you need to inspect or log UDRs for traffic across virtual networks.
- VPN gateways with Border Gateway Protocol (BGP) are transitive within Azure and on-premises networks, but they don't provide transitive access to networks connected via ExpressRoute by default. If you need transitive access to networks connected via ExpressRoute, consider [Azure Route Server](#).

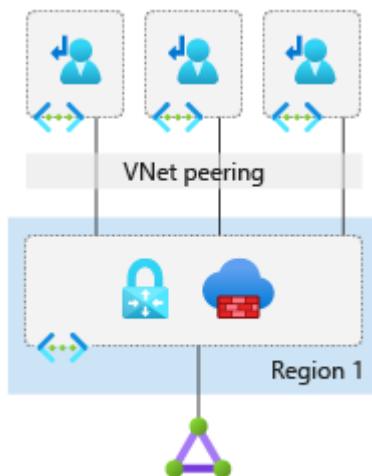
- When you connect multiple ExpressRoute circuits to the same virtual network, use connection weights and BGP techniques to ensure an optimal path for traffic between on-premises networks and Azure. For more information, see [Optimize ExpressRoute routing](#).
- Using BGP metrics to influence ExpressRoute routing is a configuration change made outside of the Azure platform. Your organization or your connectivity provider must configure the on-premises routers accordingly.
- ExpressRoute circuits with premium add-ons provide global connectivity.
- ExpressRoute has certain limits; there are a maximum number of ExpressRoute connections per ExpressRoute gateway, and ExpressRoute private peering can identify a maximum number of routes from Azure to on-premises. For more information about ExpressRoute limits, see [ExpressRoute limits](#).
- A VPN gateway's maximum aggregated throughput is 10 gigabits per second. A VPN gateway supports up to 100 site-to-site or network-to-network tunnels.
- If an NVA is part of the architecture, consider Azure Route Server to simplify dynamic routing between your network virtual appliance (NVA) and your virtual network. Azure Route Server allows you to exchange routing information directly through Border Gateway Protocol (BGP) routing protocol between any NVA that supports the BGP routing protocol and the Azure software defined network (SDN) in the Azure virtual network (VNet) without the need to manually configure or maintain route tables.

#### **Design recommendations:**

- Consider a network design based on the traditional hub-and-spoke network topology for the following scenarios:
  - A network architecture deployed within a single Azure region.
  - A network architecture that spans multiple Azure regions, with no need for transitive connectivity between virtual networks for landing zones across regions.
  - A network architecture that spans multiple Azure regions, and global virtual network peering that can connect virtual networks across Azure regions.
  - There's no need for transitive connectivity between VPN and ExpressRoute connections.

- The main hybrid connectivity method in place is ExpressRoute, and the number of VPN connections is less than 100 per VPN Gateway.
- There's a dependency on centralized NVAs and granular routing.
- For regional deployments, primarily use the hub-and-spoke topology. Use landing zone virtual networks that connect with virtual network peering to a central hub virtual network for the following scenarios:
  - Cross-premises connectivity via ExpressRoute.
  - VPN for branch connectivity.
  - Spoke-to-spoke connectivity via NVAs and UDRs.
  - Internet-outbound protection via Azure Firewall or another third-party NVA.

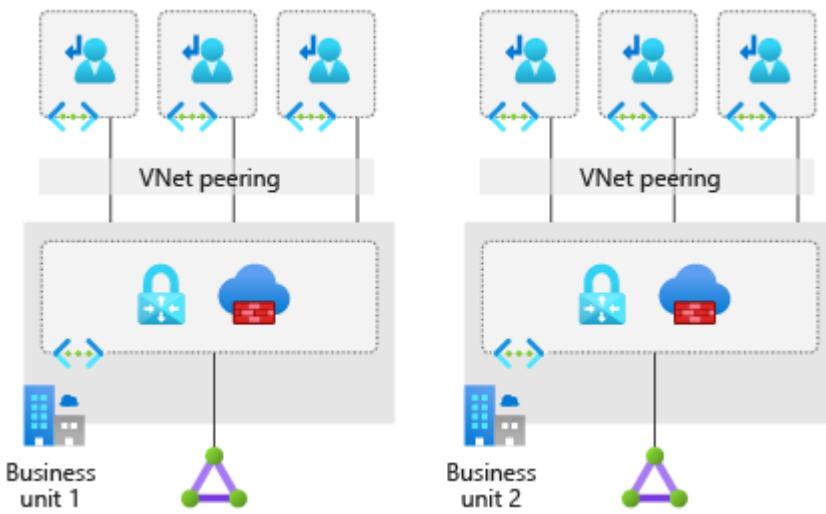
The following diagram shows the hub-and-spoke topology. This configuration allows for appropriate traffic control to meet most requirements for segmentation and inspection.



*Figure 2: Hub-and-spoke network topology.*

- Use the topology of multiple virtual networks connected with multiple ExpressRoute circuits when one of these conditions is true:
  - You need a high level of isolation.
  - You need dedicated ExpressRoute bandwidth for specific business units.
  - You've reached the maximum number of connections per ExpressRoute gateway (refer to the [ExpressRoute limits](#) article for the maximum number).

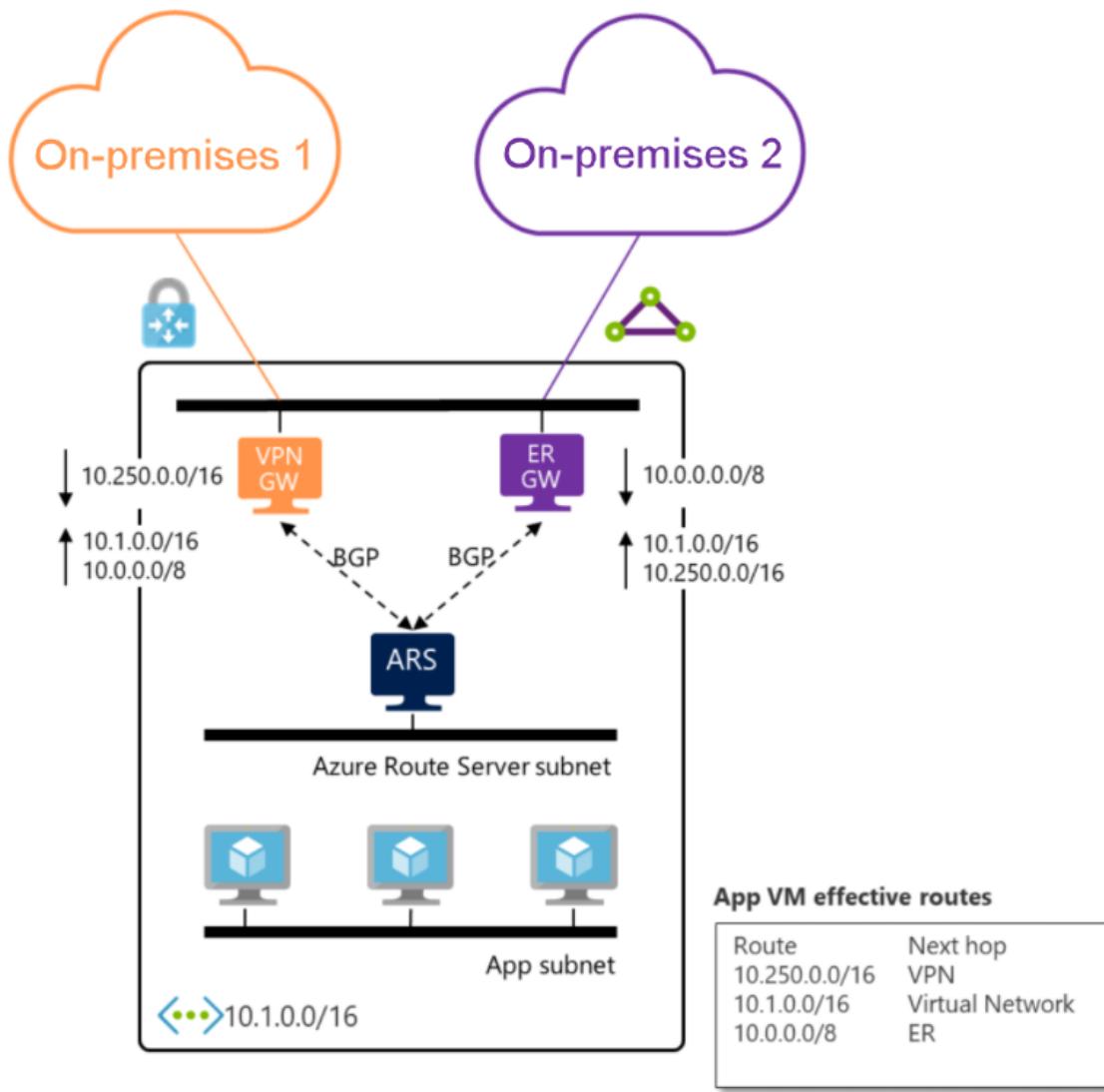
The following figure shows this topology.



*Figure 3: Multiple virtual networks connected with multiple ExpressRoute circuits.*

- Deploy a set of minimal shared services, including ExpressRoute gateways, VPN gateways (as required), and Azure Firewall or partner NVAs (as required) in the central-hub virtual network. If necessary, also deploy Active Directory domain controllers and DNS servers.
- Deploy Azure Firewall or partner NVAs for east/west or south/north traffic protection and filtering, in the central-hub virtual network.
- When you're deploying partner networking technologies or NVAs, follow the partner vendor's guidance to ensure that:
  - The vendor supports deployment.
  - The guidance supports high availability and maximum performance.
  - There are no conflicting configurations with Azure networking.
- Don't deploy Layer 7 inbound NVAs, such as Azure Application Gateway, as a shared service in the central-hub virtual network. Instead, deploy them together with the application in their respective landing zones.
- Deploy a single Azure DDoS standard protection plan in the connectivity subscription.
  - All landing zone and platform virtual networks should use this plan.
- Use your existing network, multiprotocol label switching, and SD-WAN to connect branch locations with corporate headquarters. If you don't use Azure Route Server, then there's no support for transit in Azure between ExpressRoute and VPN gateways.

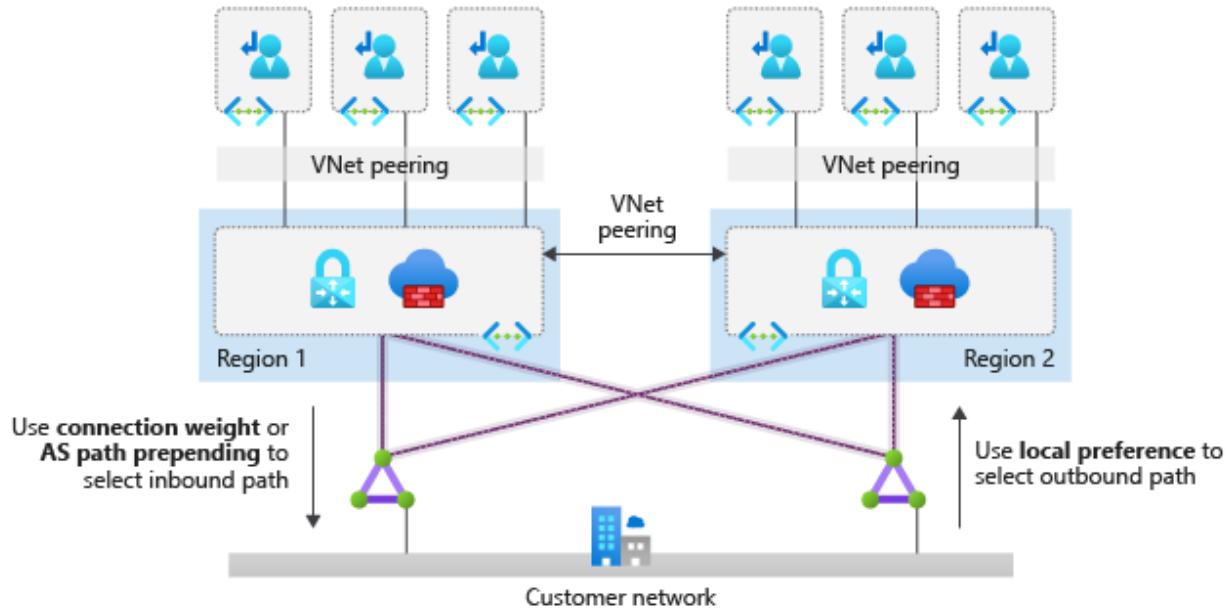
- If you need transitivity between ExpressRoute and VPN gateways in a hub-and-spoke scenario, use Azure Route Server as described in [this reference scenario](#).



- When you have hub-and-spoke networks in multiple Azure regions and a few landing zones need to connect across regions, use global virtual network peering to directly connect landing zone virtual networks that need to route traffic to each other. Depending on the communicating VM's SKU, global virtual network peering can provide high network throughput. Traffic between directly peered landing zone virtual networks bypasses NVAs within hub virtual networks. [Limitations on global virtual network peering](#) apply to the traffic.
- When you have hub-and-spoke networks in multiple Azure regions, and most landing zones need to connect across regions (or when using direct peering to bypass hub NVAs isn't compatible with your security requirements), use hub NVAs to connect hub virtual networks in each region to each other and to route traffic across regions. Global virtual network peering or ExpressRoute circuits can help to connect hub virtual networks in the following ways:

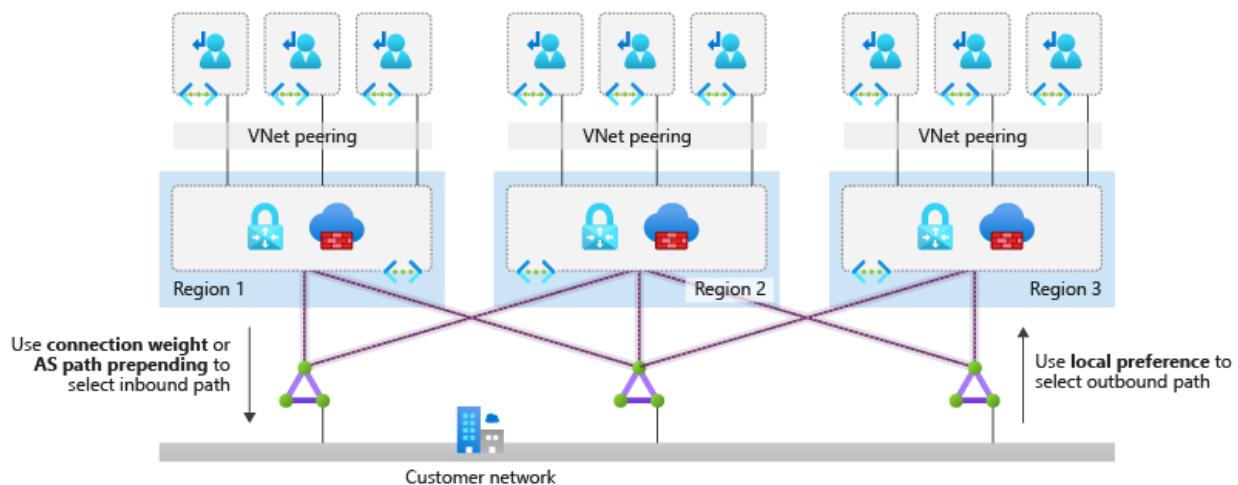
- Global virtual network peering provides a low latency and high throughput connection but generates [traffic fees](#).
- Routing via ExpressRoute might lead to increased latency (due to the MSEE hairpin), and the selected [ExpressRoute gateway SKU](#) limits the throughput.

The following figure shows both options:



*Figure 4: Options for hub-to-hub connectivity.*

- When two Azure regions need to connect, evaluate, and use global virtual network peering or the same ExpressRoute circuits to connect both hub virtual networks.
- When more than two Azure regions need to connect, then we recommend that the hub virtual networks in each region connect to the same ExpressRoute circuits. Global virtual network peering would require managing a large number of peering relationships and a complex set of user-defined routes (UDRs) across multiple virtual networks. The following diagram shows how to connect hub-and-spoke networks in three regions:



*Figure 5: ExpressRoute providing hub-to-hub connectivity between multiple regions.*

- When you use ExpressRoute circuits for cross-region connectivity, spokes in different regions communicate directly and bypass the firewall because they learn via BGP routes to the spokes of the remote hub. If you need the firewall NVAs in the hub virtual networks to inspect traffic across spokes, you must implement one of these options:
  - Create more specific route entries in the spoke UDRs for the firewall in the local hub virtual network to redirect traffic across hubs.
  - To simplify route configuration, [disable BGP propagation](#) on the spoke route tables.
- When your organization requires hub-and-spoke network architectures across more than two Azure regions and global transit connectivity between landing zones virtual networks across Azure regions, and you want to minimize network management overhead, we recommended a managed global transit network architecture that's based on [Virtual WAN](#).
- Deploy each region's hub network resources into separate resource groups, and sort them into each deployed region.
- Use [Azure Virtual Network Manager](#) to manage connectivity and security configuration of virtual networks globally across subscriptions.
- Use [Azure Monitor for Networks](#) to monitor the end-to-end state of your networks on Azure.
- You must consider the following two [limits](#) when you connect spoke virtual networks to the central hub virtual network:
  - The maximum number of virtual network peering connections per virtual network.
  - The maximum number of prefixes that ExpressRoute with private peering advertises from Azure to on-premises.

Ensure that the number of spoke virtual networks connected to the hub virtual network don't exceed these limits.

# Implement TIC 3.0 compliance

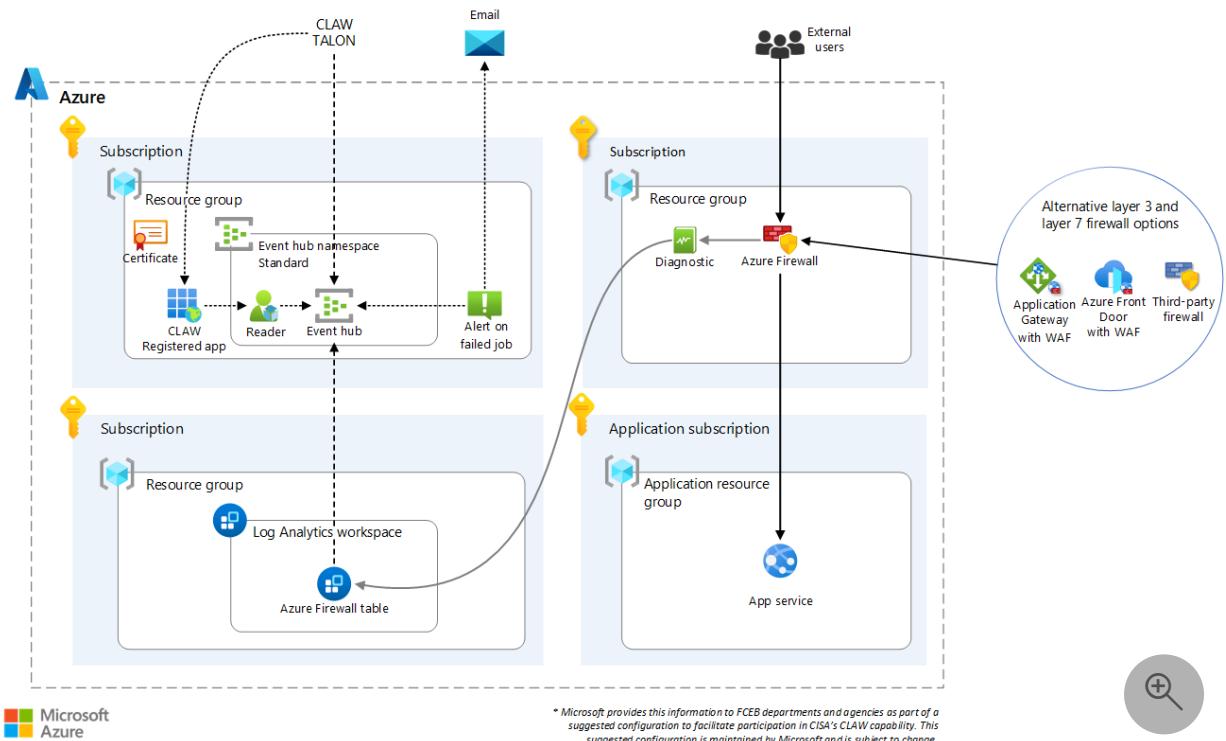
Azure Firewall   Azure Application Gateway   Azure Front Door   Azure Log Analytics   Azure Event Hubs

This article describes how to achieve Trusted Internet Connections (TIC) 3.0 compliance for internet-facing Azure applications and services. It provides solutions and resources to help government organizations meet TIC 3.0 compliance. It also describes how to deploy the required assets and how to incorporate the solutions into existing systems.

## ⓘ Note

Microsoft provides this information to Federal Civilian Executive Branch (FCEB) departments and agencies as part of a suggested configuration to facilitate participation in the Cybersecurity and Infrastructure Security Agency (CISA) Cloud Log Aggregation Warehouse (CLAW) capability. The suggested configurations are maintained by Microsoft and are subject to change.

## Architecture



Download a [Visio file](#) of this architecture.

## Dataflow

## 1. Firewall

- The firewall can be any layer 3 or layer 7 firewall.
  - Azure Firewall and some third-party firewalls, also known as Network Virtual Appliances (NVAs), are layer 3 firewalls.
  - Azure Application Gateway with Web Application Firewall (WAF) and Azure Front Door with WAF are layer 7 firewalls.
  - This article provides deployment solutions for Azure Firewall, Application Gateway with WAF, and Azure Front Door with WAF deployments.
- The firewall enforces policies, collects metrics, and logs connection transactions between web services and the users and services that access the web services.

## 2. Firewall logs

- Azure Firewall, Application Gateway with WAF, and Azure Front Door with WAF send logs to the Log Analytics workspace.
- Third-party firewalls send logs in Syslog format to the Log Analytics workspace via a Syslog forwarder virtual machine.

## 3. Log Analytics workspace

- The Log Analytics workspace is a repository for logs.
- It can host a service that provides custom analysis of the network traffic data from the firewall.

## 4. Service principal (registered application)

## 5. Azure Event Hubs Standard

## 6. CISA TALON

# Components

- Firewall. Your architecture will use one or more of the following firewalls. (For more information, see the [Alternatives](#) section of this article.)
  - [Azure Firewall](#) is a cloud-native, intelligent network firewall security service that provides enhanced threat protection for cloud workloads that run on Azure. It's a fully stateful firewall as a service with built-in high availability and unrestricted cloud scalability. It's available in two performance tiers: Standard and Premium. Azure Firewall Premium includes all the functionality of Azure Firewall Standard and provides additional features like Transport Layer Security (TLS) inspection and an intrusion detection and prevention system (IDPS).
  - [Application Gateway](#) with [WAF](#) is a regional web traffic load balancer that enables you to manage traffic to your web applications. WAF provides

enhanced centralized protection of your web applications from common exploits and vulnerabilities.

- [Azure Front Door](#) with [WAF](#) is a global web traffic load balancer that enables you to manage traffic to your web applications. It provides content delivery network (CDN) capabilities to speed up and modernize your applications. WAF provides enhanced centralized protection of your web applications from common exploits and vulnerabilities.
- A third-party firewall is an NVA that runs on an Azure virtual machine and uses firewall services from non-Microsoft vendors. Microsoft supports a large ecosystem of third-party vendors that provide firewall services.
- Logging and authentication.
  - Log Analytics is a tool that's available in the Azure portal that you can use to edit and run log queries against Azure Monitor Logs. For more information, see [Overview of Log Analytics in Azure Monitor](#).
  - [Azure Monitor](#) is a comprehensive solution for collecting, analyzing, and acting on telemetry.
  - [Microsoft Entra ID](#) provides identity services, single sign-on, and multifactor authentication across Azure workloads.
  - A [service principal](#) (registered application) is an entity that defines the access policy and permissions for a user or application in a Microsoft Entra tenant.
  - [Event Hubs Standard](#) is a modern big data streaming platform and event ingestion service.
  - CISA TALON is a CISA-operated service that runs on Azure. TALON connects to your Event Hubs service, authenticates by using a CISA-supplied certificate that's associated with your service principal, and collects logs for CLAW consumption.

## Alternatives

There are a few alternatives that you can use in these solutions:

- You can separate log collection into areas of responsibility. For example, you can send Microsoft Entra logs to a Log Analytics workspace that's managed by the identity team, and send network logs to a different Log Analytics workspace that's managed by the network team.
- The examples in this article each use a single firewall, but some organizational requirements or architectures require two or more. For example, an architecture can include an Azure Firewall instance and an Application Gateway instance with WAF. Logs for each firewall must be collected and made available for CISA TALON to collect.

- If your environment requires internet egress from Azure-based virtual machines, you can use a layer 3 solution like Azure Firewall or a third-party firewall to monitor and log the outbound traffic.

## Scenario details

TIC 3.0 moves TIC from on-premises data collection to a cloud-based approach that better supports modern applications and systems. It improves performance because you can directly access Azure applications. With TIC 2.x, you need to access Azure applications via a TIC 2.x Managed Trusted Internet Protocol Service (MTIPS) device, which slows down the response.

Routing application traffic through a firewall and logging that traffic is the core functionality demonstrated in the solutions presented here. The firewall can be Azure Firewall, Azure Front Door with WAF, Application Gateway with WAF, or a third-party NVA. The firewall helps to secure the cloud perimeter and saves logs of each transaction. Independently of the firewall layer, the log collection and delivery solution requires a Log Analytics workspace, a registered application, and an event hub. The Log Analytics workspace sends logs to the event hub.

CLAW is a CISA managed service. In late 2022, CISA released TALON. TALON is a CISA managed service that uses Azure native capabilities. An instance of TALON runs in each Azure region. TALON connects to event hubs that are managed by government agencies to pull agency firewall and Microsoft Entra logs into CISA CLAW.

For more information on CLAW, TIC 3.0, and MTIPS, see:

- [Trusted Internet Connections guidance](#)
- [TIC 3.0 core guidance documents ↗](#)
- [National Cybersecurity Protection System \(NCPS\) documents ↗](#)
- [EINSTEIN ↗](#)
- [Managed Trusted Internet Protocol Services \(MTIPS\) ↗](#)

## Potential use cases

TIC 3.0 compliance solutions are commonly used by federal organizations and government agencies for their Azure-based web applications and API services.

## Considerations

These considerations implement the pillars of the Azure Well-Architected Framework, which is a set of guiding tenets that you can use to improve the quality of a workload. For more information, see [Microsoft Azure Well-Architected Framework](#).

- Evaluate your current architecture to determine which of the solutions presented here provides the best approach to TIC 3.0 compliance.
- Contact your CISA representative to request access to CLAW.
- Use the **Deploy to Azure** buttons in this article to deploy one or more of the solutions in a test environment. Doing so should help you become familiar with the process and the deployed resources.
- See [TIC 3.0 compliance for internet-facing applications](#), a complementary article that provides more details and assets for TIC 3.0:
  - Additional information about achieving compliance.
  - ARM templates to simplify deployment.
  - Information to assist with integrating existing resources into the solution.
  - The types of logs collected for each service layer and Kusto queries for reviewing logs collected by CISA. You can use the queries for your organization's security requirements.

## Reliability

Reliability ensures your application can meet the commitments you make to your customers. For more information, see [Overview of the reliability pillar](#).

- Azure Firewall Standard and Premium integrate with availability zones to increase availability.
- Application Gateway v2 supports autoscaling and availability zones to increase reliability.
- Multi-region implementations that include load balancing services like Azure Front Door can improve reliability and resiliency.
- Event Hubs Standard and Premium provide Geo-disaster recovery pairing that enables a namespace to fail over to a secondary region.

## Security

Security provides assurances against deliberate attacks and the abuse of your valuable data and systems. For more information, see [Overview of the security pillar](#).

- When you register an enterprise application, a service principal is created. Use a naming scheme for service principals that indicates the purpose of each one.

- Perform audits to determine the activity of service principals and the status of service principal owners.
- Azure Firewall has standard policies. WAFs associated with Application Gateway and Azure Front Door have managed rule sets to help secure your web service. Start with these rule sets and build organizational policies over time based on industry requirements, best practices, and government regulations.
- Event Hubs access is authorized via Microsoft Entra managed identities and a certificate that's provided by CISA.

## Cost optimization

Cost optimization is about reducing unnecessary expenses and improving operational efficiencies. For more information, see [Overview of the cost optimization pillar](#).

The cost of each solution scales down as the resources increase. The pricing in this [Azure pricing calculator example scenario](#) is based on the Azure Firewall solution. If you change the configuration, costs might increase. With some plans, costs increase as the number of ingested logs increase.

 **Note**

Use the [Azure pricing calculator](#) to get up-to-date pricing that's based on the resources that are deployed for the selected solution.

## Operational excellence

Operational excellence covers the operations processes that deploy an application and keep it running in production. For more information, see [Overview of the operational excellence pillar](#).

- [Azure Monitor alerts](#) are built into the solutions to notify you when an upload fails to deliver logs to CLAW. You need to determine the severity of the alerts and how to respond.
- You can use ARM templates to speed up the deployment of TIC 3.0 architectures for new applications.

## Performance efficiency

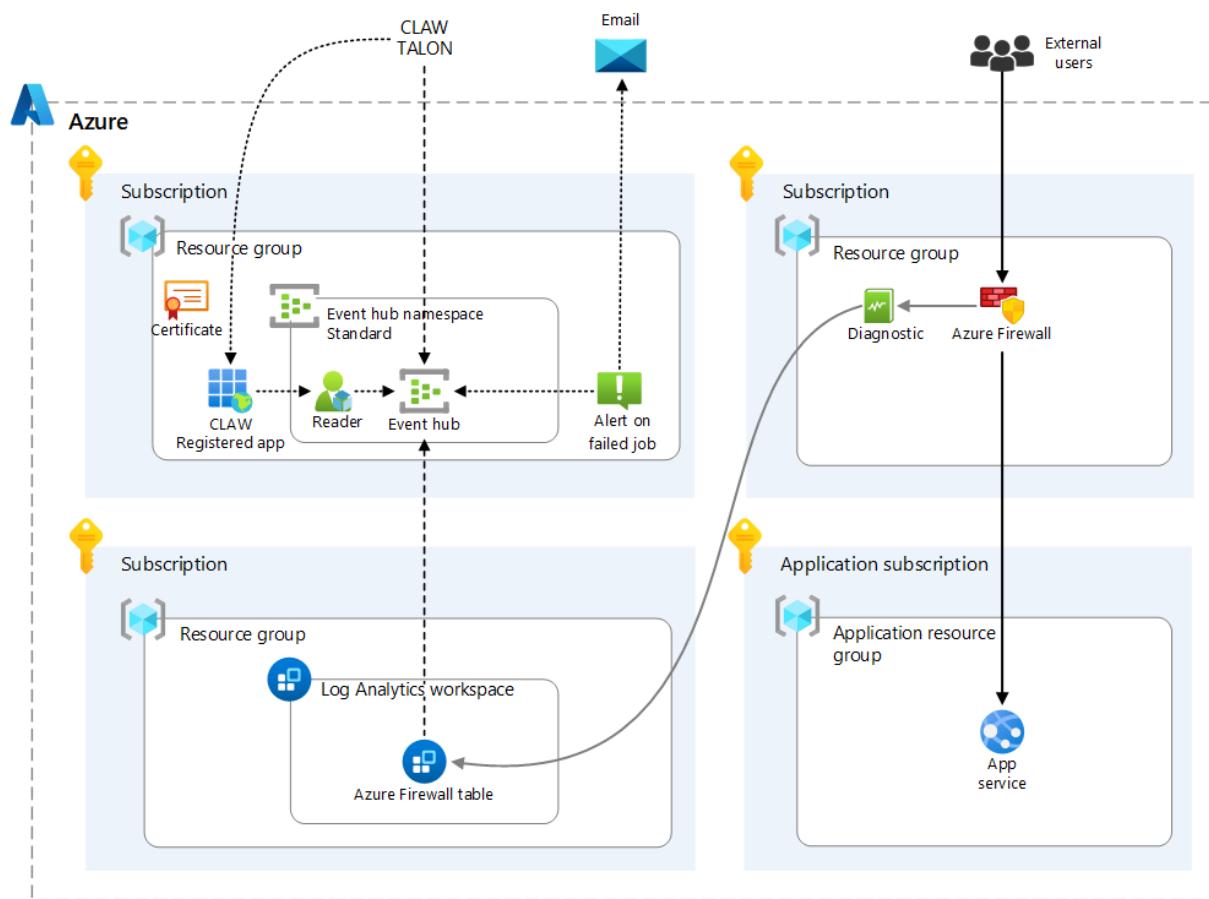
Performance efficiency is the ability of your workload to scale to meet the demands placed on it by users in an efficient manner. For more information, see [Performance](#)

efficiency pillar overview.

- Azure Firewall, Application Gateway, Azure Front Door, and Event Hubs performance scales as usage increases.
- Azure Firewall Premium allows more TCP connections than Standard and provides increased bandwidth.
- Application Gateway v2 automatically ensures that new instances are spread across fault domains and update domains.
- Azure Front Door provides caching, compression, traffic acceleration, and TLS termination to improve performance.
- Event Hubs Standard and Premium provide auto-inflate to scale up as load increases.

## Deploy an Azure Firewall solution

The following solution uses Azure Firewall to manage traffic entering your Azure application environment. The solution includes all resources for generating, collecting, and delivering logs to CLAW. It also includes an app service to track the types of telemetry collected by the firewall.



The solution includes:

- A virtual network that has separate subnets for the firewall and servers.
- A Log Analytics workspace.
- Azure Firewall with a network policy for internet access.
- Azure Firewall diagnostic settings that send logs to the Log Analytics workspace.
- A route table that's associated with the application resource group to route the app service to the firewall for the logs it generates.
- A registered application.
- An event hub.
- An alert rule that sends an email if a job fails.

For the sake of simplicity, all resources are deployed to a single subscription and virtual network. You can deploy the resources in any combination of resource groups or across multiple virtual networks.



Deploy to Azure



Deploy to Azure Government



## Post-deployment tasks

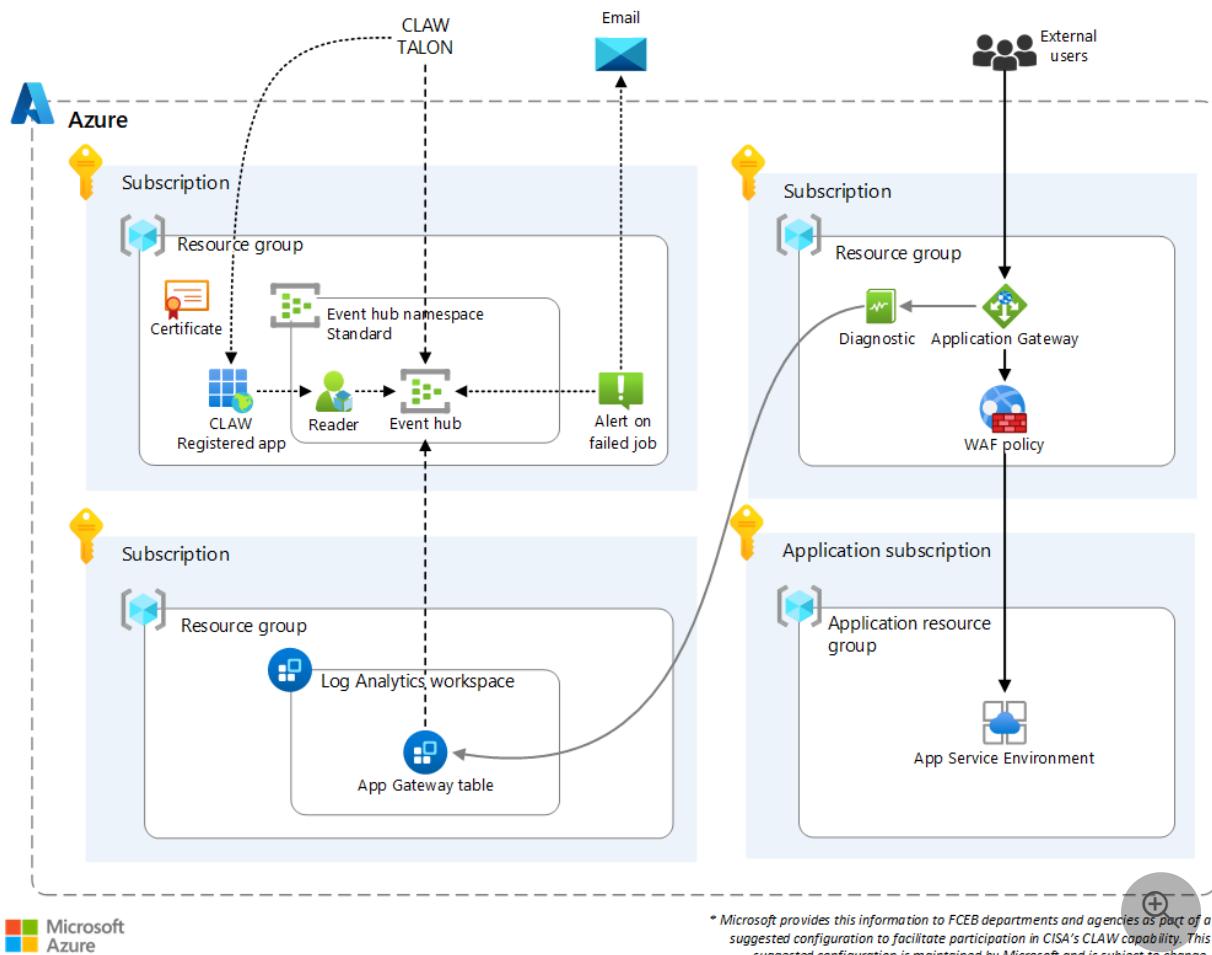
After deployment, your environment performs the firewall capabilities and logging connections. To meet compliance with TIC 3.0 policies for network telemetry collection, you need to ensure that the logs make it to CISA CLAW. The post-deployment steps finish the tasks to enable compliance. To complete these steps, you need to coordinate with CISA because CISA needs to supply a certificate to associate with your service principal. For step-by-step details, see [Post-deployment tasks](#).

You need to manually perform the following tasks after deployment. You can't complete them by using an ARM template.

- Obtain a public key certificate from CISA.
- Create a service principal (application registration).
- Add the public key certificate to the application registration.
- Assign the application the Azure Event Hubs Data Receiver role at the Event Hubs namespace scope.
- Activate the feed by sending the Azure tenant ID, application (client) ID, event hub namespace name, event hub name, and consumer group name to CISA.

# Deploy a solution that uses Application Gateway with WAF

The following solution uses Application Gateway with WAF to manage traffic entering your Azure application environment. The solution includes all resources for generating, collecting, and delivering logs to CLAW. It also includes an app service to track the types of telemetry collected by the firewall.



The solution includes:

- A virtual network that has separate subnets for the firewall and servers.
- A Log Analytics workspace.
- An Application Gateway v2 instance with WAF. The WAF is configured with bot and Microsoft managed policies.
- Application Gateway v2 diagnostic settings that send logs to the Log Analytics workspace.
- A registered application.
- An event hub.
- An alert rule that sends an email if a job fails.

For the sake of simplicity, all resources are deployed to a single subscription and virtual network. You can deploy the resources in any combination of resource groups or across multiple virtual networks.



Deploy to Azure



Deploy to Azure Government



## Post-deployment tasks

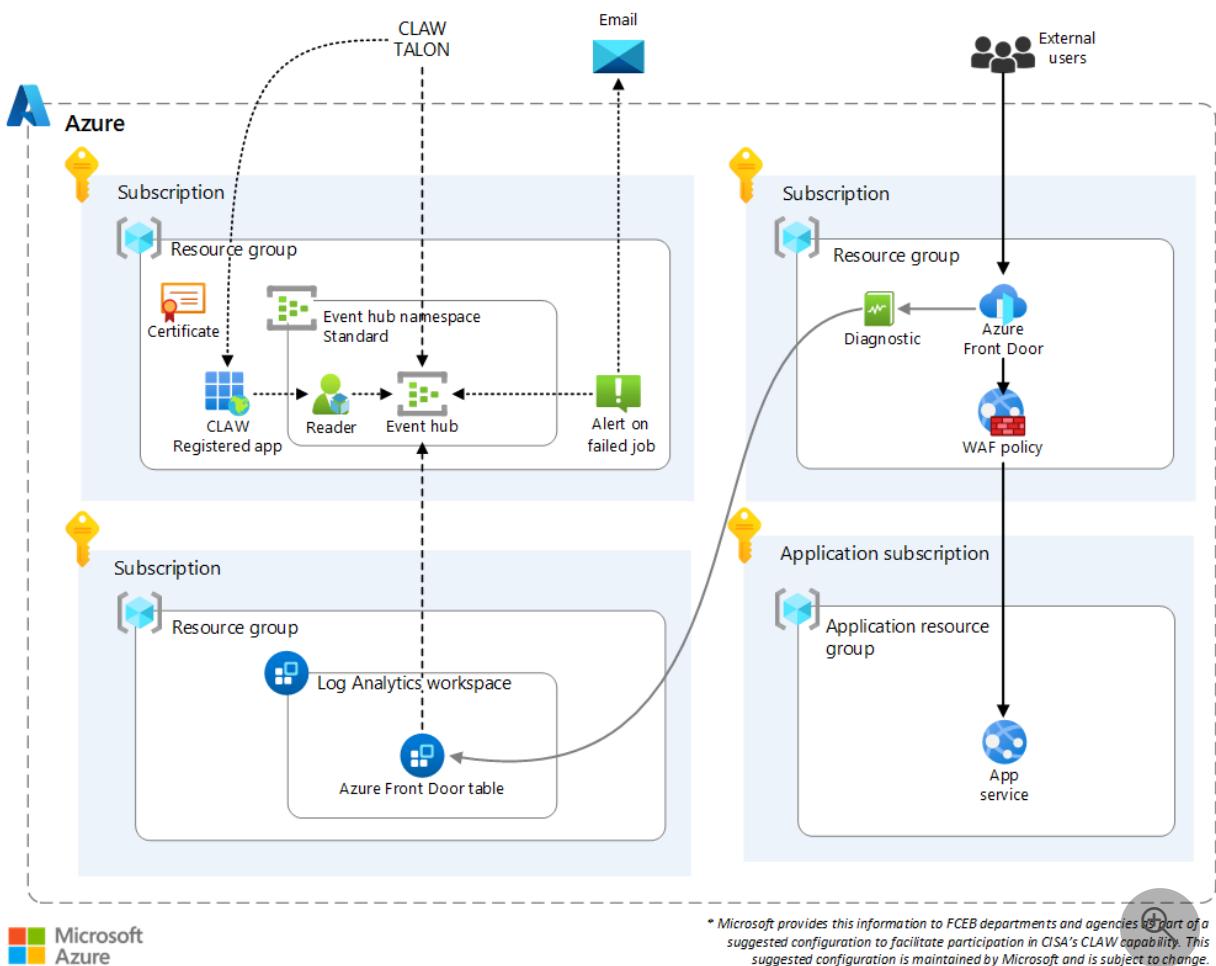
After deployment, your environment performs the firewall capabilities and logging connections. To meet compliance with TIC 3.0 policies for network telemetry collection, you need to ensure that the logs make it to CISA CLAW. The post-deployment steps finish the tasks to enable compliance. To complete these steps, you need to coordinate with CISA because CISA needs to supply a certificate to associate with your service principal. For step-by-step details, see [Post-deployment tasks](#).

You need to manually perform the following tasks after deployment. You can't complete them by using an ARM template.

- Obtain a public key certificate from CISA.
- Create a service principal (application registration).
- Add the public key certificate to the application registration.
- Assign the application the Azure Event Hubs Data Receiver role at the Event Hubs namespace scope.
- Activate the feed by sending the Azure tenant ID, application (client) ID, event hub namespace name, event hub name, and consumer group name to CISA.

## Deploy a solution that uses Azure Front Door with WAF

The following solution uses Azure Front Door with WAF to manage traffic entering your Azure application environment. The solution includes all resources for generating, collecting, and delivering logs to CLAW. It also includes an app service to track the types of telemetry collected by the firewall.



\* Microsoft provides this information to FCEB departments and agencies as part of a suggested configuration to facilitate participation in CISA's CLAW capability. This suggested configuration is maintained by Microsoft and is subject to change.

The solution includes:

- A virtual network that has separate subnets for the firewall and servers.
- A Log Analytics workspace.
- An Azure Front Door instance with WAF. The WAF is configured with bot and Microsoft managed policies.
- Azure Front Door diagnostic settings that send logs to the Log Analytics workspace.
- A registered application.
- An event hub.
- An alert rule that sends an email if a job fails.

For the sake of simplicity, all resources are deployed to a single subscription and virtual network. You can deploy the resources in any combination of resource groups or across multiple virtual networks.

[Deploy to Azure](#) [Deploy to Azure Government](#)

## Post-deployment tasks

After deployment, your environment performs the firewall capabilities and logging connections. To meet compliance with TIC 3.0 policies for network telemetry collection, you need to ensure that the logs make it to CISA CLAW. The post-deployment steps finish the tasks to enable compliance. To complete these steps, you need to coordinate with CISA because CISA needs to supply a certificate to associate with your service principal. For step-by-step details, see [Post-deployment tasks](#).

You need to manually perform the following tasks after deployment. You can't complete them by using an ARM template.

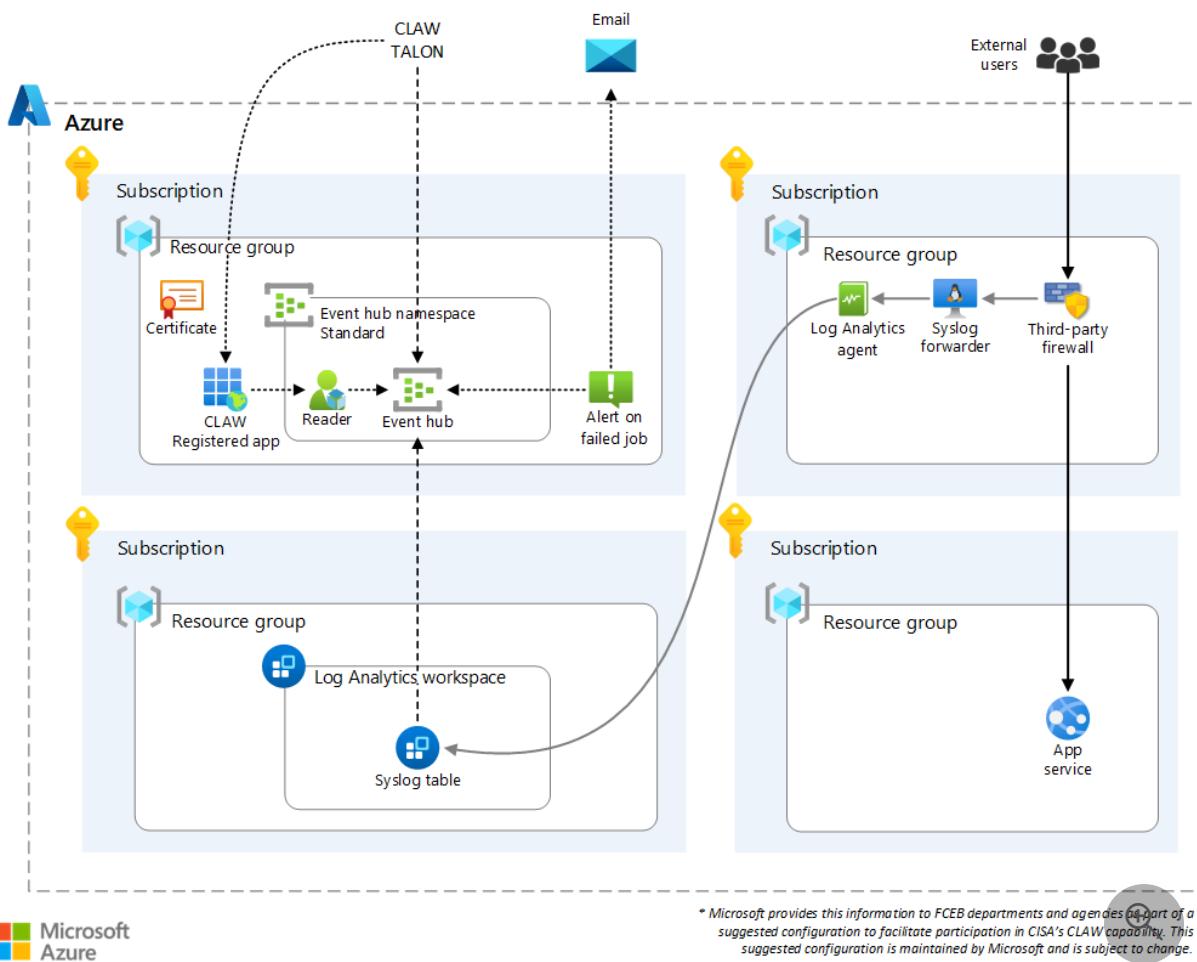
- Obtain a public key certificate from CISA.
- Create a service principal (application registration).
- Add the public key certificate to the application registration.
- Assign the application the Azure Event Hubs Data Receiver role at the Event Hubs namespace scope.
- Activate the feed by sending the Azure tenant ID, application (client) ID, event hub namespace name, event hub name, and consumer group name to CISA.

## Third-party firewall (NVA) solution

### Note

Deployment resources aren't provided for this solution. It's included only to provide guidance.

The following solution illustrates how you can use a third-party firewall to manage traffic entering your Azure application environment and implement TIC 3.0 compliance. Third-party firewalls require the use of a Syslog forwarder virtual machine. Its agents need to be registered with the Log Analytics workspace. The third-party firewall is configured to export its logs in Syslog format to the Syslog forwarder virtual machine. The agent is configured to send its logs to the Log Analytics workspace. After the logs are in the Log Analytics workspace, they're sent to Event Hubs and processed as they are in the other solutions described in this article.



\* Microsoft provides this information to FCEB departments and agencies as part of a suggested configuration to facilitate participation in CISA's CLAW capability. This suggested configuration is maintained by Microsoft and is subject to change.

## Post-deployment tasks

After deployment, your environment performs the firewall capabilities and logging connections. To meet compliance with TIC 3.0 policies for network telemetry collection, you need to ensure that the logs make it to CISA CLAW. The post-deployment steps finish the tasks to enable compliance. To complete these steps, you need to coordinate with CISA because CISA needs to supply a certificate to associate with your service principal. For step-by-step details, see [Post-deployment tasks](#).

You need to manually perform the following tasks after deployment. You can't complete them by using an ARM template.

- Obtain a public key certificate from CISA.
- Create a service principal (application registration).
- Add the public key certificate to the application registration.
- Assign the application the Azure Event Hubs Data Receiver role at the Event Hubs namespace scope.
- Activate the feed by sending the Azure tenant ID, application (client) ID, event hub namespace name, event hub name, and consumer group name to CISA.

# Contributors

This article is maintained by Microsoft. It was originally written by the following contributors.

Principal author:

- [Paul Lizer](#) | Senior Cloud Solution Architect

Other contributor:

- [Mick Alberts](#) | Technical Writer

To see non-public LinkedIn profiles, sign in to LinkedIn.

## Next steps

- [TIC 3.0 core guidance documents](#)
- [National Cybersecurity Protection System \(NCPS\) documents](#)
- [EINSTEIN](#)
- [Managed Trusted Internet Protocol Services \(MTIPS\)](#)
- [Azure Trusted Internet Connection - Extended](#)
- [Federal App Innovation - TIC 3.0](#)
- [What is Azure Firewall?](#)
- [Azure Firewall documentation](#)
- [What is Azure Application Gateway?](#)
- [Azure Front Door](#)
- [Introduction to Azure WAF](#)
- [Overview of Log Analytics in Azure Monitor](#)
- [What is Azure Event Hubs?](#)
- [Overview of alerts in Azure](#)
- [Application and service principal objects in Microsoft Entra ID](#)
- [Use the portal to create a Microsoft Entra application and service principal that can access resources](#)
- [Register an application with the Microsoft identity platform](#)
- [Assign Azure roles using the Azure portal](#)
- [Create resource groups](#)
- [How to find your Microsoft Entra tenant ID](#)
- [Collect Syslog data sources with the Log Analytics agent](#)
- [Parse text data in Azure Monitor logs](#)
- [Introduction to flow logging for network security groups](#)
- [What are managed identities for Azure resources?](#)

- Deploy and configure Azure Firewall using the Azure portal

## Related resources

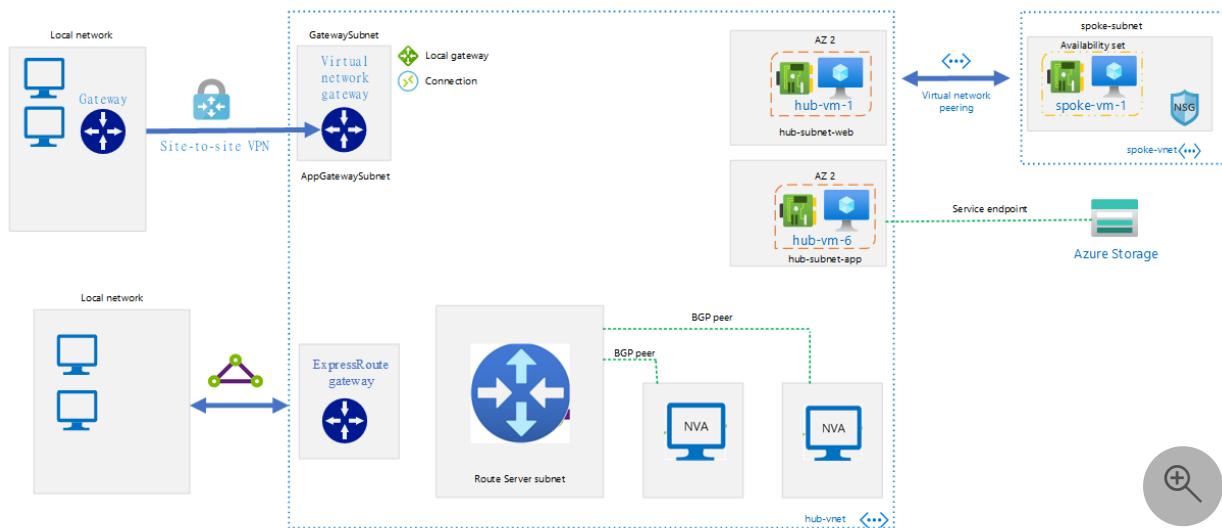
- [Implement a secure hybrid network](#)
- [Securely managed web applications](#)
- [Secure and govern workloads with network-level segmentation](#)
- [Improved-security access to multitenant web apps from an on-premises network](#)

# Update route tables by using Azure Route Server

Azure ExpressRoute   Azure Storage   Azure Virtual Network   Azure VPN Gateway

This article presents a solution for managing the dynamic routing between NVAs and virtual networks. At the core of the solution is Azure Route Server. This service simplifies the configuration, maintenance, and deployment of NVAs in your virtual network. When you use Route Server, you no longer need to manually update NVA route tables when your virtual network addresses change.

## Architecture



Download a [Visio file](#) of this architecture.

## Workflow

- This hub-and-spoke architecture has a hub virtual network and one spoke virtual network. The hub virtual network has multiple subnets, each containing virtual machines (VMs).
- The address space of each virtual network defines address ranges. For each of those ranges, Azure creates a route with the address prefix of that range. Azure adds those routes to route tables. Each virtual network has multiple subnets, and each subnet has a network interface card (NIC) that controls connectivity. Azure injects each virtual network's route table into the subnets' NICs.

You can't create or remove these default system routes. But you can:

- Override some system routes with [custom routes](#).
- Configure Azure to add [optional default routes](#) to specific subnets.
- Local networks use Azure VPN Gateway and an ExpressRoute gateway to connect to the hub virtual network in a coexisting configuration. When you add the VPN gateway, routes with the gateway as the next route get added to the route tables. When you add ExpressRoute, the route tables are also updated. These routes propagate to all subnets.
- The border gateway protocol (BGP) makes the exchange of IP addresses between on-premises and Azure components possible. This protocol directs packets between autonomous systems. Such systems are small networks or huge pools of routers that a single organization runs.
- A virtual network peering exists between the hub virtual network and the spoke virtual network. When you create the peering, Azure updates the route table. Specifically, Azure adds a route for each address range that's in the hub address space or the spoke address space. These routes propagate to all subnets.
- A subnet in the hub virtual network uses a service endpoint for Azure Storage. Azure adds a public IP address for Storage to that subnet's route table.
- The hub virtual network contains two NVAs. The NVAs might be gateways, software-defined wide-area networks (SD-WANs), or security appliance firewalls. Route Server exchanges the NVA, network application, and gateway routes by:
  - Creating an instance of Azure Virtual Machine Scale Sets. Each VM in the scale set has an IP address. As with gateway IP addresses, Route Server has access to the VM IP addresses.
  - Establishing BGP peers between each NVA and a VM in the scale set.
  - Injecting the VM IP addresses into all route tables in the virtual network and connected networks.

There's no need to:

- Manually add user-defined routes.
- Manually create route tables.
- Link route tables to the subnet to propagate the routes.
- Update route tables when IP addresses change.

## Components

- [Route Server](#) simplifies dynamic routing between NVAs that support BGP and virtual networks. This service eliminates the administrative overhead of maintaining

route tables.

- [Virtual Network](#) is the fundamental building block for private networks in Azure. Azure resources like VMs can securely communicate with each other, the internet, and on-premises networks through Virtual Network.
- [Virtual network peering](#) connects two or more Azure virtual networks. Peerings provide low-latency, high-bandwidth connections between resources in different virtual networks. Traffic between VMs in peered virtual networks uses only the Microsoft private network.
- [VPN Gateway](#) is a specific type of virtual network gateway. You can use VPN Gateway to send encrypted traffic:
  - Between an Azure virtual network and an on-premises location over the public internet.
  - Between Azure virtual networks over the Azure backbone network.
- [ExpressRoute](#) extends on-premises networks into the Microsoft cloud. By using a connectivity provider, ExpressRoute establishes private connections to cloud components like Azure services and Microsoft 365.
- A [service endpoint](#) provides secure and direct connectivity to an Azure service from private IP addresses in a virtual network. The service endpoint provides the identity of the virtual network to the Azure service. So the virtual network resources don't need public IP addresses to access the service, and the endpoint protects the service by allowing only traffic from the specified virtual network. The connections use optimized routes over the Azure backbone network.
- An NVA is a virtual appliance that offers networking capabilities such as firewall security and load balancing.
- [Azure Storage](#) is a cloud storage solution that includes object, file, disk, queue, and table storage. Services include hybrid storage solutions and tools for transferring, sharing, and backing up data.

## Alternatives

- In this solution, you don't have to connect the service endpoint to Storage. You can use other Azure services instead. For a list of services that you can secure with service endpoints, see [Virtual Network service endpoints](#).
- Instead of using Route Server, you can add user-defined routes to each subnet's route table. For more information about user-defined routes, see [User-defined in](#)

## Scenario details

Network routing is the process of determining the path that traffic takes across networks to reach a destination. Route tables list network topology information that's useful for determining routing paths.

When your virtual network contains a network virtual appliance (NVA), you need to manually configure and update your route tables.

This article presents a solution for managing the dynamic routing between NVAs and virtual networks. At the core of the solution is Azure Route Server. This service simplifies the configuration, maintenance, and deployment of NVAs in your virtual network. When you use Route Server, you no longer need to manually update NVA route tables when your virtual network addresses change.

## Potential use cases

This solution applies to scenarios that:

- Use dual-homed networks. Besides typical hub-and-spoke network topologies, Router Server also supports dual-homed network topologies. This type of configuration peers a spoke virtual network with two or more hub virtual networks. For detailed information, see [About dual-homed network with Azure Route Server](#).
- Connect NVAs to Azure ExpressRoute. Some virtual networks contain Route Server, an ExpressRoute gateway, and an NVA. By default, Route Server doesn't propagate the NVA routes to ExpressRoute. Route Server also doesn't propagate ExpressRoute routes to the NVA. You can get ExpressRoute and the NVA to exchange routes by turning on route exchange functionality in Route Server. For detailed information, see [About Azure Route Server support for ExpressRoute and Azure VPN](#).
- Use Azure to connect to the internet from an on-premises system. Organizations that lack good internet access might use this configuration. Systems that have already migrated internet proxies to Azure are other possibilities. Route Server makes this setup possible.

## Considerations

Consider these points when implementing this solution:

- Route Server establishes connections and exchanges routes. It doesn't transfer data packets. As a result, the VMs that Route Server runs in its back end don't require significant CPU power or computational power.
- When you deploy Route Server, create a subnet called `RouteServerSubnet` that uses an IPv4 subnet mask of `/27`. Place Route Server in that subnet.
- In Azure gateways, the Basic pricing tier doesn't support coexisting ExpressRoute and VPN Gateway connections. For other limitations with coexisting configurations, see [Limits and limitations](#).
- There's no limit on the number of service endpoints that you can use in a virtual network. But some Azure services, such as Storage, enforce limits on the number of subnets that you can use to secure the resource. For more information, see [Next steps in Virtual Network service endpoints](#).

When considering this solution, also keep in mind the points in the following sections.

## Availability

Route Server is a fully managed service that offers high availability. For this service's availability guarantee, see [SLA for Azure Route Server](#).

## Scalability

Most components in this solution are managed services that automatically scale. But there are a few exceptions:

- Route Server can advertise at most 200 routes to ExpressRoute or a VPN gateway.
- Route Server can support at most 2,000 VMs per virtual network, including peered virtual networks.

## Security

- For guidance on improving the security of your applications and data on Azure, see [Overview of the Azure Security Benchmark \(v1\)](#).
- For guidance from Azure Security Benchmark version 1.0 that's specific to Virtual Network, see [Azure security baseline for Virtual Network](#).

## Resiliency

This solution uses only managed components. At a regional level, all these components are automatically resilient. Route Server offers high availability. When you deploy Route Server in an Azure region that supports availability zones, your implementation has zone-level redundancy. For more information about availability zones, see [Regions and availability zones](#).

## Cost optimization

To estimate the cost of implementing this solution, see the [Azure pricing calculator](#). For general information about reducing unnecessary expenses, see [Overview of the cost optimization pillar](#).

The following sections discuss pricing information for the solution's components.

### Route Server

Currently, there's no upfront cost or termination fee for Route Server. For pricing information, see [Azure Route Server pricing](#).

### Virtual Network

You can use Virtual Network free of charge. With an Azure subscription, you can create up to 50 virtual networks across all regions. Traffic that's within a virtual network's boundaries is free. As a result, there's no charge for communication between two VMs in the same virtual network.

### VPN Gateway

When you use VPN Gateway, all inbound traffic is free. You're charged only for outbound traffic. Internet bandwidth costs apply with VPN outbound traffic. For more information, see [VPN Gateway pricing](#).

### ExpressRoute

ExpressRoute data transfers that are inbound are free of charge. For outbound data transfer, you're charged a predetermined rate. A fixed monthly port fee also applies. For more information, see [Azure ExpressRoute pricing](#).

### Service endpoints

There's no charge for using service endpoints.

## NVAs

NVAs are charged based on the appliance that you use. You're also charged for the Azure VMs that you deploy and the underlying infrastructure resources that you consume, such as storage and networking. For more information, see [Linux Virtual Machines Pricing ↗](#).

## Next steps

- Quickstart: Create and configure Route Server using the Azure portal
- About Azure Route Server support for ExpressRoute and Azure VPN
- Azure Route Server FAQ
- Azure road map ↗
- Networking blog ↗
- SLA for Azure Route Server ↗
- What is Azure Route Server?

## Related resources

- [Azure Firewall architecture overview](#)
- [Choose between virtual network peering and VPN gateways](#)
- [Build solutions for high availability using availability zones](#)
- [Deploy highly available NVAs](#)
- [Zero-trust network for web applications with Azure Firewall and Application Gateway](#)

# Hub-spoke network topology with Azure Virtual WAN

Azure Virtual WAN

This hub-spoke architecture provides an alternate solution to the reference architectures [hub-spoke network topology in Azure](#) and [implement a secure hybrid network](#).

The *hub* is a virtual network in Azure that acts as a central point of connectivity to your on-premises network. The *spokes* are virtual networks that peer with the hub and can be used to isolate workloads. Traffic flows between the on-premises data center(s) and the hub through an ExpressRoute or VPN gateway connection. The main differentiator of this approach is the use of [Azure Virtual WAN](#) (VWAN) to replace hubs as a managed service.

This architecture includes the benefits of standard hub-spoke network topology and introduces new benefits:

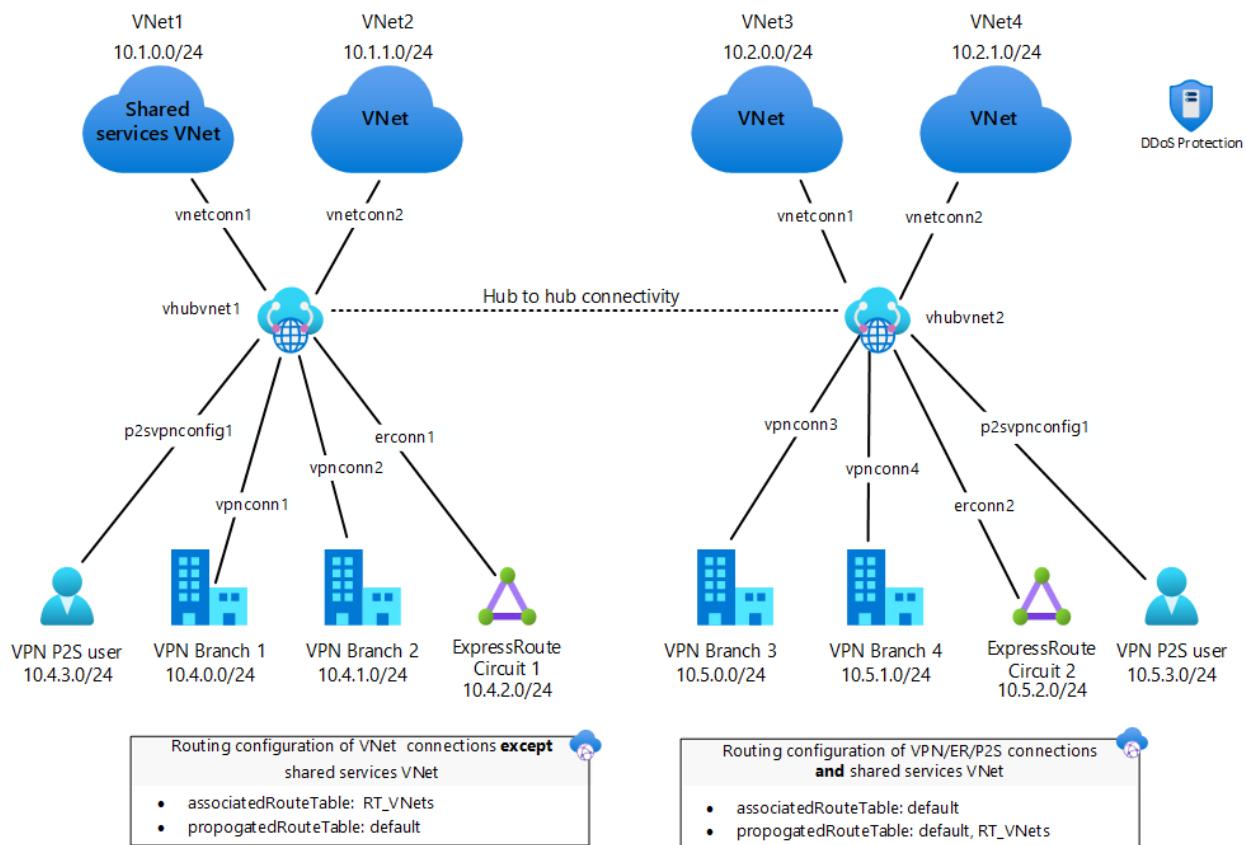
- **Less operational overhead** by replacing existing hubs with a fully managed VWAN service.
- **Cost savings** by using a managed service and removing the necessity of network virtual appliance.
- **Improved security** by introducing centrally managed secured Hubs with Azure Firewall and VWAN to minimize security risks related to misconfiguration.
- **Separation of concerns** between central IT (SecOps, InfraOps) and workloads (DevOps).

## Potential use cases

Typical uses for this architecture include cases in which:

- Connectivity among workloads requires central control and access to shared services.
- An enterprise requires central control over security aspects, such as a firewall, and requires segregated management for the workloads in each spoke.

# Architecture



Download a [Visio file](#) of this architecture.

The architecture consists of:

- **On-premises network.** A private local area network (LAN) running within an organization.
- **VPN device.** A device or service that provides external connectivity to the on-premises network.
- **VPN virtual network gateway or ExpressRoute gateway.** The virtual network gateway enables the virtual network to connect to the VPN device, or ExpressRoute circuit, used for connectivity with your on-premises network.
- **Virtual WAN hub.** The [Virtual WAN](#) is used as the hub in the hub-spoke topology. The hub is the central point of connectivity to your on-premises network, and a place to host services that can be consumed by the different workloads hosted in the spoke virtual networks.
- **Secured virtual hub.** A Virtual WAN hub with associated security and routing policies configured by Azure Firewall Manager. A secured virtual hub comes with a

built-in routing so there is no need to configure user-defined routes.

- **Gateway subnet.** The virtual network gateways are held in the same subnet.
- **Spoke virtual networks.** One or more virtual networks that are used as spokes in the hub-spoke topology. Spokes can be used to isolate workloads in their own virtual networks and are managed separately from other spokes. Each workload might include multiple tiers, with multiple subnets connected through Azure load balancers.
- **Virtual network peering.** Two virtual networks can be connected using a VNet peering connection. Peering connections are nontransitive, low-latency connections between virtual networks. Once peered, virtual networks exchange traffic by using the Azure backbone, without the need for a router. In a hub-spoke network topology, you use virtual network peering to connect the hub to each spoke. Azure Virtual WAN enables transitivity among hubs, which is not possible solely using peering.

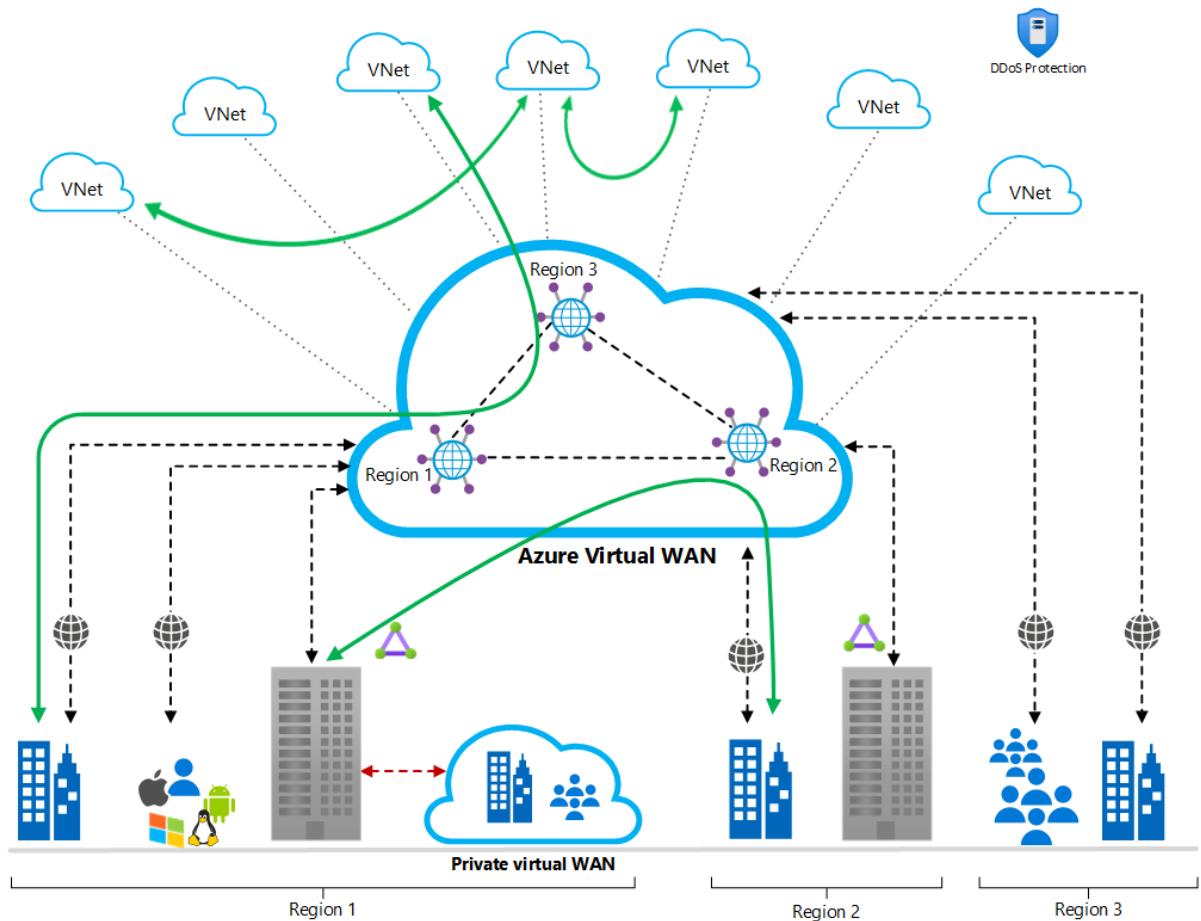
## Components

- [Azure Virtual Network ↗](#)
- [Azure Virtual WAN ↗](#)
- [Azure VPN Gateway ↗](#)
- [Azure ExpressRoute ↗](#)
- [Azure Firewall ↗](#)

## Alternatives

A hub-spoke architecture can be achieved two ways: a customer-managed hub infrastructure, or a Microsoft-managed hub infrastructure. In either case, spokes are connected to the hub using virtual network peering.

## Advantages



[Download a Visio file](#) of this architecture.

This diagram illustrates a few of the advantages that this architecture can provide:

- A full meshed hub among Azure Virtual Networks
- Branch to Azure connectivity
- Branch to Branch connectivity
- Mixed use of VPN and Express Route
- Mixed use of user VPN to the site
- VNET to VNET connectivity

## Recommendations

The following recommendations apply to most scenarios. Follow them, unless you have a specific requirement that overrides them.

## Resource groups

The hub and each spoke can be implemented in different resource groups, and, even better, in different subscriptions. When you peer virtual networks in different

subscriptions, both subscriptions can be associated to the same or a different Microsoft Entra tenant. This allows for a decentralized management of each workload, while sharing services maintained in the hub.

## Virtual WAN

Create a Standard Virtual WAN if you have a requirement for any of the following:

- Scaling for higher throughputs
- Private Connectivity (requires Premium Circuit in Global Reach location)
- ExpressRoute VPN Interconnect
- Integrated monitoring with [Azure Monitor](#) (Metrics and Resource Health)

Standard Virtual WANs are by default connected in a full mesh. Standard Virtual WAN supports any-to-any connectivity (Site-to-Site VPN, VNet, ExpressRoute, Point-to-site endpoints) in a single hub as well as across hubs. Basic virtual WAN supports only Site-to-Site VPN connectivity, branch-to-branch connectivity, and branch-to-VNet connectivity in a **single hub**.

## Virtual WAN Hub

A virtual hub is a Microsoft-managed virtual network. The hub contains various service endpoints to enable connectivity. The hub is the core of your network in a region. There can be multiple hubs per Azure region. For more information, see [Virtual WAN FAQ](#).

When you create a hub using the Azure portal, it creates a virtual hub VNet and a virtual hub VPN gateway. A Virtual WAN Hub requires an address range minimum of /24. This IP address space will be used for reserving a subnet for gateway and other components.

## Secured virtual hub

A virtual hub can be created as a secured virtual hub or converted to a secure one anytime after creation. For additional information, see [Secure your virtual hub using Azure Firewall Manager](#).

## GatewaySubnet

For more information about setting up the gateway, see the following reference architectures, depending on your connection type:

- Hybrid network using ExpressRoute
- Hybrid network using a VPN Gateway

For greater availability, you can use ExpressRoute plus a VPN for failover. See [Connect an on-premises network to Azure using ExpressRoute with VPN failover](#).

A hub-spoke topology cannot be used without a gateway, even if you don't need connectivity with your on-premises network.

## Virtual network peering

Virtual network peering is a nontransitive relationship between two virtual networks. However, Azure Virtual WAN allows spokes to connect with each other without having a dedicated peering among them.

However, if you have several spokes that need to connect with each other, you will run out of possible peering connections very quickly due to the limitation on the number of virtual network peerings per virtual network. (For more information, see [Networking limits](#).) In this scenario, Azure VWAN will solve this problem with its out-of-box functionality. For additional information, see [Global transit network architecture and Virtual WAN](#).

You can also configure spokes to use the hub gateway to communicate with remote networks. To allow gateway traffic to flow from spoke to hub, and connect to remote networks, you must:

- Configure the peering connection in the hub to **allow gateway transit**.
- Configure the peering connection in each spoke to **use remote gateways**.
- Configure all peering connections to **allow forwarded traffic**.

For additional information, see [Choose between virtual network peering and VPN gateways](#).

## Hub extensions

To support network-wide shared services like DNS resources, custom NVAs, Azure Bastion, and others, implement each service following the [virtual hub extension pattern](#). Following this model, you can build and operate single-responsibility extensions to individually expose these business-critical, shared services that you're otherwise unable to deploy directly in a virtual hub.

# Considerations

## Operations

Azure VWAN is a managed service provided by Microsoft. From a technology standpoint, it is not completely different from a customer-managed hub infrastructure. Azure Virtual WAN simplifies overall network architecture by offering a mesh network topology with transitive network connectivity among spokes. Monitoring of Azure VWAN can be achieved using Azure Monitor. Site-to-site configuration and connectivity between on-premises networks and Azure can be fully automated.

## Reliability

Azure Virtual WAN handles routing, which helps to optimize network latency among spokes as well as assure predictability of latency. Azure Virtual WAN also provides reliable connectivity among different Azure regions for the workloads spanning multiple regions. With this setup, end-to-end flow within Azure becomes more visible.

## Performance

With the help of Azure Virtual WAN, lower latency among spokes and across regions can be achieved. Azure Virtual WAN enables you to scale up to 20Gbps aggregate throughput.

## Scalability

Azure Virtual WAN provides a full mesh connectivity among spokes by preserving the ability to restrict traffic based on needs. With this architecture it is possible to have large-scale site-to-site performance. Moreover, you can create a global transit network architecture by enabling any-to-any connectivity between globally distributed sets of cloud workloads.

## Security

Hubs in Azure VWAN can be converted to secure HUBs by leveraging Azure Firewall. User-defined routes (UDRs) can still be leveraged in the same way to achieve network isolation. Azure VWAN enables encryption of traffic between the on-premises networks and Azure virtual networks over ExpressRoute.

Azure DDoS Protection, combined with application-design best practices, provides enhanced DDoS mitigation features to provide more defense against DDoS attacks. You should enable [Azure DDOS Protection](#) on any perimeter virtual network.

## Spoke connectivity and shared services

Connectivity among spokes is already achieved using Azure Virtual WAN. However, using UDRs in the spoke traffic is useful to isolate virtual networks. Any shared service can also be hosted on the same Virtual WAN as a spoke.

## Virtual network peering - Hub connection

Virtual network peering is a nontransitive relationship between two virtual networks. While using Azure Virtual WAN, virtual network peering is managed by Microsoft. Each connection added to a hub will also configure virtual network peering. With the help of Virtual WAN, all spokes will have a transitive relationship.

## Cost optimization

A customer-managed hub infrastructure introduces management cost to underlying Azure resources. To achieve a transitive connectivity with a predictable latency, you must have a Network Virtual Appliance (NVA) or Azure Firewall deployed in each hub. Using Azure Firewall with either choice will lower the cost compared to an NVA. Azure Firewall costs are the same for both options. There is an extra cost for Azure Virtual WAN; however, it is much less costly than managing your own hub infrastructure.

For more information, see [Virtual WAN pricing](#).

## Contributors

*This article is maintained by Microsoft. It was originally written by the following contributors.*

Principal author:

- [Yunus Emre Alpozen](#) | Program Architect Cross-Workload

*To see non-public LinkedIn profiles, sign in to LinkedIn.*

## Next steps

Learn more:

- [Hub-spoke network topology in Azure](#)
- [Design a hybrid Domain Name System solution with Azure](#)
- [Implement a secure hybrid network](#)
- [What is Azure ExpressRoute?](#)
- [Connect an on-premises network to Azure using ExpressRoute](#)
- [Firewall and Application Gateway for virtual networks](#)
- [Extend an on-premises network using VPN](#)
- [Secure and govern workloads with network level segmentation](#)

## Related resources

- [Strengthen your security posture with Azure ↗](#)
- [Virtual Network ↗](#)
- [Azure ExpressRoute ↗](#)
- [VPN Gateway ↗](#)
- [Azure Firewall ↗](#)

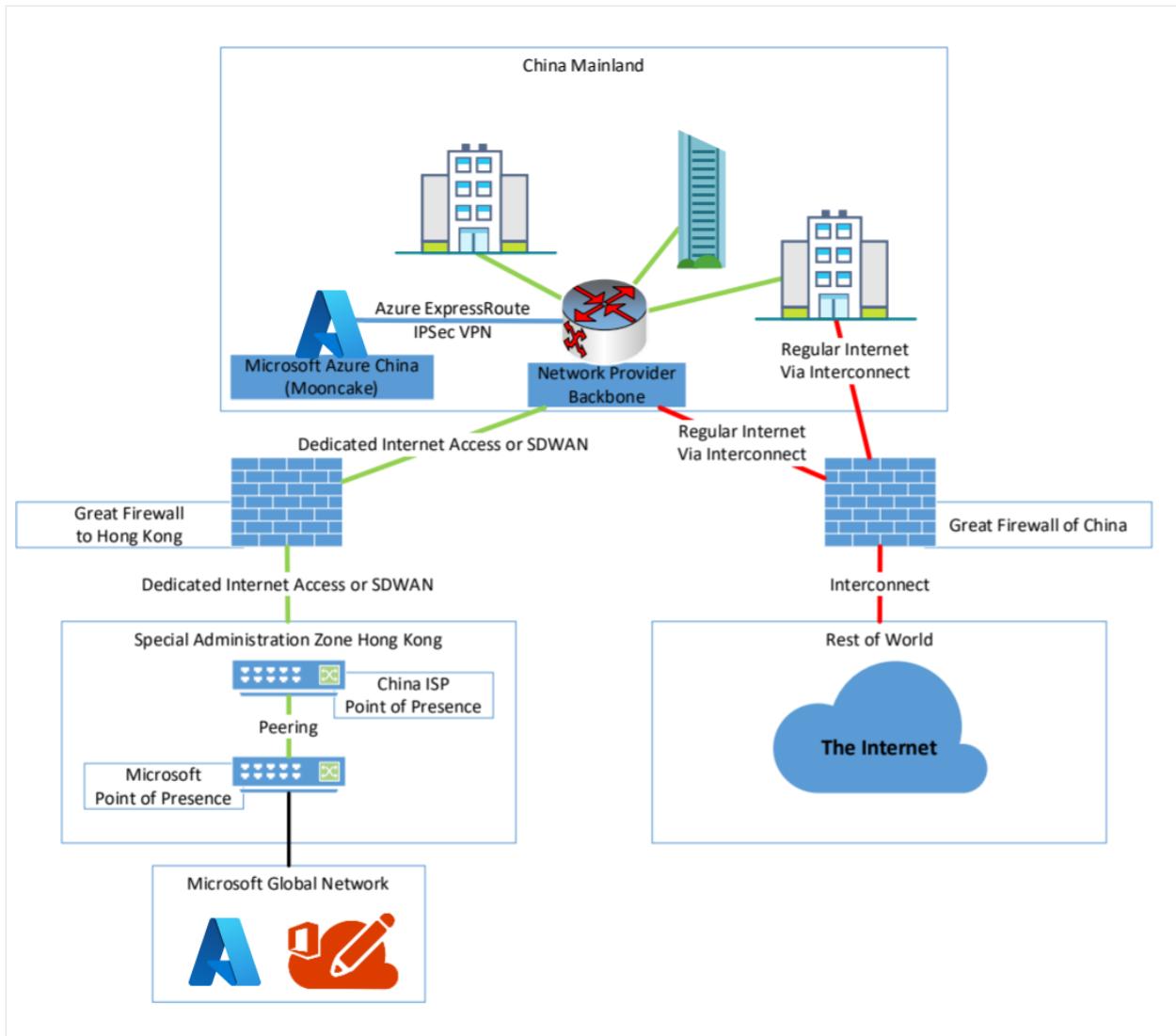
# Interconnect with China using Azure Virtual WAN and Secure Hub

Article • 02/15/2023

When looking at common automotive, manufacturing, logistics industries, or other institutes like embassies, there is often the question about how to improve interconnection with China. Those improvements are mostly relevant for using Cloud Services like Microsoft 365, Azure Global Services, or interconnect branches inside of China with a customer backbone.

In most of the cases, customers are struggling with high latencies, low bandwidth, unstable connection, and high costs connecting to outside of China (for example, Europe or the United States).

A reason for these struggles is the "Great Firewall of China", which protects the Chinese part of the Internet and filters traffic to China. Nearly all traffic running from People's Republic of China to outside of China, except the special administration zones like Hong Kong and Macau, passes the Great Firewall. The traffic running through Hong Kong and Macau doesn't hit the Great Firewall in full force, it's handled by a subset of the Great Firewall.



Using Virtual WAN, a customer can establish a more performant and stable connection to Microsoft Cloud Services and a connection to their enterprise network without breaking the Chinese cybersecurity law.

## Requirements and workflow

If you want to stay compliant to the Chinese cybersecurity law, you need to meet a set of certain conditions.

First, you need to work together with a network and ISP who owns an ICP (Internet Content Provider) license for China. In most cases, you'll end up with one of the following providers:

- China Telecom Global Ltd.
- China Mobile Ltd.
- China Unicom Ltd.
- PCCW Global Ltd.
- Hong Kong Telecom Ltd.

Depending on the provider and your needs, you now need to purchase one of the following network connectivity services to interconnect your branches within China.

- A MPLS/IPVPN Network
- A Software Defined WAN (SDWAN)
- Dedicated Internet Access

Next, you need to agree with that provider to give a breakout to the Microsoft Global Network and its Edge Network in Hong Kong, not in Beijing or Shanghai. In this case, Hong Kong is very important because of its physical connection and location to China.

While most customers think using Singapore for interconnect is the best case because it looks nearer to China when looking on the map, this isn't true. When you follow network fiber maps, nearly all network connects go through Beijing, Shanghai, and Hong Kong. This makes Hong Kong a better location choice to interconnect to China.

Depending on the provider, you may get different service offerings. The table below shows an example of providers and the service they offer, based on information at the time this article was written.

Service	Provider examples
MPLS/IPVPN Network	PCCW, China Telecom Global
SDWAN	PCCW, China Telecom Global
Dedicated Internet Access	PCCW, Hong Kong Telecom, China Mobil

With your provider, you can agree on which of the following two solutions to use to reach the Microsoft global backbone:

- Getting a Microsoft Azure ExpressRoute terminated in Hong Kong. That would be the case for the use of MPLS/IPVPN. Currently, only the only ICP license provider with ExpressRoute to Hong Kong is China Telecom Global. However, they can also talk to the other providers if they leverage Cloud Exchange Providers like Megaport or InterCloud. For more information, see [ExpressRoute connectivity providers](#).
- Using a Dedicated Internet Access directly at one of the following Internet Exchange Points, or using a private network interconnect.

The following list shows Internet Exchanges possible in Hong Kong:

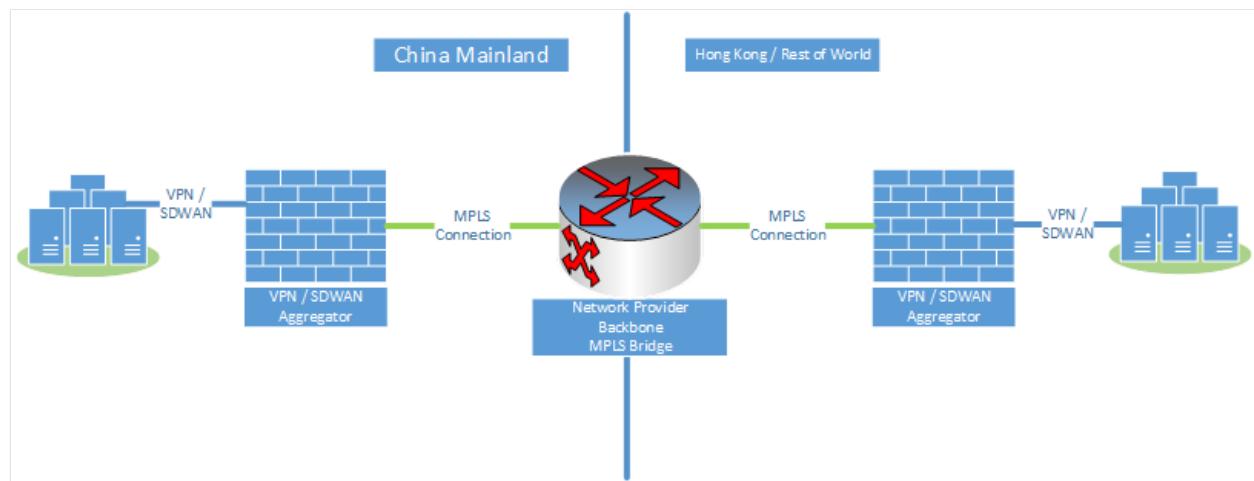
- AMS-IX Hong Kong
- BBIX Hong Kong

- Equinix Hong Kong
- HKIX

When using this connect, your next BGP hop for Microsoft Services must be Microsoft Autonomous System Number (AS#) 8075. If you use a single location or SDWAN solution, that would be the choice of connection.

With the current changes regarding interconnects between China and Hong Kong SAR, most of these network providers build an MPLS bridge between China and Hong Kong SAR.

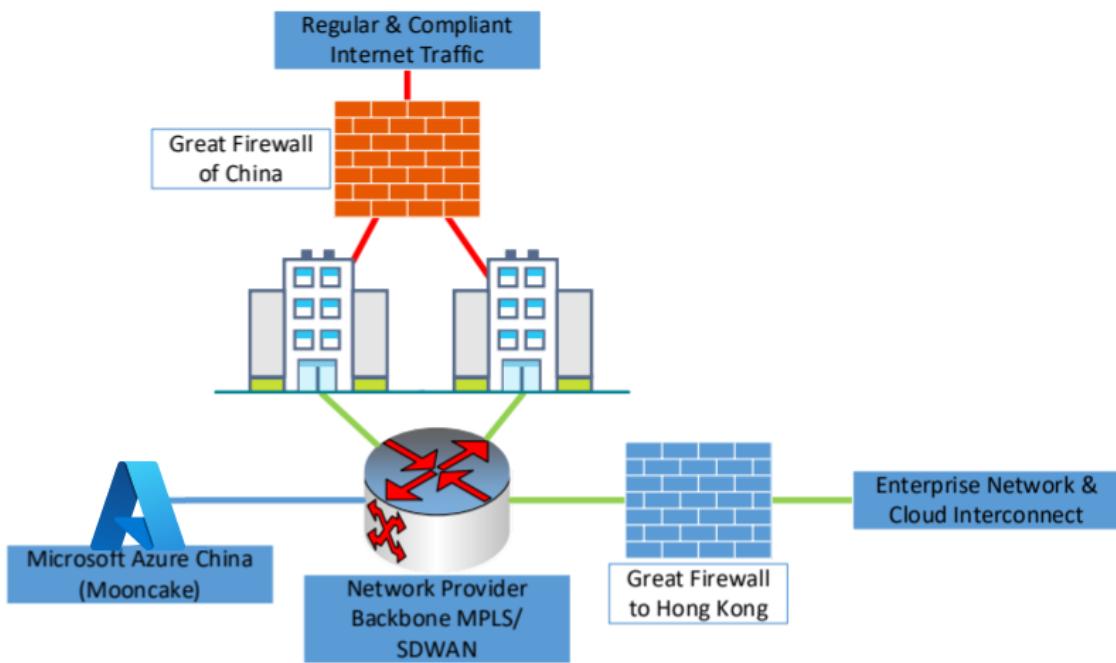
You can see that site-to-site VPN connections inside China are allowed and are mostly stable. The same applies for the site-to-site connections between branches in the rest of the world. Providers now create a VPN/SDWAN Aggregation on both sides and bridge via MPLS between them.



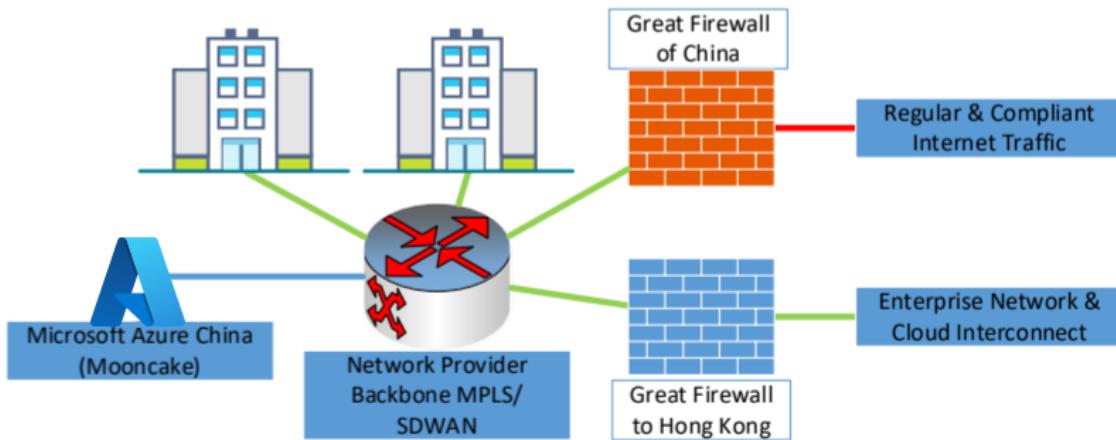
Either way, we still recommend that you have a second and regular internet breakout into China. This is to split the traffic between enterprise traffic to cloud services like Microsoft 365 and Azure, and by-law regulated internet traffic.

A compliant network architecture within China could look like the following example:

### Multiple Branches using WAN Backbone & Internet Breakout per Branch



### Multiple Branches using WAN Backbone & central managed Internet Breakout

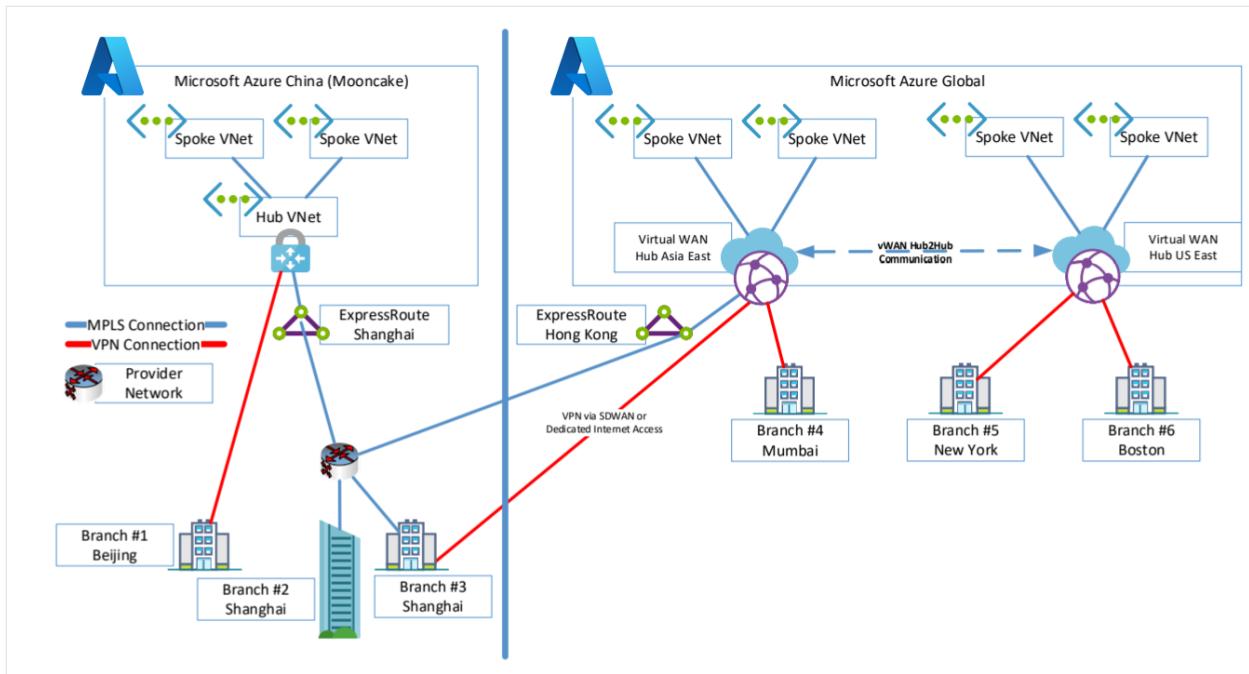


In this example, having an interconnect with the Microsoft Global Network in Hong Kong, you can now start to leverage the [Azure Virtual WAN Global Transit Architecture](#) and additional services, like Azure secure Virtual WAN hub, in order to consume services and interconnect to your branches and datacenter outside China.

## Hub-to-hub communication

In this section, we use Virtual WAN hub-to-hub communication to interconnect. In this scenario, you create a new Virtual WAN hub resource to connect to a Virtual WAN hub in Hong Kong, other regions you prefer, a region where you already have Azure resources, or where want to connect.

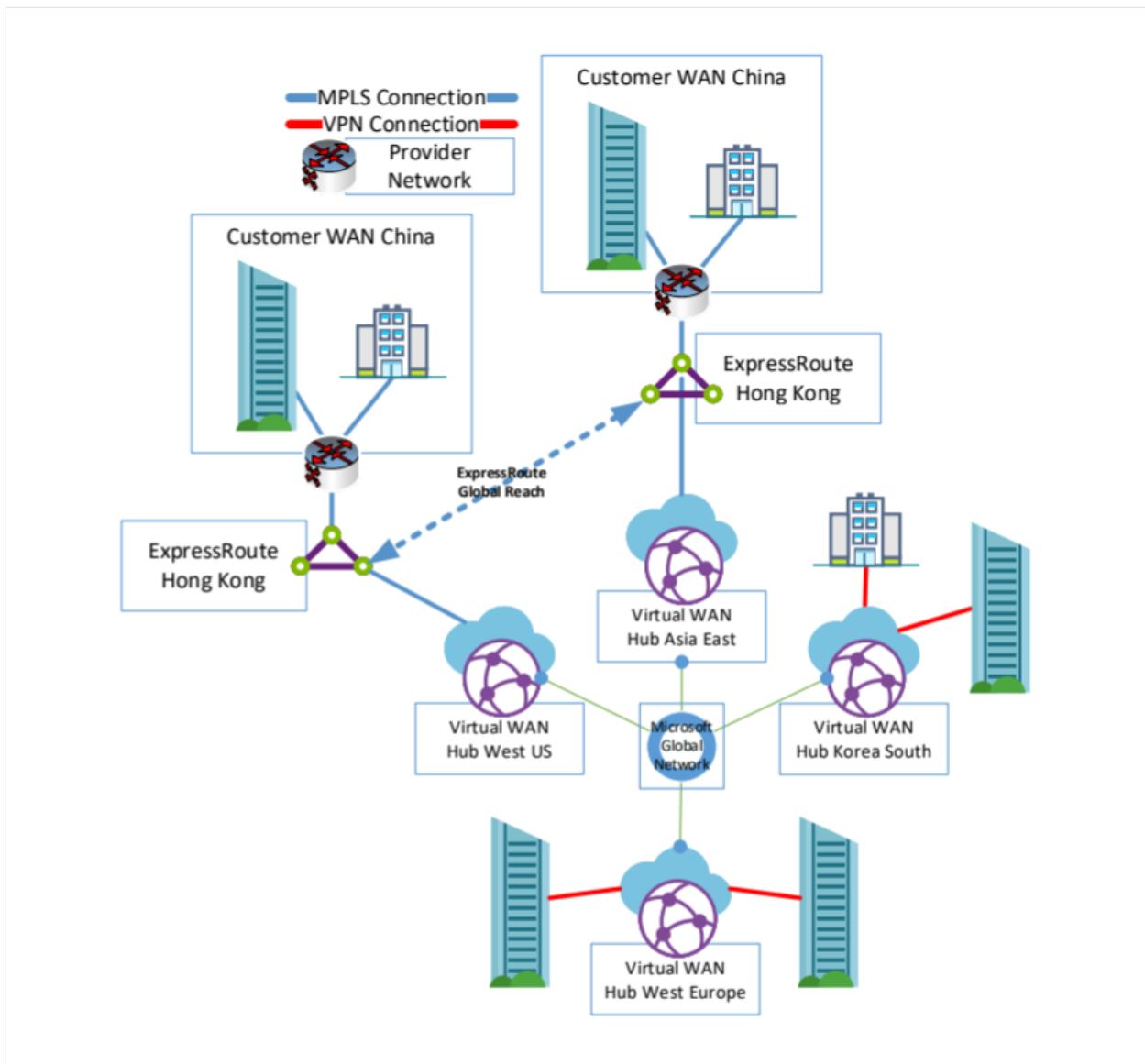
A sample architecture could look like following example:



In this example, the China branches connect to Azure Cloud China and each other by using VPN or MPLS connections. Branches that need to be connected to Global Services use MPLS or Internet-based services that are connected directly to Hong Kong. If you want to use ExpressRoute in Hong Kong and in the other region, you need to configure [ExpressRoute Global Reach](#) to interconnect both ExpressRoute Circuits.

ExpressRoute Global Reach isn't available in some regions. If you need to interconnect with Brazil or India, for example, you need to leverage [Cloud Exchange Providers](#) to provide the routing services.

The figure below shows both examples for this scenario.

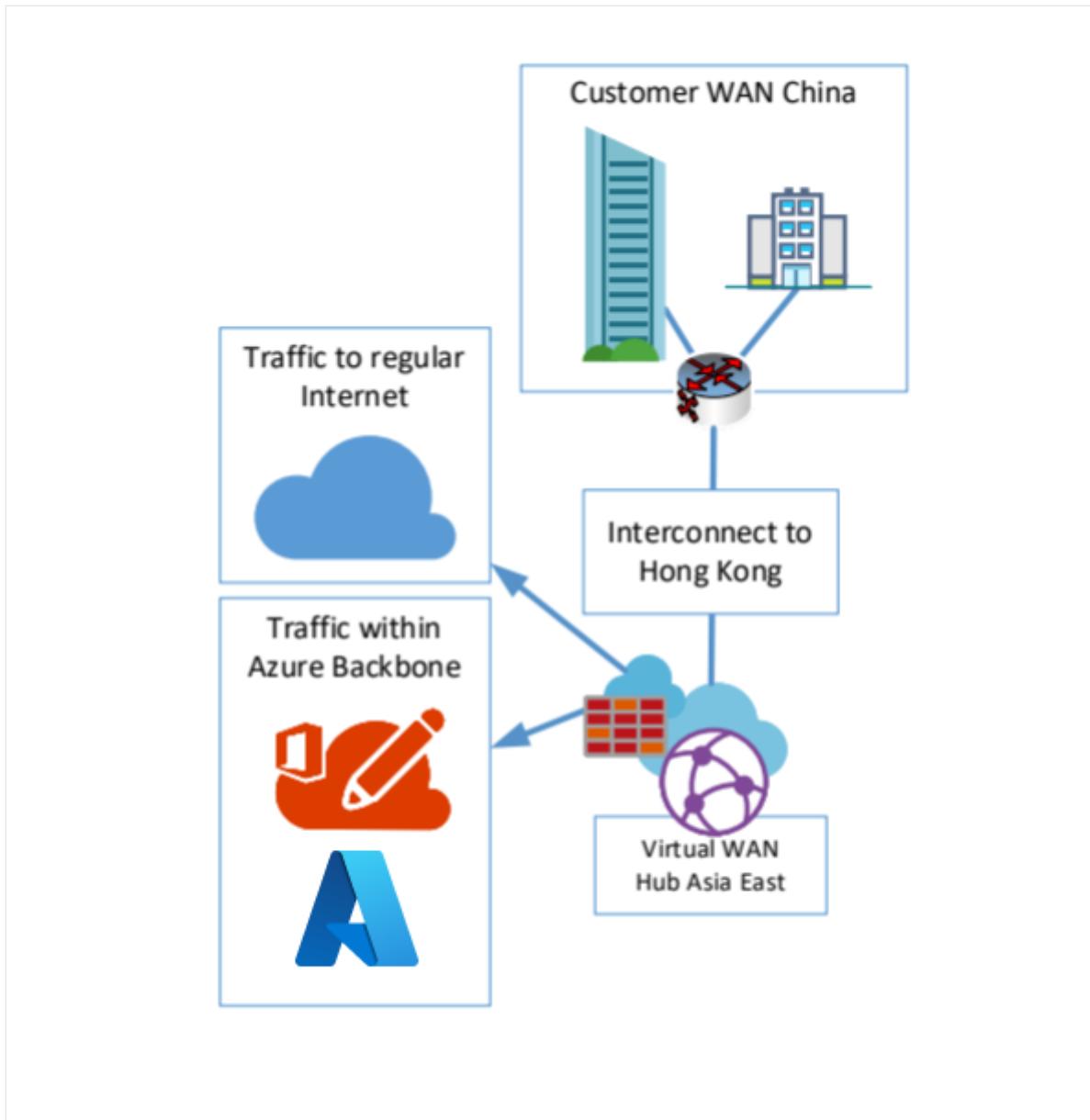


## Secure Internet breakout for Microsoft 365

Another consideration is network security and logging for the entry point between China and the Virtual WAN established backbone component, and the customer backbone. In most cases, there is a need to breakout to the Internet in Hong Kong to directly reach the Microsoft Edge Network and, with that, the Azure Front Door Servers used for Microsoft 365 Services.

For both scenarios with Virtual WAN, you would leverage the [Azure Virtual WAN secured hub](#). Using Azure Firewall Manager, you can change a regular Virtual WAN hub to a secured hub, and then deploy and manage an Azure Firewall within that hub.

The following figure shows an example of this scenario:



## Architecture and traffic flows

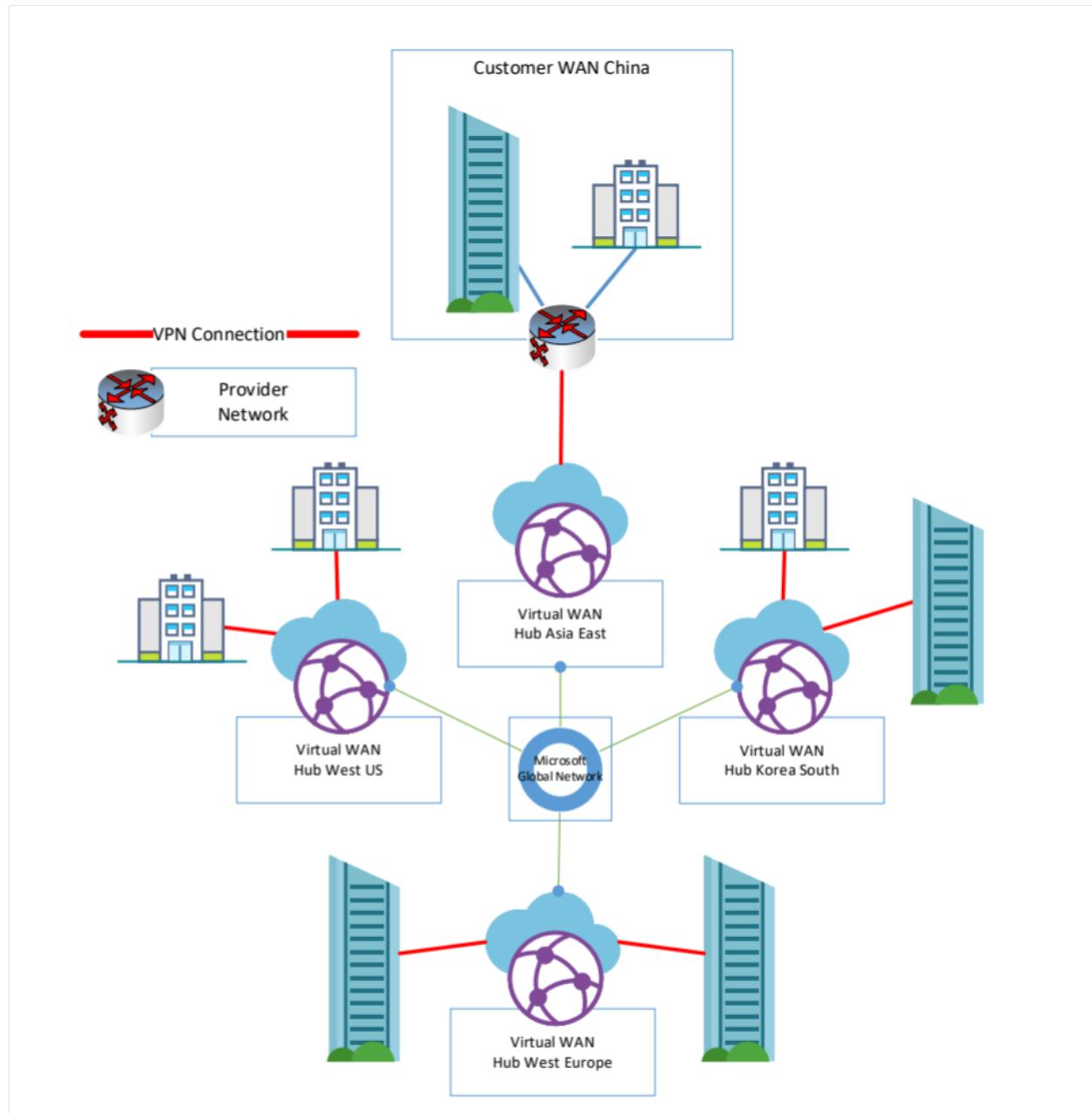
Depending on your choice regarding the connection to Hong Kong, the overall architecture may change slightly. This section shows three available architectures in different combination with VPN or SDWAN and/or ExpressRoute.

All of these options make use of Azure Virtual WAN secured hub for direct Microsoft 365 connectivity in Hong Kong. These architectures also support the compliance requirements for [Microsoft 365 Multi-Geo](#) and keep that traffic near the next Azure Front Door location. As a result, it's also an improvement for the usage of Microsoft 365 out of China.

When using Azure Virtual WAN together with Internet connections, every connection can benefit from additional services like [Microsoft Azure Peering Services \(MAPS\)](#). MAPS was built to optimize traffic coming to the Microsoft Global Network from 3rd Party Internet Service Providers.

## Option 1: SDWAN or VPN

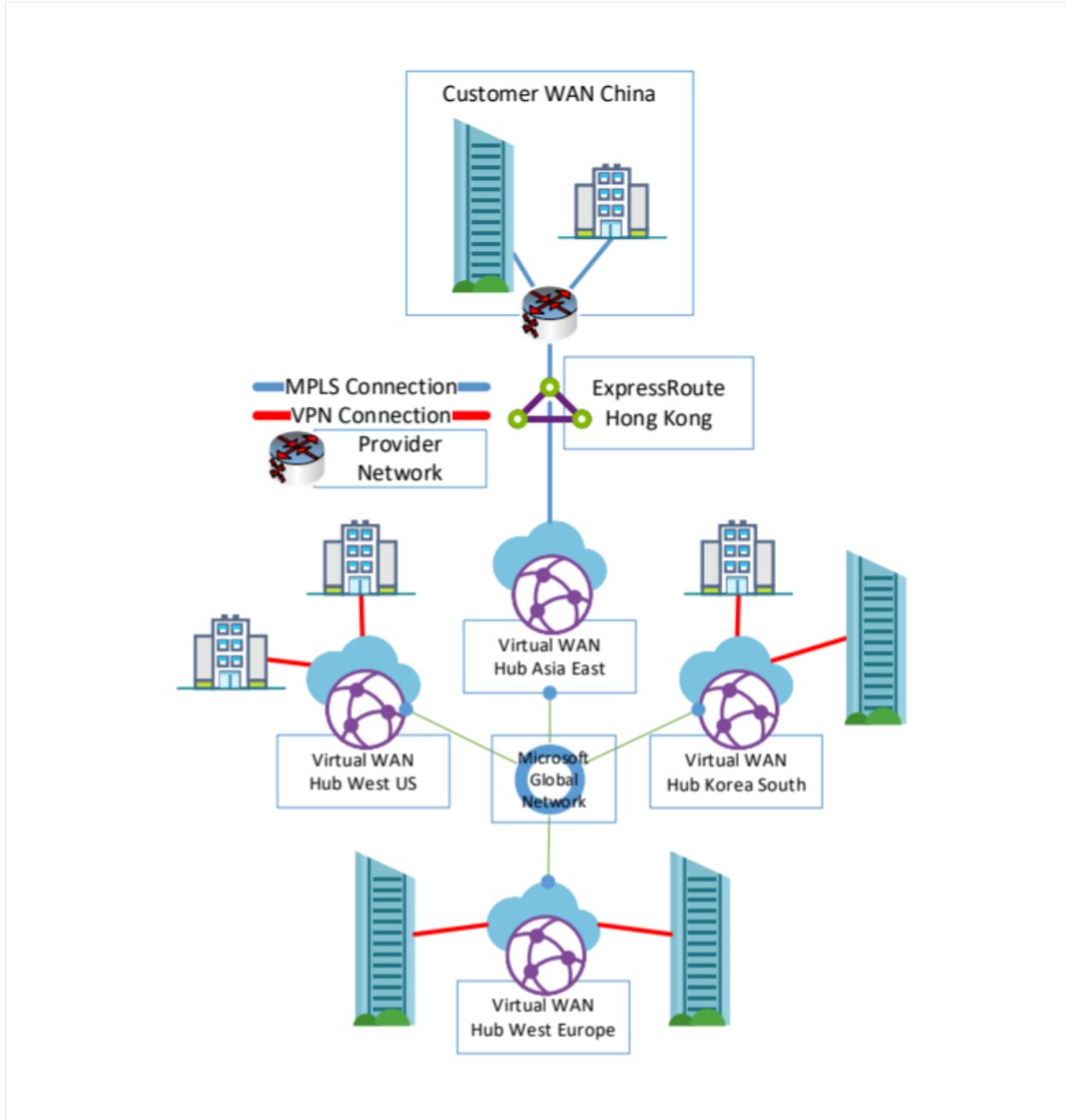
This section discusses a design that uses SDWAN or VPN to Hong Kong and to other branches. This option shows the use and traffic flow when using pure Internet connection on both sites of the Virtual WAN backbone. In this case, the connection is brought to Hong Kong using dedicated Internet access, or an ICP provider SDWAN solution. Other branches are using pure Internet or SDWAN Solutions as well.



In this architecture, every site is connected to the Microsoft Global Network by using VPN and Azure Virtual WAN. The traffic between the sites and Hong Kong is transmitted through the Microsoft Network and only uses regular Internet connection on the last mile.

## Option 2: ExpressRoute and SDWAN or VPN

This section discusses a design that uses ExpressRoute in Hong Kong and other Branches with VPN/SDWAN Branches. This option shows the use of ExpressRoute terminated in Hong Kong and other branches connected via SDWAN or VPN. ExpressRoute in Hong Kong is currently limited to a short list of Providers, which you can find in the list of [Express Route Partners](#).



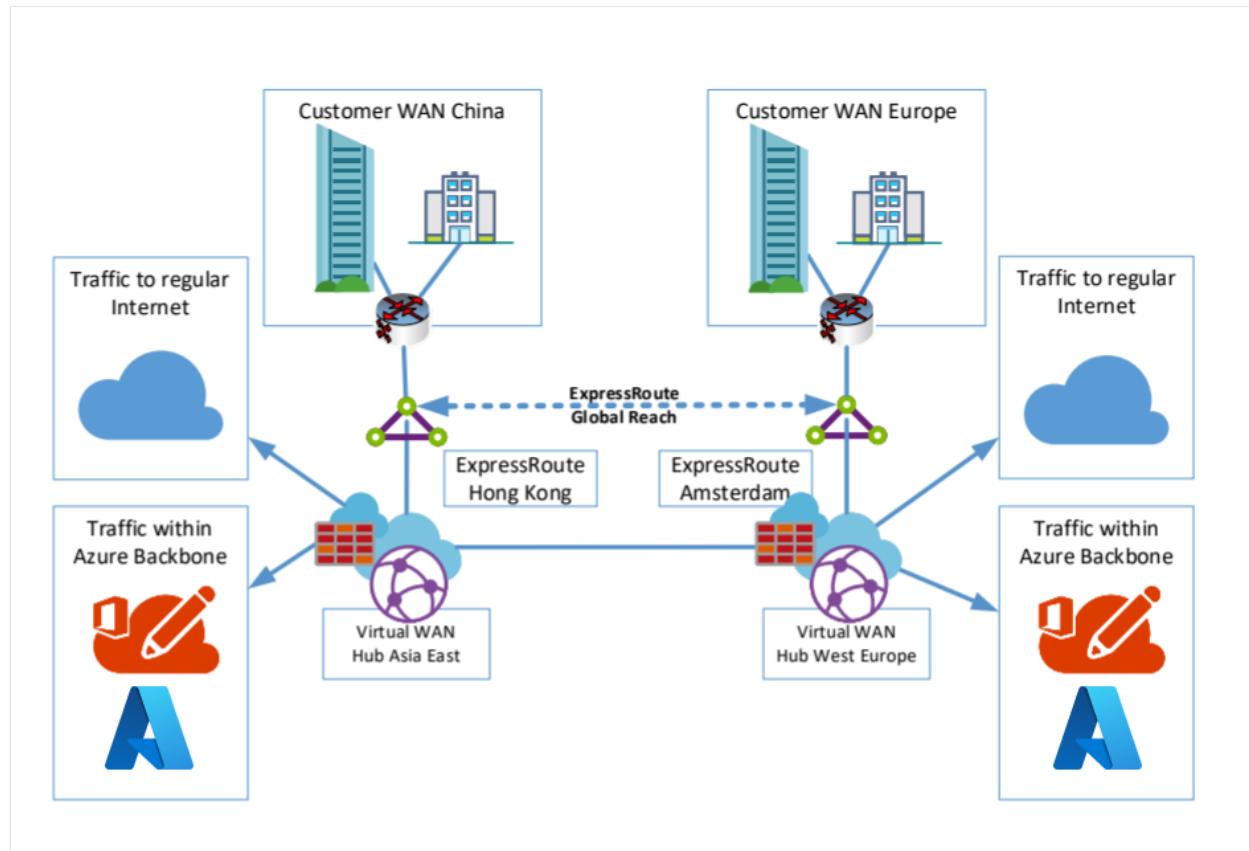
There are also options to terminate ExpressRoute from China, for example, in South Korea or Japan. But, given compliance, regulation, and latency, Hong Kong is currently the best choice.

## Option 3: ExpressRoute only

This section discusses a design that where ExpressRoute is used for Hong Kong and other Branches. This option shows the interconnect using ExpressRoute on both ends.

Here you have a different traffic flow than the other. The Microsoft 365 traffic will flow to the Azure virtual WAN secured hub and from there to the Microsoft Edge Network and the Internet.

The traffic that goes to the interconnected branches or from them to the locations in China will follow a different approach within that architecture. Currently virtual WAN doesn't support ExpressRoute to ExpressRoute transit. The traffic will leverage ExpressRoute Global Reach or the 3rd Party interconnect without passing the virtual WAN Hub. It will directly flow from one Microsoft Enterprise Edge (MSEE) to another.



Currently ExpressRoute Global Reach isn't available in every country/region, but you can configure a solution using Azure Virtual WAN.

You can, for example, configure an ExpressRoute with Microsoft Peering and connect a VPN tunnel through that peering to Azure Virtual WAN. Now you have enabled, again, the transit between VPN and ExpressRoute without Global Reach and 3rd party provider and service, such as Megaport Cloud.

## Next steps

See the following articles for more information:

- [Global Transit network architecture with Azure Virtual WAN](#)

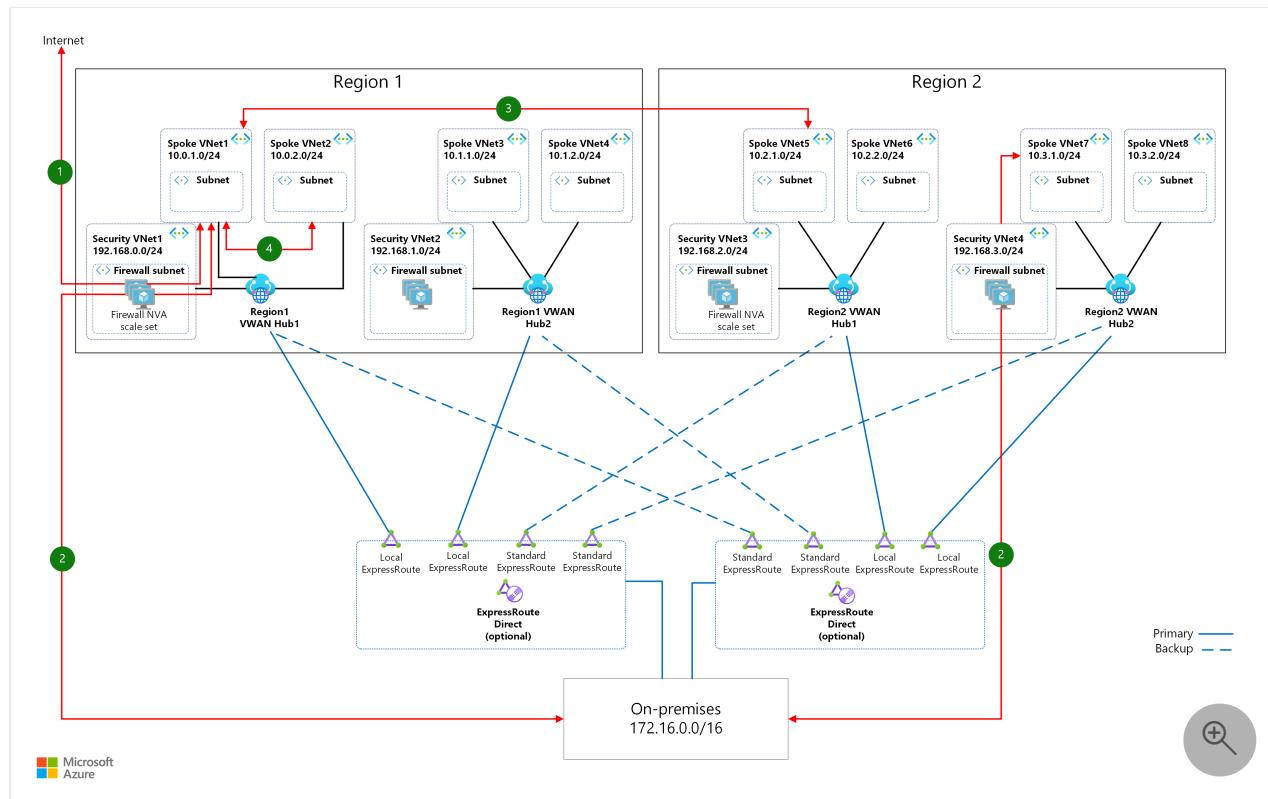
- Create a Virtual WAN hub
- Configure a Virtual WAN secured hub
- Azure Peering Service Preview Overview

# Massive-scale VWAN architecture design

Azure Virtual WAN   Azure Virtual Machine Scale Sets   Azure ExpressRoute

This example workload shows an Azure Virtual WAN deployment with multiple hubs per region. To improve availability and scalability, each hub peers to geographically dispersed, redundant Azure ExpressRoute circuits. This architecture is for exceptionally large and critical workloads. It supports business units and applications that reside on spoke virtual networks. The spoke virtual networks often have security requirements for internet-to-spoke or spoke-to-spoke connectivity.

## Architecture



Download a [Visio file](#) of this architecture.

## Workflow

The following workflow corresponds to the previous diagram:

1. Traffic from the spoke virtual networks to the internet routes through the network virtual appliance (NVA) firewalls in the security virtual networks that are attached to the same hub as the spoke.
2. The NVAs that are connected to the same hub as the spoke source or destination inspect all traffic between the spoke virtual networks and on-premises. This routing optimizes

performance and retains secure traffic between on-premises and Azure.

3. Traffic between spokes that reside on different hubs follow the path *spoke > hub > hub > spoke*. If spoke owners want more inspection, they must implement it within their spokes. This traffic doesn't traverse ExpressRoute connections, and security virtual network NVAs don't inspect it.
4. Spoke-to-spoke traffic on the same hub follows the path *spoke > hub > spoke*. Security virtual network NVAs don't inspect this traffic.

## Components

- [ExpressRoute](#) is a service that provides a private connection between your on-premises environment and Azure resources.
- [Virtual WAN](#) is a networking service that provides optimized and automated branch to branch connectivity through Azure. It provides transit for networking and routing via ExpressRoute between your on-premises resources and your Azure resources.
  - Custom route tables optimize routing in the solution, so network-to-network traffic can bypass the firewalls. Traffic between networks and on-premises environments remains inspected.
  - Labels simplify the routing by eliminating the need to extensively propagate the routes of individual networks to all route tables.
- [NVAs](#) are virtual machines that control routing to manage the flow of network traffic. This architecture uses NVAs. Large organizations with established investment in firewall technology and management often require NVAs.

## Alternatives

An alternative is a hub-and-spoke virtual network model with Azure route servers. You can have better performance than the 50-Gbps limit per hub. This alternative has better performance limits but more complexity. For more information, see [Hub-spoke network topology in Azure](#).

As another alternative, ExpressRoute Direct splits off ExpressRoute circuits into local and standard circuits. This service can optimize cost if the necessary bandwidth is sufficient to justify using ExpressRoute Direct.

## Scenario details

This deployment maximizes the scalability of Virtual WAN by using multiple Virtual WAN hubs per region. To find the number of virtual network connections that each hub can support, you subtract the total number of Virtual WAN hubs in your solution from 500. In this solution with four hubs, each hub can support 496 virtual network connections. Performance scales linearly with the number of hubs, so this solution provides exceptional performance and virtual network scaling.

This solution uses an open bow-tie design for ExpressRoute connectivity to the Virtual WAN hub. Each hub has two geographically dispersed ExpressRoute circuits. This design solves many problems and enables the use of NVAs.

ExpressRoute is a preferred path for Virtual WAN because traffic can travel between two spokes that are attached to different hubs, for instance between Spoke VNet1 and Spoke VNet5. If the design is a complete bow tie with a single ExpressRoute circuit that connects to Region1 VWAN Hub1 and Region2 VWAN Hub1, traffic between the spokes starts at Spoke VNet1 and then goes to Region1 VWAN Hub1. It goes down the ExpressRoute circuit and then back up the ExpressRoute path to Region2 VWAN Hub1 and then to Spoke VNet5. The open bow-tie design eliminates that path and enables the spoke-to-hub-to-hub-to-spoke path.

This solution uses different ExpressRoute circuits, so you can use the local ExpressRoute SKU for all their standard operating traffic. The disaster recovery path is rarely used and is a standard circuit SKU, which optimizes the bandwidth cost in the solution.

Traffic can use the NVA in the security virtual network that's attached to the same hub as the virtual network where the source of the traffic resides. During an ExpressRoute failure, the backup path continues to use the local NVA. The backup path simplifies routing, optimizes performance by avoiding inspection in multiple regions, and minimizes the risk of asymmetric routes by limiting complexity.

Custom NVA design allows routing flexibility by using customer-defined route tables in Virtual WAN.

This deployment provides highly redundant ExpressRoute connectivity for each hub. Highly redundant NVAs are attached to each hub.

## Region1 Hub1 route tables

The following tables show the defined routing options for Region1 Hub1.

### Default (Hub1)

[+] Expand table

Destination	Next hop	Associated	Propagated	Labels
10.0.0.0/16	SecurityVNet1Connection/NVA internal IP address	Branches	Branches	default, Hub1Default

### Spokes (Hub1)

[+] Expand table

Destination	Next hop	Associated	Propagated	Labels
172.16.0.0/16	SecurityVNet1Connection/NVA internal IP address	Spoke VNet1, Spoke VNet2	-	AllWorkloadSpokes
0.0.0.0/0	SecurityVNet1Connection/NVA internal IP address	Spoke VNet1, Spoke VNet2	-	AllWorkloadSpokes

## Security (Hub1)

[\[+\] Expand table](#)

Destination	Next hop	Associated	Propagated	Labels
-	-	Security VNet1	-	Hub1SecuritySpokes, AllSecuritySpokes

## Region1 Hub2 route tables

The following tables show the defined routing options for Region1 Hub2.

### Default route table (Hub2)

[\[+\] Expand table](#)

Destination	Next hop	Associated	Propagated	Labels
10.1.0.0/16	SecurityVNet2Connection/NVA internal IP address	Branches	Branches	default, Hub2Default

### Spokes (Hub2)

[\[+\] Expand table](#)

Destination	Next hop	Associated	Propagated	Labels
172.16.0.0/16	SecurityVNet2Connection/NVA internal IP address	Spoke VNet3, Spoke VNet4	-	AllWorkloadSpokes
0.0.0.0/0	SecurityVNet2Connection/NVA internal IP address	Spoke VNet3, Spoke VNet4	-	AllWorkloadSpokes

## Security (Hub2)

[\[+\] Expand table](#)

Destination	Next hop	Associated	Propagated	Labels
-	-	Security VNet2	-	Hub2SecuritySpokes, AllSecuritySpokes

## Region2 Hub1 route tables

The following tables show the defined routing options for Region2 Hub1.

### Default (Hub3)

[\[+\] Expand table](#)

Destination	Next hop	Associated	Propagated	Labels
10.2.0.0/16	SecurityVNet3Connection/NVA internal IP address	Branches	Branches	default, Hub3Default

### Spokes (Hub3)

[\[+\] Expand table](#)

Destination	Next hop	Associated	Propagated	Labels
172.16.0.0/16	SecurityVNet3Connection/NVA internal IP address	Spoke VNet5, Spoke VNet6	-	AllWorkloadSpokes
0.0.0.0/0	SecurityVNet3Connection/NVA internal IP address	Spoke VNet5, Spoke VNet6	-	AllWorkloadSpokes

## Security (Hub3)

[\[+\] Expand table](#)

Destination	Next hop	Associated	Propagated	Labels
-	-	Security VNet3	-	Hub3SecuritySpokes, AllSecuritySpokes

## Region2 Hub2 route tables

The following tables show the defined routing options for Region2 Hub2.

### Default (Hub4)

[\[+\] Expand table](#)

Destination	Next hop	Associated	Propagated	Labels
10.3.0.0/16	SecurityVNet4Connection/NVA internal IP address	Branches	Branches	default, Hub4Default

### Spokes (Hub4)

[\[+\] Expand table](#)

Destination	Next hop	Associated	Propagated	Labels
172.16.0.0/16	SecurityVNet4Connection/NVA internal IP address	Spoke VNet7, Spoke VNet8	-	AllWorkloadSpokes
0.0.0.0/0	SecurityVNet3Connection/NVA internal IP address	Spoke VNet7, Spoke VNet8	-	AllWorkloadSpokes

## Security (Hub4)

[Expand table](#)

Destination	Next hop	Associated	Propagated	Labels
-	-	Security VNet4	-	Hub4SecuritySpokes, AllSecuritySpokes

## Labels

[Expand table](#)

Label	Propagated virtual network connections
AllWorkloadSpokes	SpokeVNet1Connection, SpokeVNet2Connection, SpokeVNet3Connection, SpokeVNet4Connection, SpokeVNet5Connection, SpokeVNet6Connection, SpokeVNet7Connection, SpokeVNet8Connection, SecurityVNet1Connection, SecurityVNet2Connection, SecurityVNet3Connection, SecurityVNet4Connection
AllSecuritySpokes	SpokeVNet1Connection, SpokeVNet2Connection, SpokeVNet3Connection, SpokeVNet4Connection, SpokeVNet5Connection, SpokeVNet6Connection, SpokeVNet7Connection, SpokeVNet8Connection
Hub1Default	SecurityVNet1Connection
Hub2Default	SecurityVNet2Connection

Label	Propagated virtual network connections
Hub3Default	SecurityVNet3Connection
Hub4Default	SecurityVNet4Connection
Hub1SecuritySpokes	SpokeVNet1Connection, SpokeVNet2Connection, SecurityVNet1Connection
Hub2SecuritySpokes	SpokeVNet3Connection, SpokeVNet4Connection, SecurityVNet2Connection
Hub3SecuritySpokes	SpokeVNet5Connection, SpokeVNet6Connection, SecurityVNet3Connection
Hub4SecuritySpokes	SpokeVNet7Connection, SpokeVNet8Connection, SecurityVNet4Connection

This network architecture integrates seamlessly with the Cloud Adoption Framework for Virtual WAN. The Virtual WAN service, ExpressRoute connections, firewalls, and, in this case, security virtual networks are in the connectivity subscription. The workloads, network security groups, and spoke virtual networks are in the workload or application owner's separate landing zone subscriptions.

For more information, see [Virtual WAN network topology](#).

## Potential use cases

This design is applicable to any business of sufficient size and footprint in Azure. The business might use this design to:

- Replace existing multiprotocol label switching (MPLS) or Virtual WAN third-party deployments.
- Connect massive-scale cloud environments to on-premises environments.
- Support various business units and applications with disparate requirements and ownership within one tenant.

## Recommendations

### ExpressRoute

- Customers with massive-scale networks often have previously established connectivity points and require high bandwidth for their circuits. If you migrate from a large-scale MPLS, such as NetBond, and require over 40-Gbps circuit connectivity, you can take advantage of your network infrastructure and establish ExpressRoute Direct. ExpressRoute Direct supports MACsec encryption for high-security workloads.

- For cost optimization, use local ExpressRoute connections to peer the primary ExpressRoute circuit to the regional hub of choice. The backup ExpressRoute circuit should use standard ExpressRoute connections.

## Spoke

- **Internet egress:** Egress internet traffic should route through the local NVA firewall that's connected to the same hub as the source virtual network for that traffic.
- **Internet ingress inspection:** Customers can inspect ingress internet connectivity for the spoke workloads. They can use Azure Application Gateway or Azure Front Door for WAF inspection of traffic into the spokes. Source network address translation (SNAT) is required to avoid routing conflicts with the 0.0.0.0/0 route that's advertised by the Virtual WAN hub.
- **Network security groups:** Use network security groups to customize the security of the application that resides in your spoke virtual network.

## NVA

- **Redundancy:** Follow a best practice architecture for NVA deployment redundancy. Use multiple virtual machines or scale sets and load balancers to provide front end and back end support.

## Virtual WAN hub routing

- Spoke virtual network connections should only propagate to route table labels and not to specific route tables. This practice simplifies approaches that use infrastructure as code.
- Each hub should have its own default hub label to allow and limit propagation of the security virtual network routes to only that hub's default route table. If you use the built-in default label, it propagates across all hubs.
- Each hub should have a route table label for that hub's security virtual network. This practice streamlines infrastructure as code because virtual network connections propagate to the label instead of a specific route table.

## Considerations

These considerations implement the pillars of the Azure Well-Architected Framework, which is a set of guiding tenets that can be used to improve the quality of a workload. For more information, see [Microsoft Azure Well-Architected Framework](#).

## Reliability

Reliability ensures your application can meet the commitments you make to your customers. For more information, see [Overview of the reliability pillar](#).

This workload optimizes high availability with Virtual WAN, redundant ExpressRoute circuits, and scale sets for NVAs. This combination results in the redundancy that's necessary for highly critical workloads.

## Security

Security provides assurances against deliberate attacks and the abuse of your valuable data and systems. For more information, see [Overview of the security pillar](#).

This workload provides firewall inspection between Azure and on-premises systems and inspection for outgoing internet traffic from Azure. For inbound internet traffic, consider Azure Front Door or Application Gateway. Use SNAT to avoid routing conflicts.

## Cost optimization

Cost optimization is about looking at ways to reduce unnecessary expenses and improve operational efficiencies. For more information, see [Overview of the cost optimization pillar](#).

For costs of Azure components, see the [Azure pricing calculator](#). Pricing for this solution is based on factors such as:

- The Azure services that are used.
- The ExpressRoute sizing.
- The Virtual WAN sizing and data traffic quantities that each hub processes.
- The NVA pricing.

This workload prioritizes performance and availability over low cost. But using ExpressRoute Local for primary connections optimizes cost because it limits bandwidth expenses. If you want to compromise performance and reliability to optimize cost, you can reduce the number of ExpressRoute circuits and firewalls. When you reduce these resources, it reduces cost but traverses the Virtual WAN hubs with less efficiency when you connect to on-premises or cloud destinations.

## Operational excellence

Operational excellence covers the operations processes that deploy an application and keep it running in production. For more information, see [Overview of the operational excellence pillar](#).

This design is compatible with Terraform and infrastructure as code. It requires the Terraform Azure API provider for deployment because of Virtual WAN lag in feature availability.

## Performance efficiency

Performance efficiency is the ability of your workload to scale to meet the demands placed on it by users in an efficient manner. For more information, see [Performance efficiency pillar overview](#).

This network is highly performant. Even if a connection fails, performance and routing use the best available path.

## Deploy this scenario

The following steps establish the Virtual WAN service, hubs, spoke virtual networks, and ExpressRoute connections. For a tutorial, see [Create an ExpressRoute association to Azure Virtual WAN](#).

1. Create a Virtual WAN service.
2. Deploy multiple hubs and an ExpressRoute gateway in each hub.
3. Deploy the required number of workload spoke virtual networks to support your workload and connect them to the desired hubs.
4. Establish connections between your ExpressRoute circuits and your hubs.
5. Deploy one security virtual network for each hub.
6. Deploy the NVA of your choosing and configure the firewall. Use NVA-specific documentation for this step. To establish the route tables and labels, use the example in [How to configure virtual hub routing: Azure portal - Azure Virtual WAN](#).
7. Verify the routing.

## Contributors

*This article is maintained by Microsoft. It was originally written by the following contributors.*

Principal authors:

- [Ethan Haslett](#) | Senior Cloud Solution Architect
- [John Poetzinger](#) | Senior Cloud Solution Architect

Other contributors:

- [Jimmy Avila](#) | Senior Cloud Solution Architect
- [Andrew Delosky](#) | Principal Cloud Solution Architect
- [Robert Lightner](#) | Senior Cloud Solution Architect
- [Rodrigo Santos](#) | Principal Cloud Solution Architect

*To see non-public LinkedIn profiles, sign in to LinkedIn.*

## Next steps

- [About virtual hub routing](#)
- [Route traffic through NVAs by using custom settings](#)
- [About ExpressRoute connections in Azure Virtual WAN](#)
- [What is Azure Virtual WAN](#)

## Related resources

- [Hub-spoke network topology with Azure Virtual WAN](#)
- [Migrate to Azure Virtual WAN](#)
- [Virtual WAN architecture optimized for department-specific requirements](#)
- [Virtual WAN network topology](#)

# Migrate to Azure Virtual WAN

Article • 12/14/2022

Azure Virtual WAN lets companies simplify their global connectivity in order to benefit from the scale of the Microsoft global network. This article provides technical details for companies that want to migrate from an existing customer-managed hub-and-spoke topology, to a design that leverages Microsoft-managed Virtual WAN hubs.

For information about the benefits that Azure Virtual WAN enables for enterprises adopting a cloud-centric modern enterprise global network, see [Global transit network architecture and Virtual WAN](#).

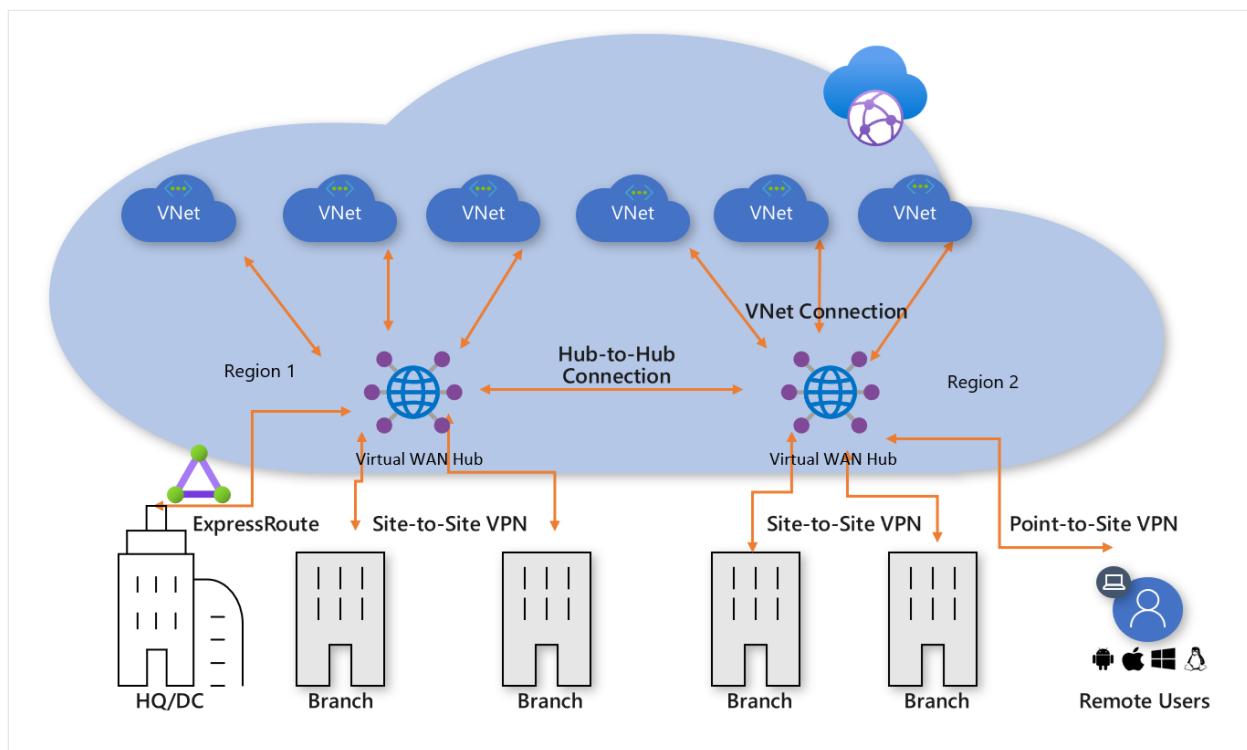


Figure: Azure Virtual WAN

The Azure hub-and-spoke connectivity model has been adopted by thousands of our customers to leverage the default transitive routing behavior of Azure Networking in order to build simple and scalable cloud networks. Azure Virtual WAN builds on these concepts and introduces new capabilities that allow global connectivity topologies, not only between on-premises locations and Azure, but also allowing customers to leverage the scale of the Microsoft network to augment their existing global networks.

This article shows how to migrate an existing customer-managed hub-and-spoke environment, to a topology that is based on Azure Virtual WAN.

## Scenario

Contoso is a global financial organization with offices in both Europe and Asia. They are planning to move their existing applications from an on-premises data center in to Azure and have built out a foundation design based on the customer-managed hub-and-spoke architecture, including regional hub virtual networks for hybrid connectivity. As part of the move to cloud-based technologies, the network team has been tasked with ensuring that their connectivity is optimized for the business moving forward.

The following figure shows a high-level view of the existing global network including connectivity to multiple Azure regions.

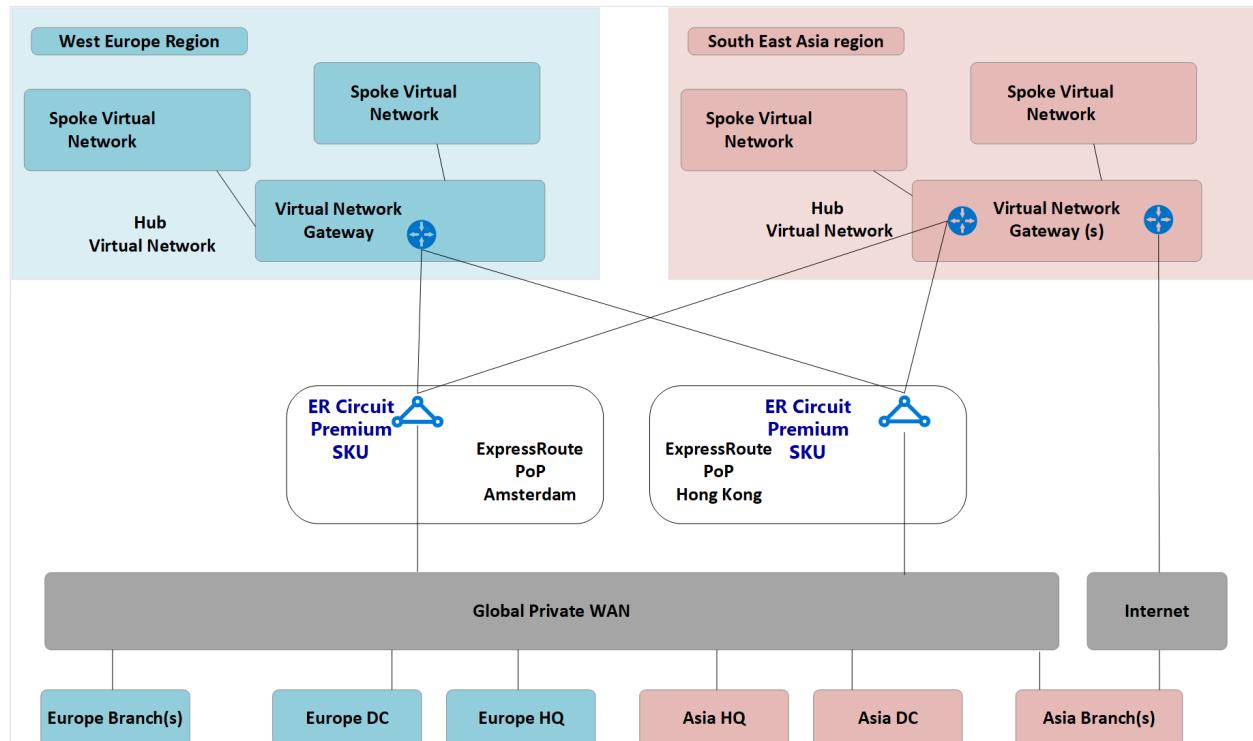


Figure: Contoso existing network topology

The following points can be understood from the existing network topology:

- A hub-and-spoke topology is used in multiple regions including ExpressRoute circuits for connectivity back to a common private Wide Area Network (WAN).
- Some of these sites also have VPN tunnels directly in to Azure to reach applications hosted within the cloud.

## Requirements

The networking team has been tasked with delivering a global network model that can support the Contoso migration to the cloud and must optimize in the areas of cost, scale, and performance. In summary, the following requirements are to be met:

- Provide both head quarter (HQ) and branch offices with optimized path to cloud hosted applications.

- Remove the reliance on existing on-premises data centers (DC) for VPN termination while retaining the following connectivity paths:
  - **Branch-to-VNet:** VPN connected offices must be able to access applications migrated to the cloud in the local Azure region.
  - **Branch-to-Hub to Hub-to-VNet:** VPN connected offices must be able to access applications migrated to the cloud in the remote Azure region.
  - **Branch-to-branch:** Regional VPN connected offices must be able to communicate with each other and ExpressRoute connected HQ/DC sites.
  - **Branch-to-Hub to Hub-to-branch:** Globally separated VPN connected offices must be able to communicate with each other and any ExpressRoute connected HQ/DC sites.
  - **Branch-to-Internet:** Connected sites must be able to communicate with the Internet. This traffic must be filtered and logged.
  - **VNet-to-VNet:** Spoke virtual networks in the same region must be able to communicate with each other.
  - **VNet-to-Hub to Hub-to-VNet:** Spoke virtual networks in the different regions must be able to communicate with each other.
- Provide the ability for Contoso roaming users (laptop and phone) to access company resources while not on the corporate network.

## Azure Virtual WAN architecture

The following figure shows a high-level view of the updated target topology using Azure Virtual WAN to meet the requirements detailed in the previous section.

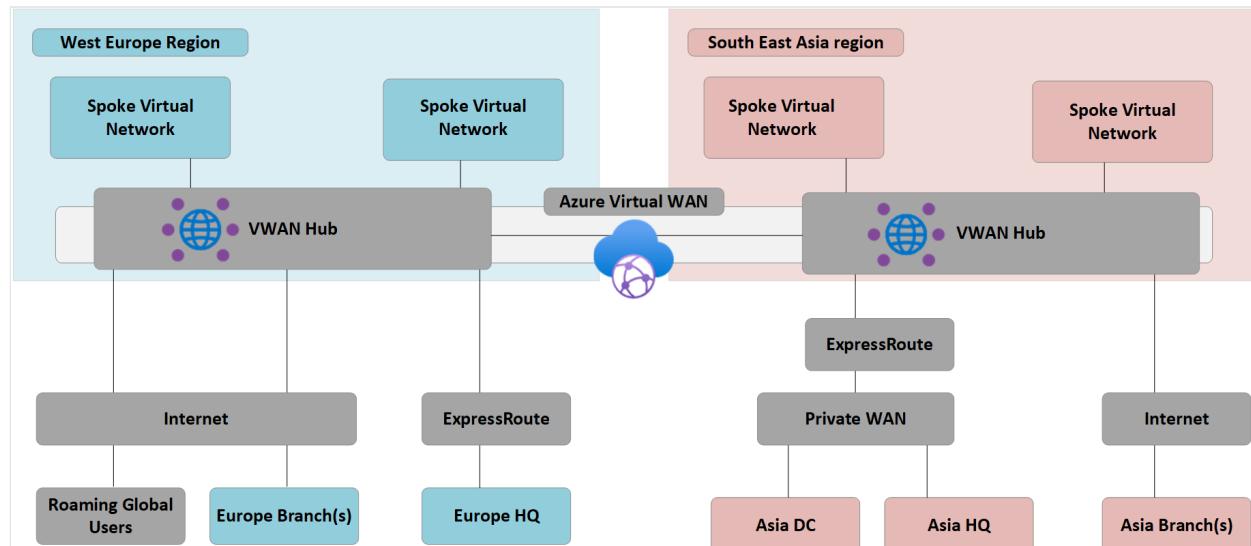


Figure: Azure Virtual WAN architecture

Summary:

- HQ in Europe remains ExpressRoute connected, Europe on-premises DC are fully migrated to Azure and now decommissioned.
- Asia DC and HQ remain connected to Private WAN. Azure Virtual WAN now used to augment the local carrier network and provide global connectivity.
- Azure Virtual WAN hubs deployed in both West Europe and South East Asia Azure regions to provide connectivity hub for ExpressRoute and VPN connected devices.
- Hubs also provide VPN termination for roaming users across multiple client types using OpenVPN connectivity to the global mesh network, allowing access to not only applications migrated to Azure, but also any resources remaining on-premises.
- Internet connectivity for resources within a virtual network provided by Azure Virtual WAN.

Internet connectivity for remote sites also provided by Azure Virtual WAN. Local internet breakout supported via partner integration for optimized access to SaaS services such as Microsoft 365.

## Migrate to Virtual WAN

This section shows the various steps for migrating to Azure Virtual WAN.

### Step 1: Single region customer-managed hub-and-spoke

The following figure shows a single region topology for Contoso prior to the rollout of Azure Virtual WAN:

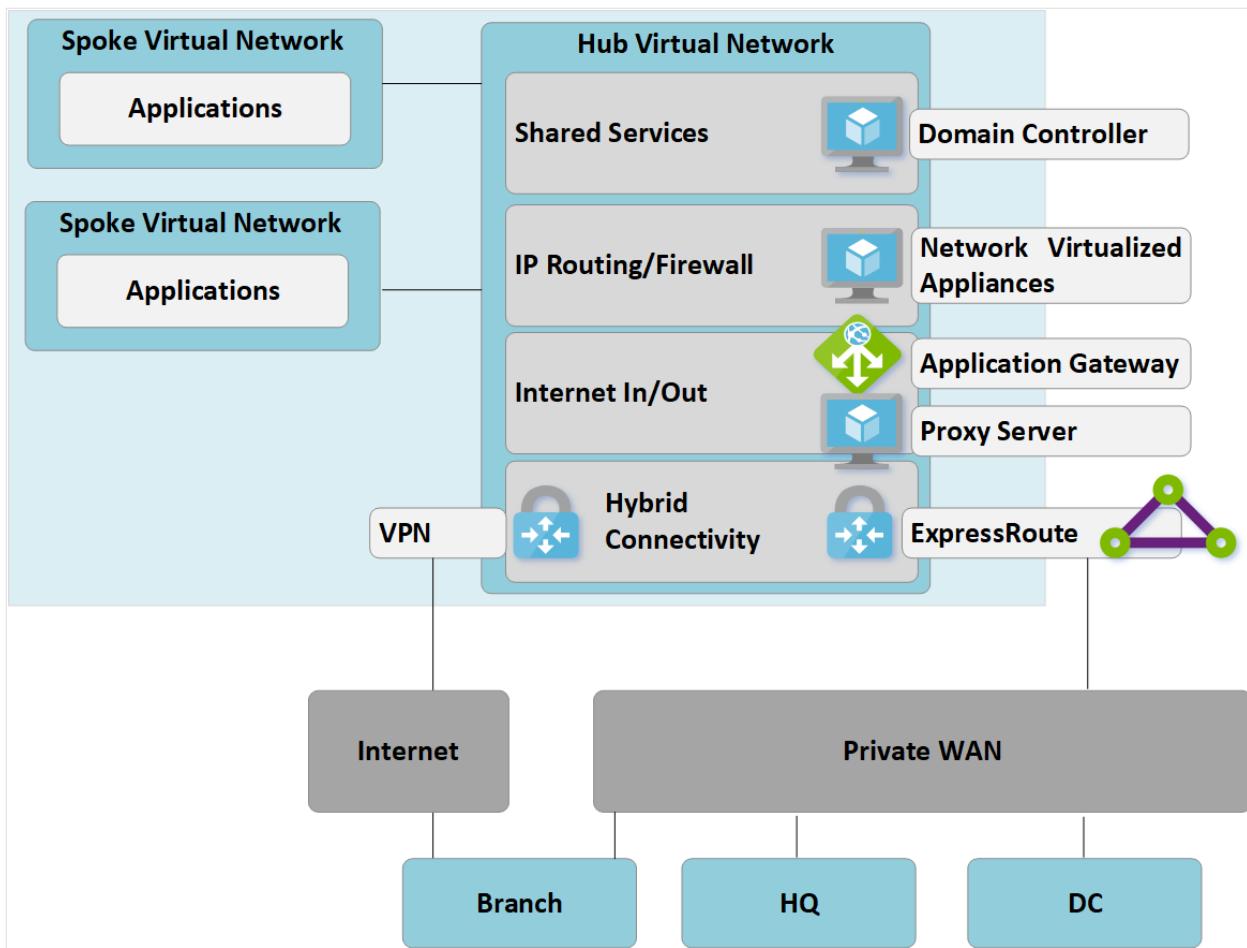


Figure 1: Single region manual hub-and-spoke

In keeping with the hub-and-spoke approach, the customer-managed hub virtual network contains several function blocks:

- Shared services (any common function required by multiple spokes). Example: Contoso uses Windows Server domain controllers on Infrastructure-as-a-service (IaaS) virtual machines.
- IP/Routing firewall services are provided by a third-party network virtual appliance, enabling spoke-to-spoke layer-3 IP routing.
- Internet ingress/egress services including Azure Application Gateway for inbound HTTPS requests and third-party proxy services running on virtual machines for filtered outbound access to internet resources.
- ExpressRoute and VPN virtual network gateway for connectivity to on-premises networks.

## Step 2: Deploy Virtual WAN hubs

Deploy a Virtual WAN hub in each region. Set up the Virtual WAN hub with VPN and ExpressRoute functionality as described in the following articles:

- [Tutorial: Create a Site-to-Site connection using Azure Virtual WAN](#)
- [Tutorial: Create an ExpressRoute association using Azure Virtual WAN](#)

## ⚠ Note

Azure Virtual WAN must be using the Standard SKU to enable some of the traffic paths shown in this article.

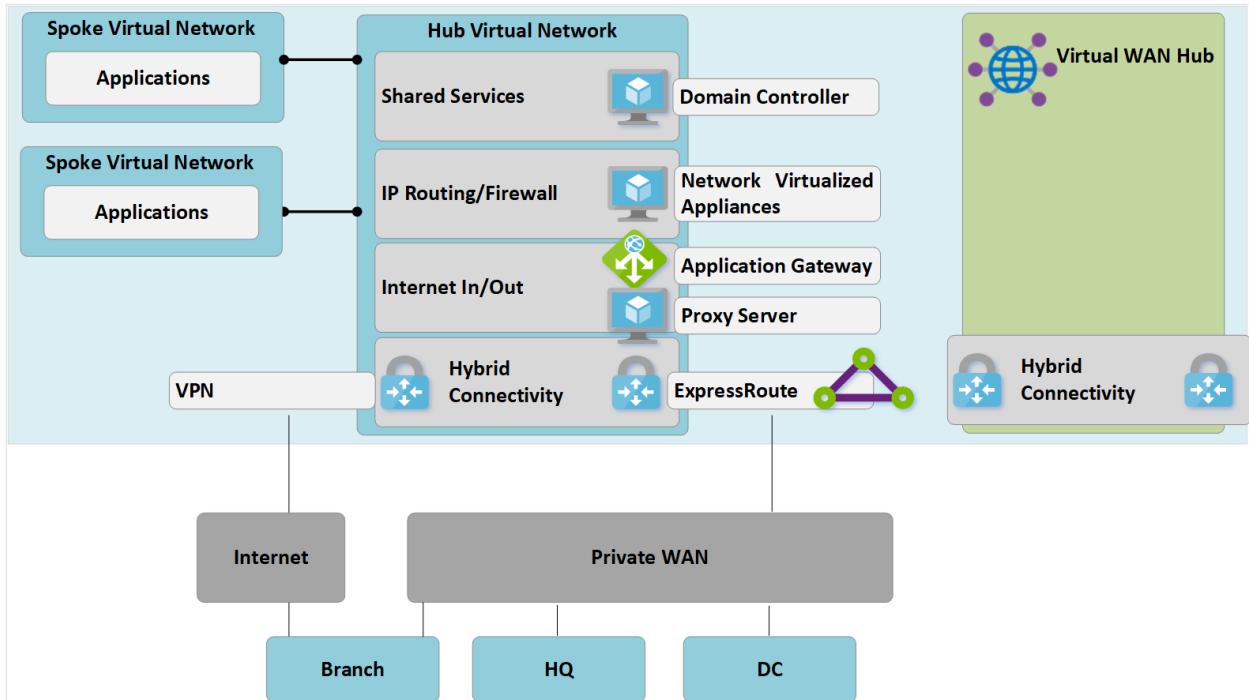


Figure 2: Customer-managed hub-and-spoke to Virtual WAN migration

## Step 3: Connect remote sites (ExpressRoute and VPN) to Virtual WAN

Connect the Virtual WAN hub to the existing ExpressRoute circuits and set up Site-to-site VPNs over the Internet to any remote branches.

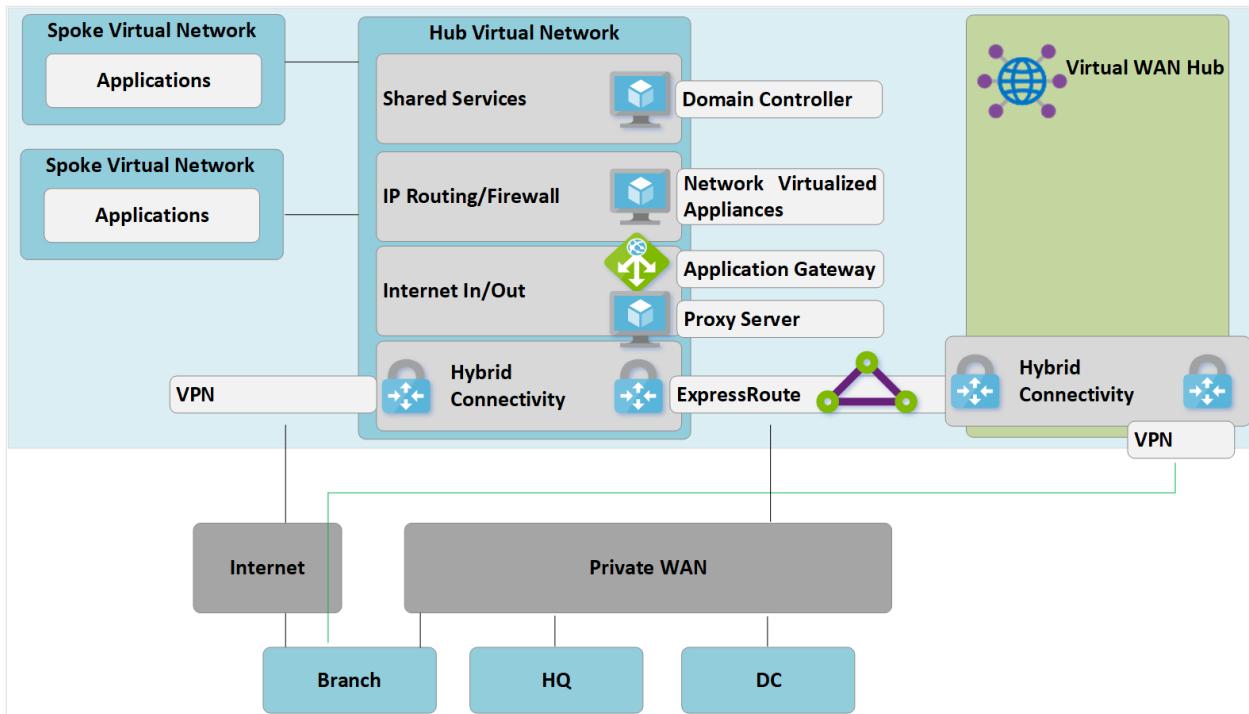
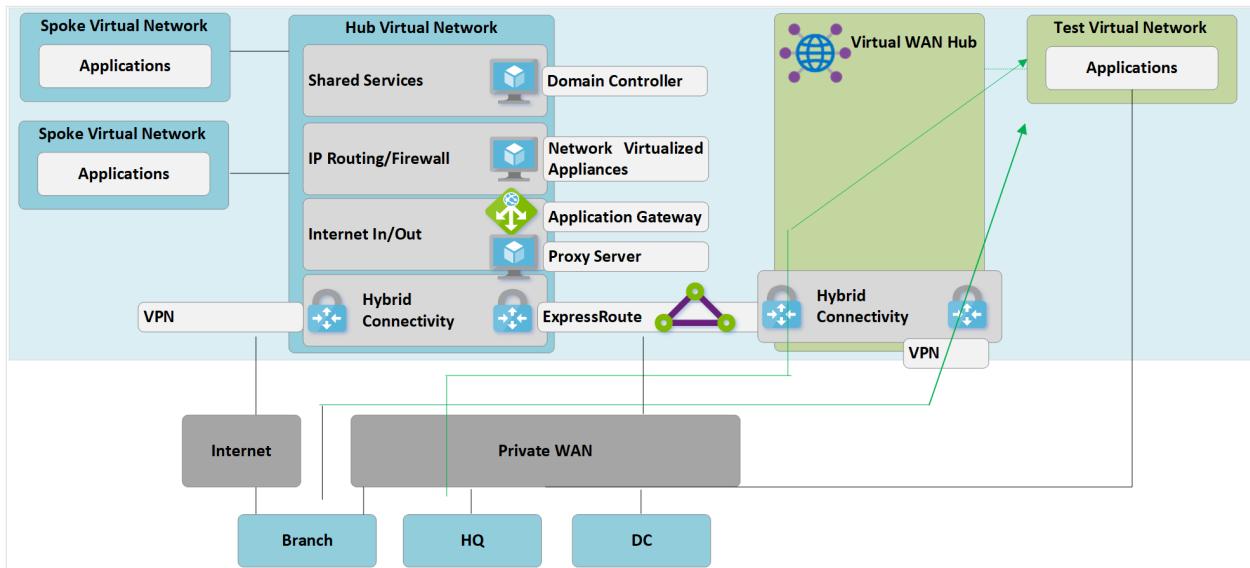


Figure 3: Customer-managed hub-and-spoke to Virtual WAN migration

At this point, on-premises network equipment will begin to receive routes reflecting the IP address space assigned to the Virtual WAN-managed hub VNet. Remote VPN-connected branches at this stage will see two paths to any existing applications in the spoke virtual networks. These devices should be configured to continue to use the tunnel to the customer-managed hub to ensure symmetrical routing during the transition phase.

## Step 4: Test hybrid connectivity via Virtual WAN

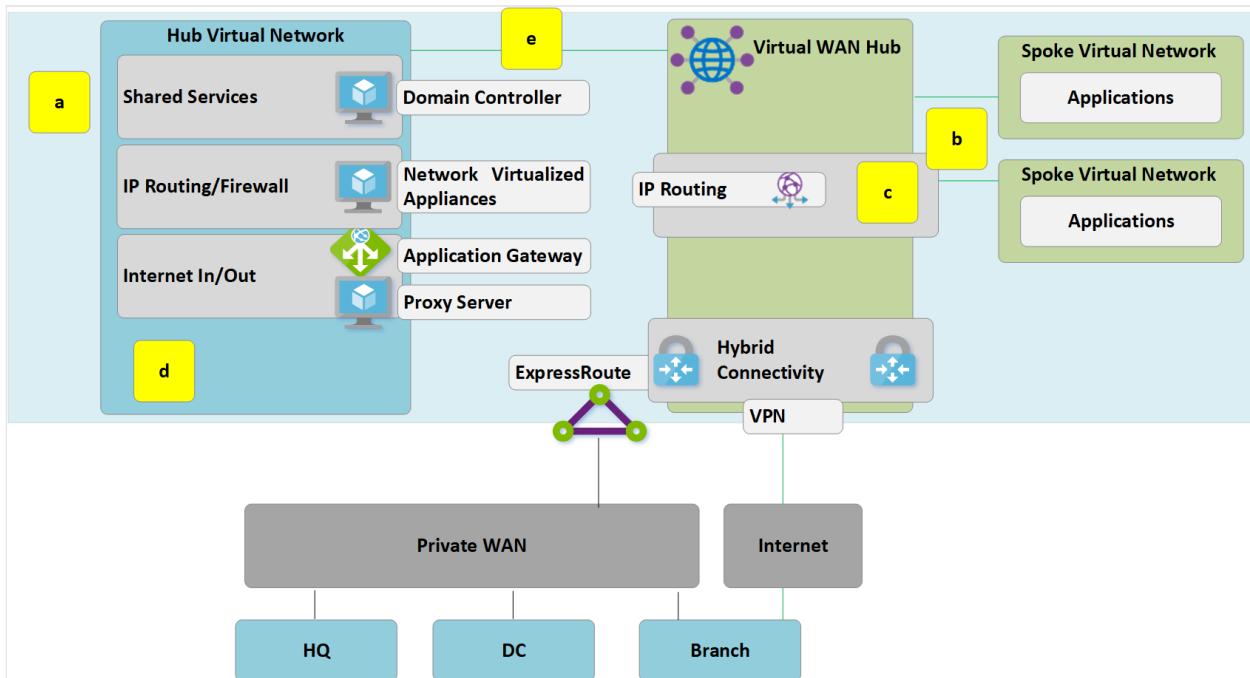
Prior to using the managed Virtual WAN hub for production connectivity, we recommend that you set up a test spoke virtual network and Virtual WAN VNet connection. Validate that connections to this test environment work via ExpressRoute and Site to Site VPN before continuing with the next steps.



**Figure 4: Customer-managed hub-and-spoke to Virtual WAN migration**

At this stage, it's important to recognize that both the original customer-managed hub virtual network and the new Virtual WAN Hub are both connected to the same ExpressRoute circuit. Due to this, we have a traffic path that can be used to enable spokes in both environments to communicate. For example, traffic from a spoke that is attached to the customer-managed hub virtual network will traverse the MSEE devices used for the ExpressRoute circuit to reach any spoke connected via a VNet connection to the new Virtual WAN hub. This allows a staged migration of spokes in Step 5.

## Step 5: Transition connectivity to virtual WAN hub



**Figure 5: Customer-managed hub-and-spoke to Virtual WAN migration**

- Delete the existing peering connections from Spoke virtual networks to the old customer-managed hub. Access to applications in spoke virtual networks is unavailable

until steps a-c are complete.

- b. Connect the spoke virtual networks to the Virtual WAN hub via VNet connections.
- c. Remove any user-defined routes (UDR) previously used within spoke virtual networks for spoke-to-spoke communications. This path is now enabled by dynamic routing available within the Virtual WAN hub.
- d. Existing ExpressRoute and VPN Gateways in the customer-managed hub are now decommissioned to permit the next step (e).
- e. Connect the old customer-managed hub (hub virtual network) to the Virtual WAN hub via a new VNet connection.

## Step 6: Old hub becomes shared services spoke

We have now redesigned our Azure network to make the Virtual WAN hub the central point in our new topology.

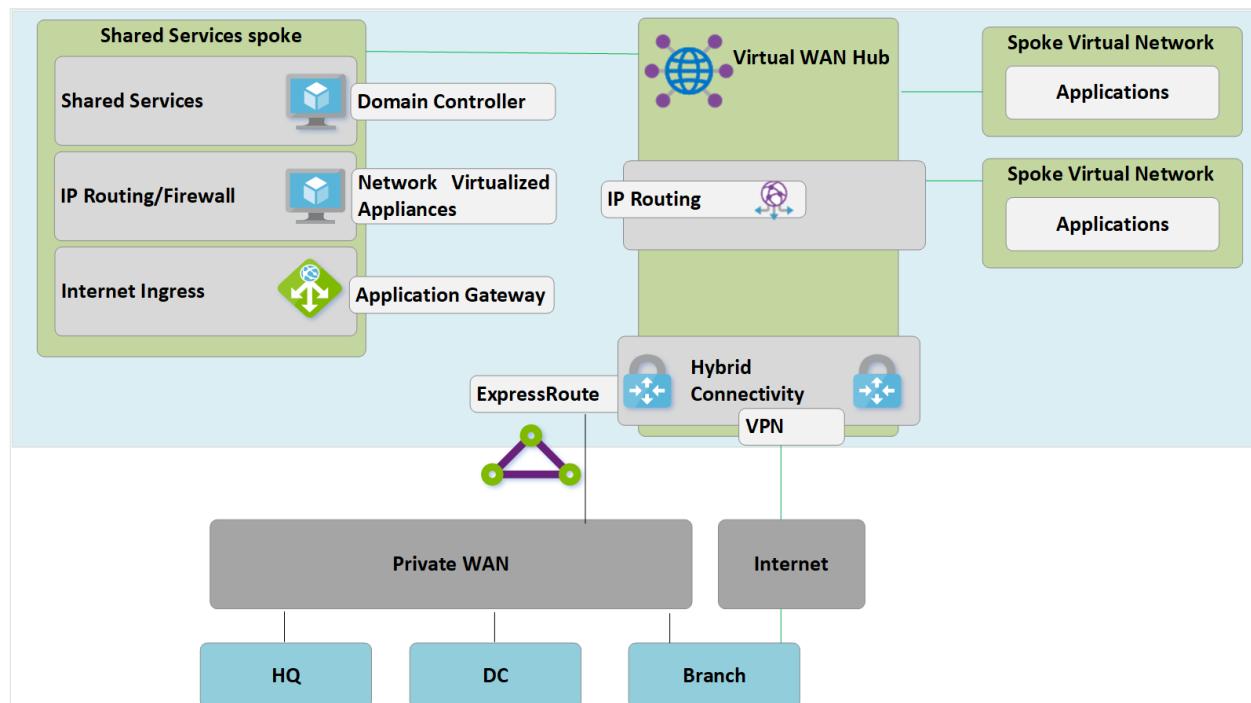


Figure 6: Customer-managed hub-and-spoke to Virtual WAN migration

Because the Virtual WAN hub is a managed entity and doesn't allow deployment of custom resources such as virtual machines, the shared services block now exists as a spoke virtual network and hosts functions such as internet ingress via Azure Application Gateway or network virtualized appliance. Traffic between the shared services environment and backend virtual machines now transits the Virtual WAN-managed hub.

## Step 7: Optimize on-premises connectivity to fully utilize Virtual WAN

At this stage, Contoso has mostly completed their migrations of business applications into the Microsoft Cloud, with only a few legacy applications remaining within the on-premises DC.

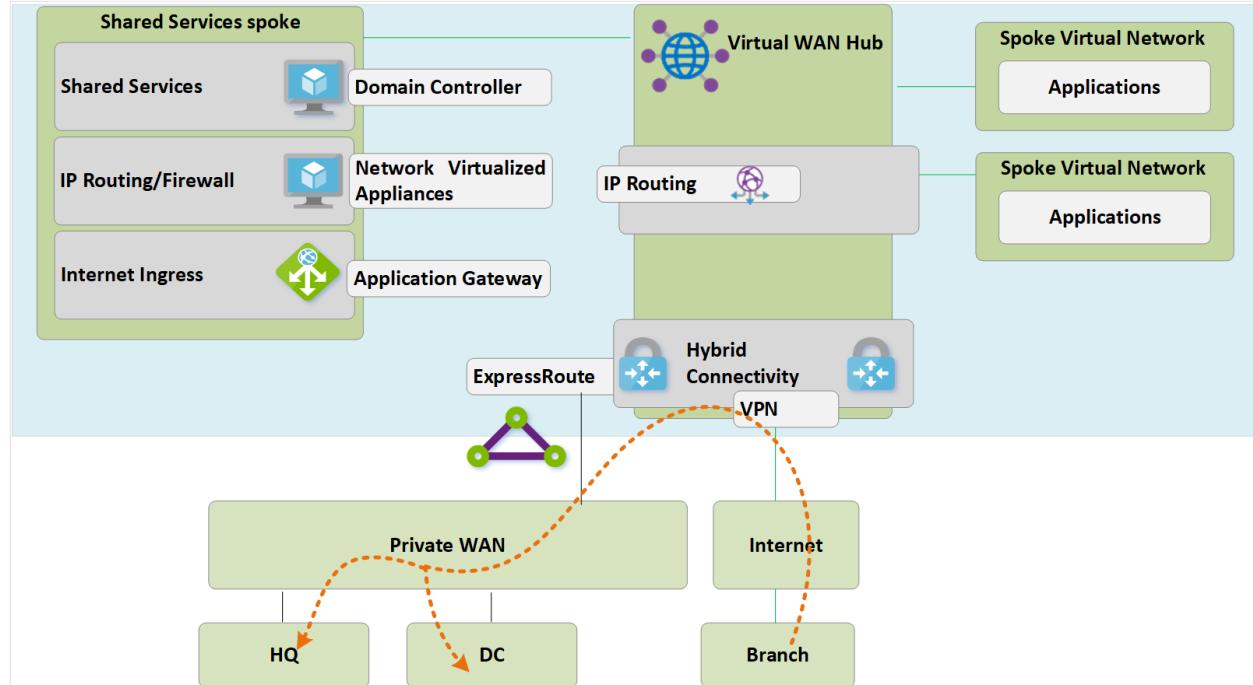


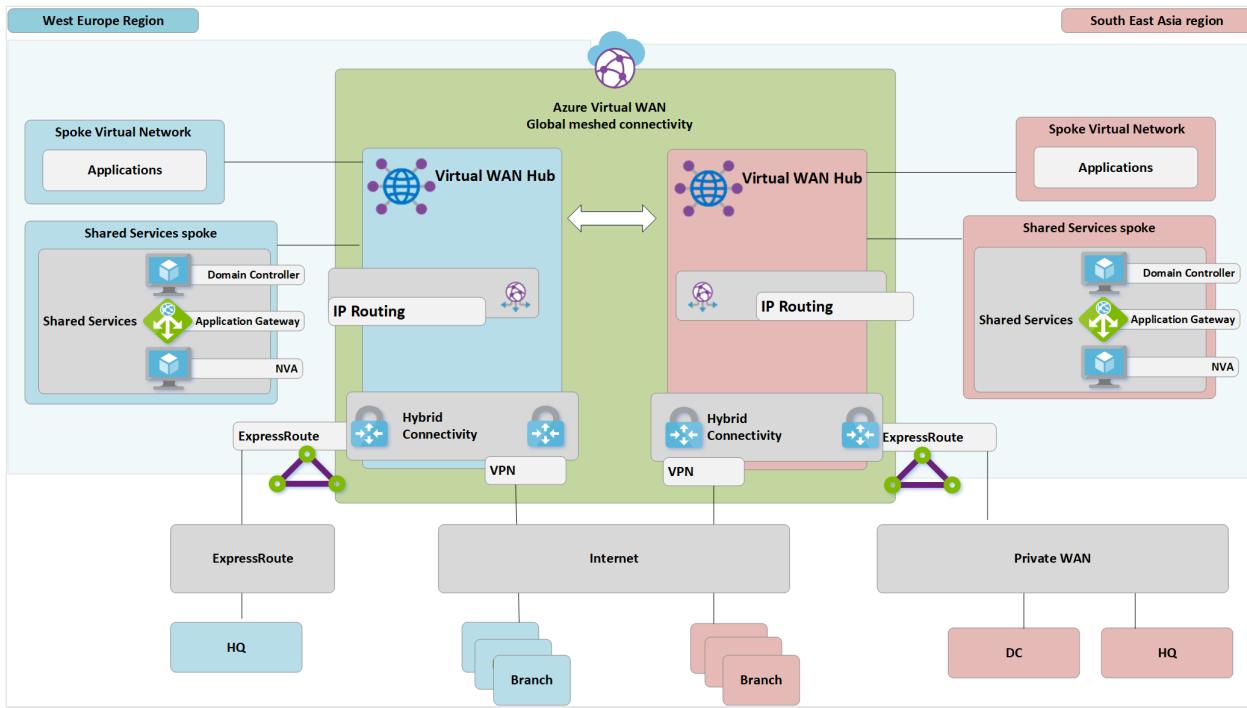
Figure 7: Customer-managed hub-and-spoke to Virtual WAN migration

To leverage the full functionality of Azure Virtual WAN, Contoso decides to decommission their legacy on-premises VPN connections. Any branches continuing to access HQ or DC networks are able to transit the Microsoft global network using the built-in transit routing of Azure Virtual WAN.

### ⓘ Note

ExpressRoute Global Reach is required for customers that want to leverage the Microsoft backbone to provide ExpressRoute to ExpressRoute transit (not shown Figure 7.).

## End-state architecture and traffic paths



**Figure: Dual region Virtual WAN**

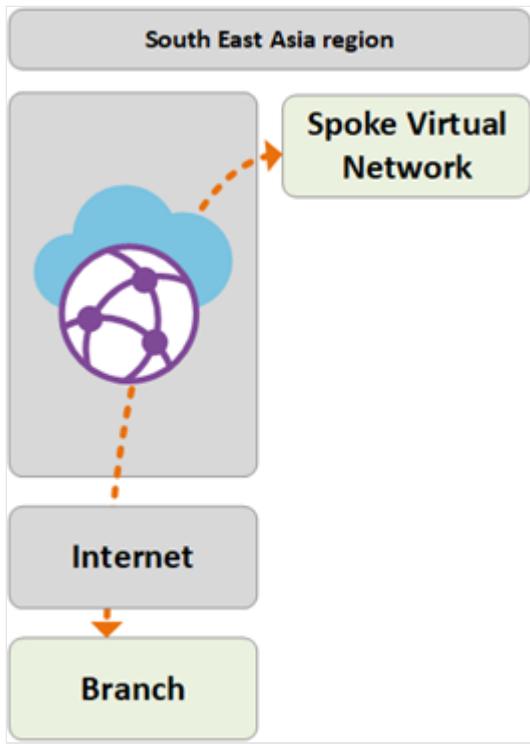
This section provides a summary of how this topology meets the original requirements by looking at some example traffic flows.

## Path 1

Path 1 shows traffic flow from a S2S VPN connected branch in Asia to an Azure VNet in the South East Asia region.

The traffic is routed as follows:

- Asia branch is connected via resilient S2S BGP enabled tunnels into South East Asia Virtual WAN hub.
- Asia Virtual WAN hub routes traffic locally to connected VNet.

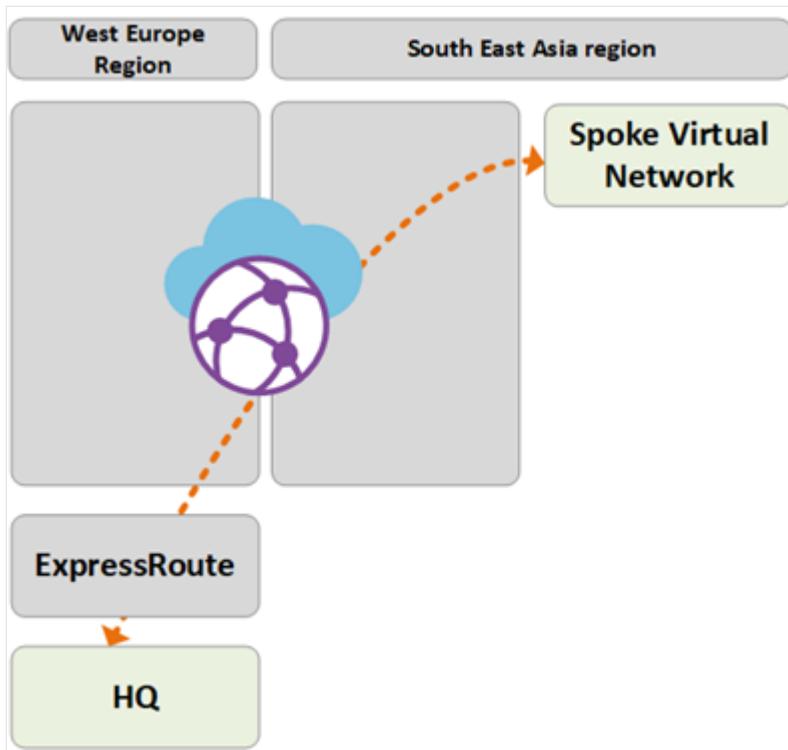


## Path 2

Path 2 shows traffic flow from the ExpressRoute connected European HQ to an Azure VNet in the South East Asia region.

The traffic is routed as follows:

- European HQ is connected via ExpressRoute circuit into West Europe Virtual WAN hub.
- Virtual WAN hub-to-hub global connectivity enables transit of traffic to VNet connected in remote region.

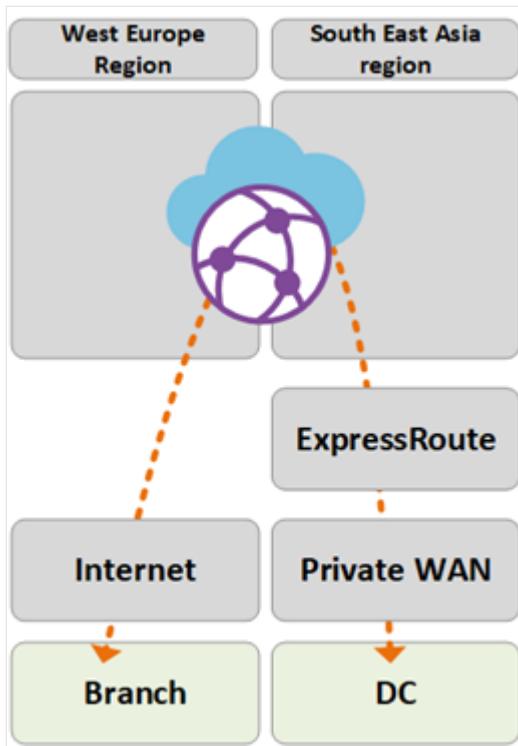


## Path 3

Path 3 shows traffic flow from the Asia on-premises DC connected to Private WAN to a European S2S connected Branch.

The traffic is routed as follows:

- Asia DC is connected to local Private WAN carrier.
- ExpressRoute circuit locally terminates in Private WAN connects to the South East Asia Virtual WAN hub.
- Virtual WAN hub-to-hub global connectivity enables transit of traffic.

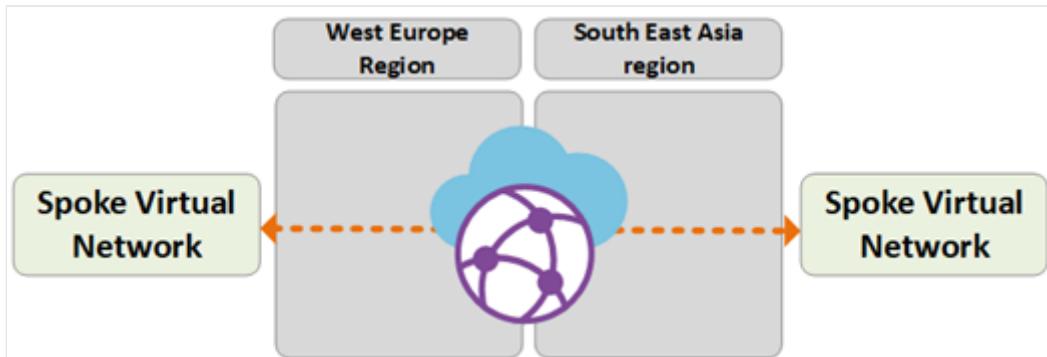


## Path 4

Path 4 shows traffic flow from an Azure VNet in South East Asia region to an Azure VNet in West Europe region.

The traffic is routed as follows:

- Virtual WAN hub-to-hub global connectivity enables native transit of all connected Azure VNets without further user config.

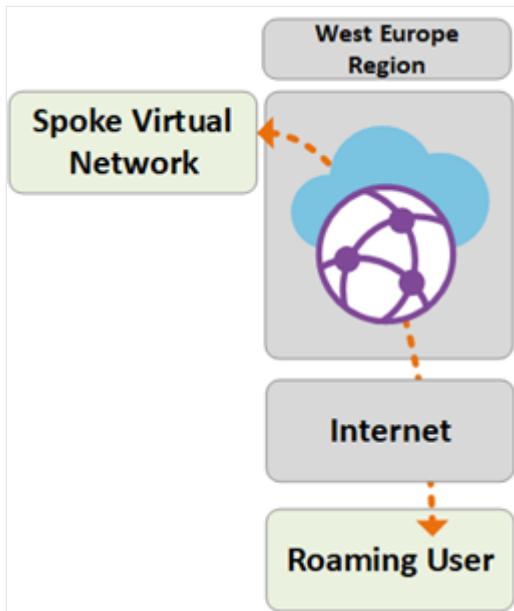


## Path 5

Path 5 shows traffic flow from roaming VPN (P2S) users to an Azure VNet in the West Europe region.

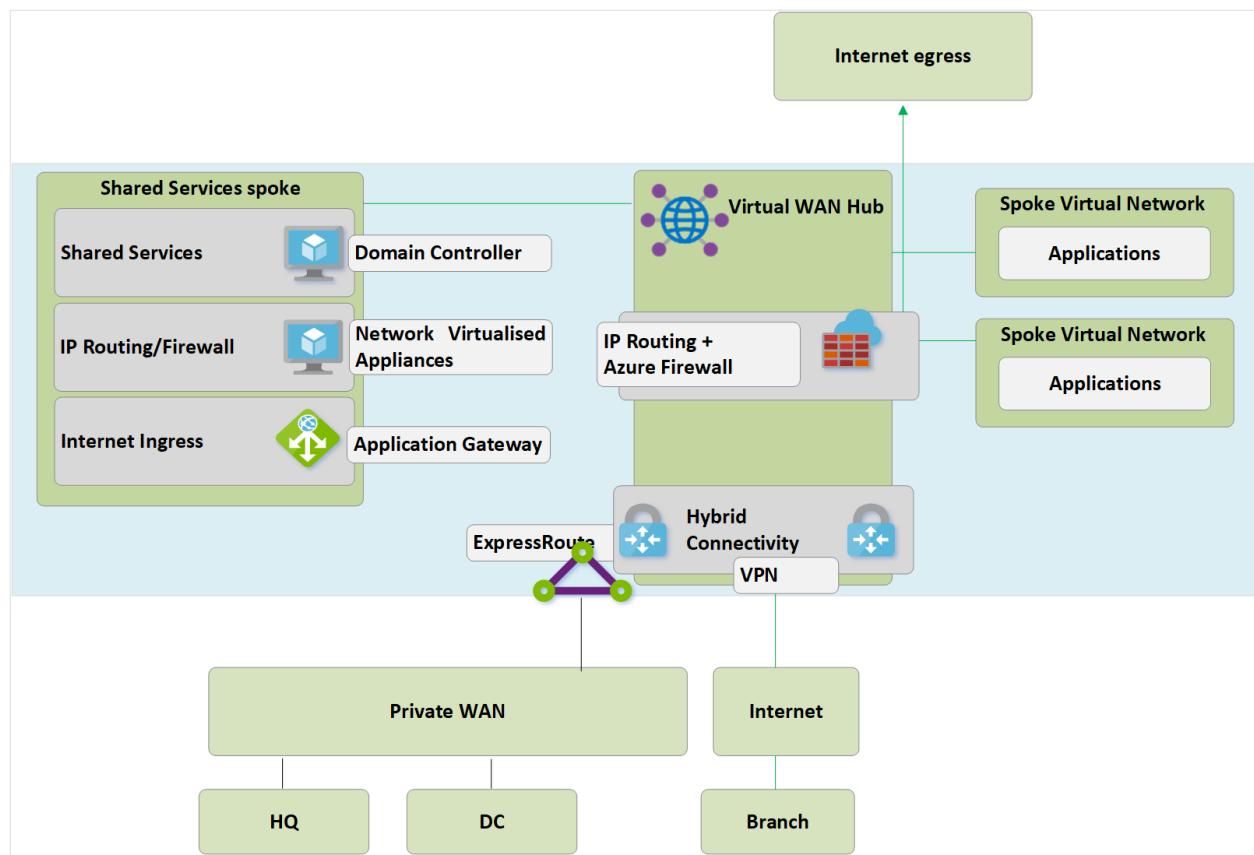
The traffic is routed as follows:

- Laptop and mobile device users use the OpenVPN client for transparent connectivity in to the P2S VPN gateway in West Europe.
- West Europe Virtual WAN hub routes traffic locally to connected VNet.



## Security and policy control via Azure Firewall

Contoso has now validated connectivity between all branches and VNets in line with the requirements discussed earlier in this article. To meet their requirements for security control and network isolation, they need to continue to separate and log traffic via the hub network. Previously this function was performed by a network virtual appliance (NVA). Contoso also wants to decommission their existing proxy services and utilize native Azure services for outbound Internet filtering.



**Figure: Azure Firewall in Virtual WAN (Secured Virtual hub)**

The following high-level steps are required to introduce Azure Firewall into the Virtual WAN hubs to enable a unified point of policy control. For more information about this process and the concept of Secure Virtual Hubs, see [Azure Firewall Manager](#).

1. Create Azure Firewall policy.
2. Link firewall policy to Azure Virtual WAN hub. This step allows the existing Virtual WAN hub to function as a secured virtual hub, and deploys the required Azure Firewall resources.

**① Note**

There are constraints relating to use of secured virtual hubs, including inter-region traffic. For more information, see [Firewall Manager - known issues](#).

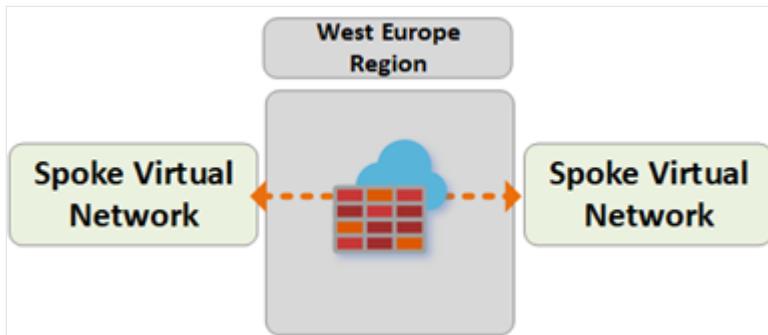
The following paths show the connectivity paths enabled by using Azure secured virtual hubs:

## Path 6

Path 6 shows secure traffic flow between VNets within the same region.

The traffic is routed as follows:

- Virtual Networks connected to the same Secured Virtual Hub now route traffic to via the Azure Firewall.
- Azure Firewall can apply policy to these flows.

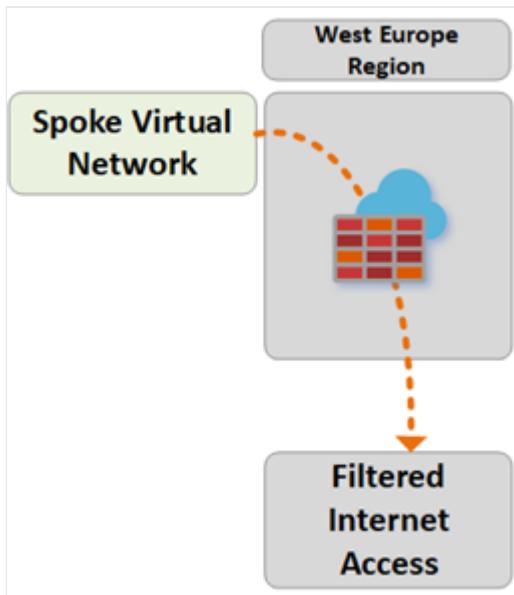


## Path 7

Path 7 shows traffic flow from an Azure VNet to the Internet or third-party Security Service.

The traffic is routed as follows:

- Virtual Networks connected to the Secure Virtual Hub can send traffic to public, destinations on the Internet, using the Secure Hub as a central point of Internet access.
- This traffic can be filtered locally using Azure Firewall FQDN rules, or sent to a third-party security service for inspection.

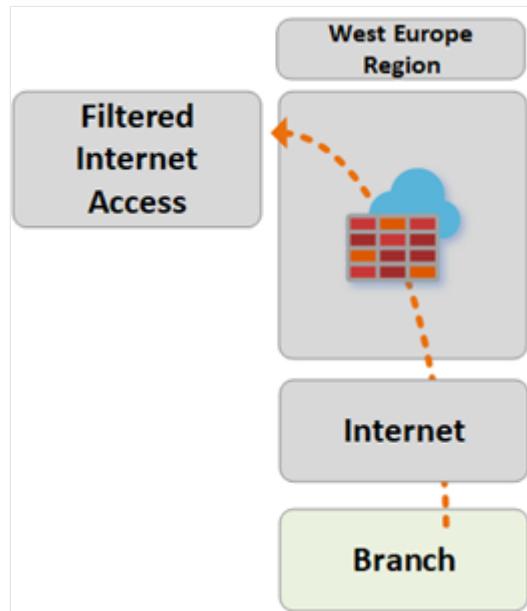


## Path 8

Path 8 shows traffic flow from branch-to-Internet or third-party Security Service.

The traffic is routed as follows:

- Branches connected to the Secure Virtual Hub can send traffic to public destinations on the Internet by using the Secure Hub as a central point of Internet access.
- This traffic can be filtered locally using Azure Firewall FQDN rules, or sent to a third-party security service for inspection.



## Next steps

- Learn more about [Azure Virtual WAN](#).
- [Configure Virtual WAN for Azure NetApp Files](#)

# SD-WAN connectivity architecture with Azure Virtual WAN

Article • 08/24/2023

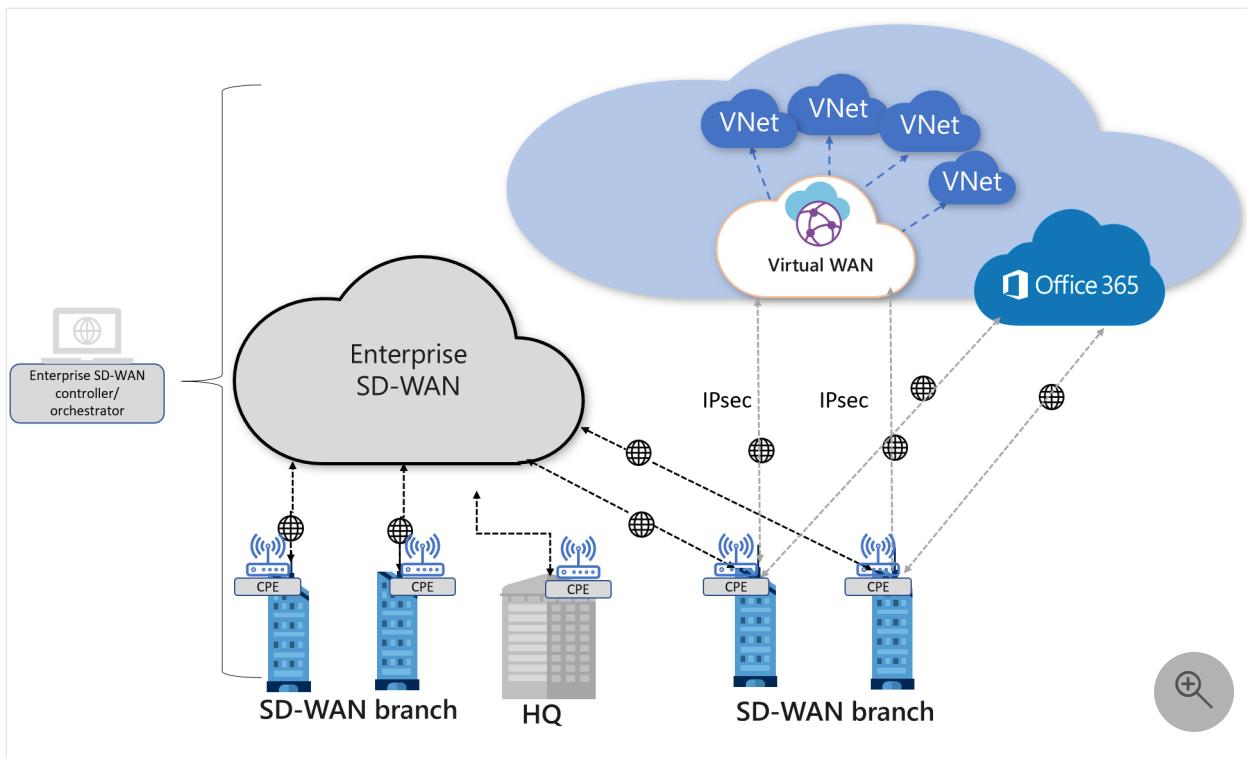
Azure Virtual WAN is a networking service that brings together many cloud connectivity and security services with a single operational interface. These services include branch (via Site-to-site VPN), remote user (Point-to-site VPN), private (ExpressRoute) connectivity, intra-cloud transitive connectivity for VNets, VPN and ExpressRoute interconnectivity, routing, Azure Firewall, and encryption for private connectivity.

Although Azure Virtual WAN is a cloud-based SD-WAN that provides a rich suite of Azure first-party connectivity, routing, and security services, Azure Virtual WAN also is designed to enable seamless interconnection with premises-based SD-WAN and SASE technologies and services. Many such services are offered by our [Virtual WAN](#) ecosystem and Azure Networking Managed Services partners ([MSPs](#)). Enterprises that are transforming their private WAN to SD-WAN have options when interconnecting their private SD-WAN with Azure Virtual WAN. Enterprises can choose from these options:

- Direct Interconnect model
- Direct Interconnect model with NVA-in-VWAN-hub
- Indirect Interconnect model
- Managed Hybrid WAN model using their favorite managed service provider [MSP](#)

In all of these cases, the interconnection of Virtual WAN with SD-WAN is similar from the connectivity side, but may vary on the orchestration and operational side.

## Direct Interconnect model



In this architecture model, the SD-WAN branch customer-premises equipment (CPE) is directly connected to Virtual WAN hubs via IPsec connections. The branch CPE may also be connected to other branches via the private SD-WAN, or use Virtual WAN for branch to branch connectivity. Branches that need to access their workloads in Azure will be able to directly and securely access Azure via the IPsec tunnel(s) that are terminated in the Virtual WAN hub(s).

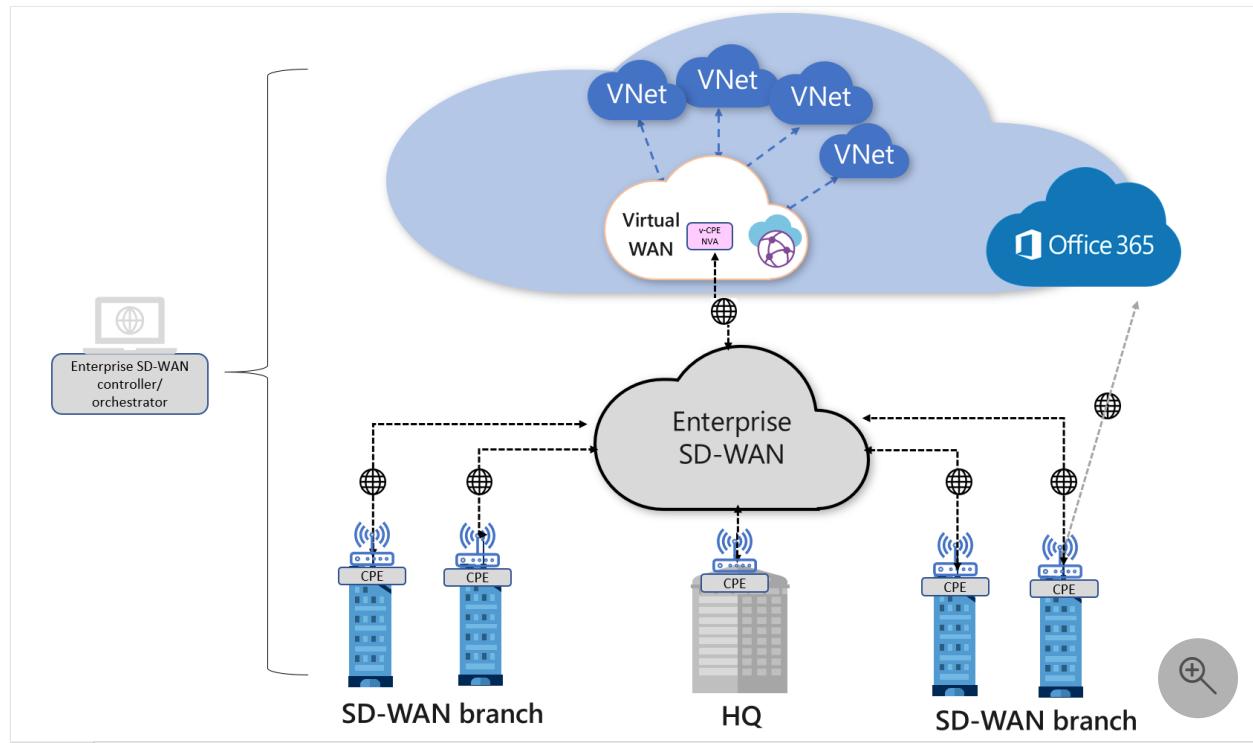
SD-WAN CPE partners can enable automation in order to automate the normally tedious and error-prone IPsec connectivity from their respective CPE devices. Automation allows the SD-WAN controller to talk to Azure via the Virtual WAN API to configure the Virtual WAN sites, and push necessary IPsec tunnel configuration to the branch CPEs. See [Automation guidelines](#) for the description of Virtual WAN interconnection automation by various SD-WAN partners.

The SD-WAN CPE continues to be the place where traffic optimization and path selection is implemented and enforced.

In this model, some vendor proprietary traffic optimization based on real-time traffic characteristics may not be supported because the connectivity to Virtual WAN is over IPsec and the IPsec VPN is terminated on the Virtual WAN VPN gateway. For example, dynamic path selection at the branch CPE is feasible due to the branch device exchanging various network packet information with another SD-WAN node, hence identifying the best link to use for various prioritized traffic dynamically at the branch. This feature may be useful in areas where last mile optimization (branch to the closest Microsoft POP) is required.

With Virtual WAN, users can get Azure Path Selection, which is policy-based path selection across multiple ISP links from the branch CPE to Virtual WAN VPN gateways. Virtual WAN allows for the setup of multiple links (paths) from the same SD-WAN branch CPE; each link represents a dual tunnel connection from a unique public IP of the SD-WAN CPE to two different instances of Azure Virtual WAN VPN gateway. SD-WAN vendors can implement the most optimal path to Azure, based on traffic policies set by their policy engine on the CPE links. On the Azure end, all connections coming in are treated equally.

## Direct Interconnect model with NVA-in-VWAN-hub



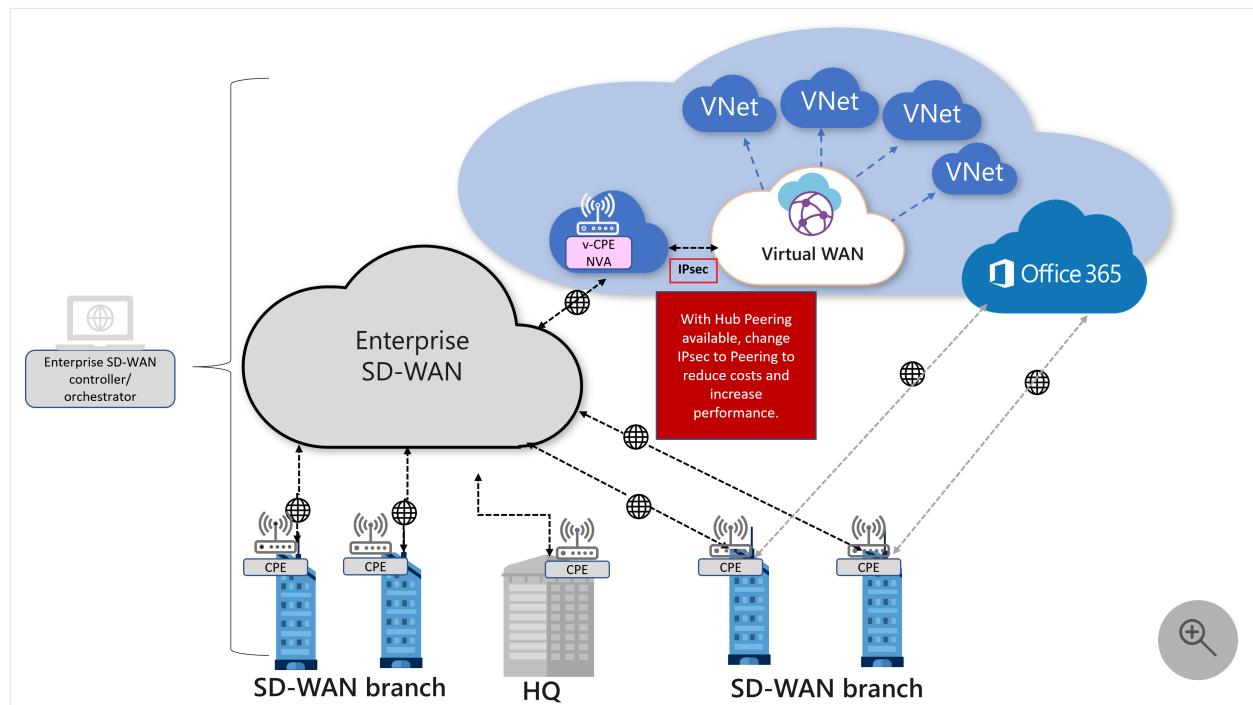
This architecture model supports the deployment of a third-party [Network Virtual Appliance \(NVA\)](#) directly into the virtual hub. This allows customers who want to connect their branch CPE to the same brand NVA in the virtual hub so that they can take advantage of proprietary end-to-end SD-WAN capabilities when connecting to Azure workloads.

Several Virtual WAN Partners have worked to provide an experience that configures the NVA automatically as part of the deployment process. Once the NVA has been provisioned into the virtual hub, any additional configuration that may be required for the NVA must be done via the NVA partners portal or management application. Direct access to the NVA isn't available. The NVAs that are available to be deployed directly into the Azure Virtual WAN hub are engineered specifically to be used in the virtual hub.

For partners that support NVA in VWAN hub and their deployment guides, please see the [Virtual WAN Partners](#) article.

The SD-WAN CPE continues to be the place where traffic optimization and path selection is implemented and enforced. In this model, vendor proprietary traffic optimization based on real-time traffic characteristics is supported because the connectivity to Virtual WAN is via the SD-WAN NVA in the hub.

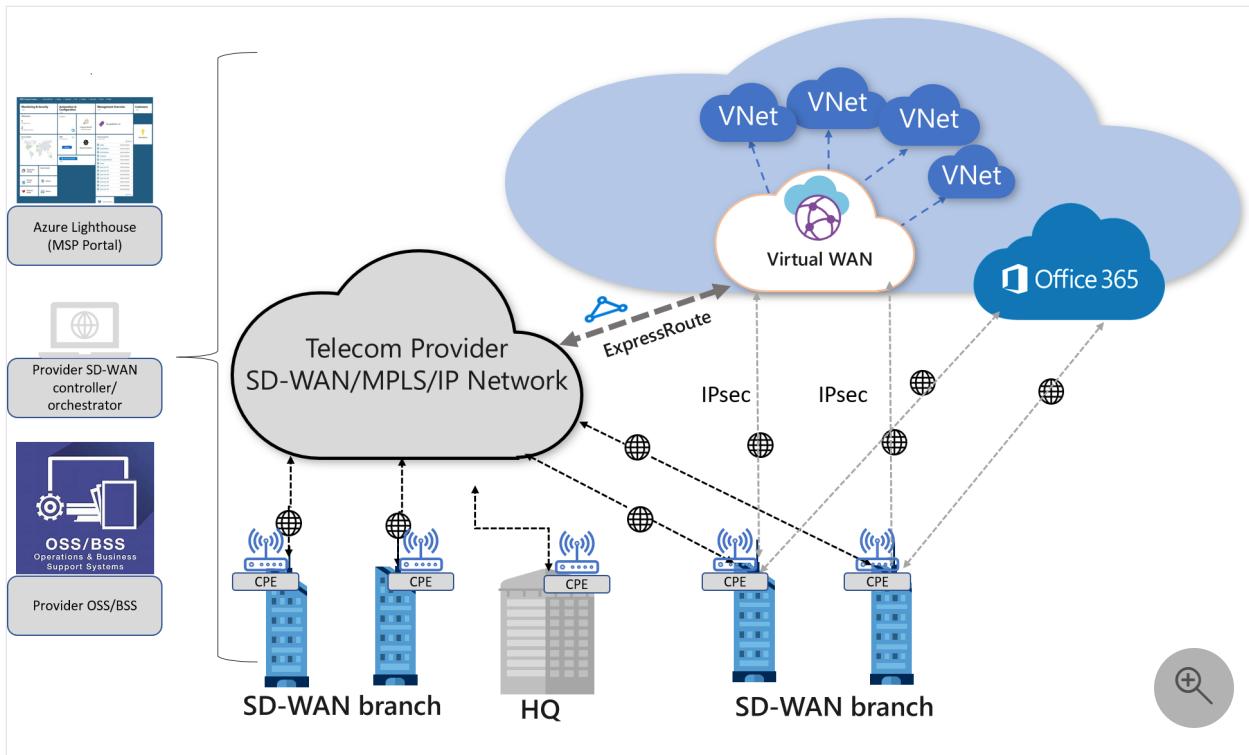
## Indirect Interconnect model



In this architecture model, SD-WAN branch CPEs are indirectly connected to Virtual WAN hubs. As the figure shows, an SD-WAN virtual CPE is deployed in an enterprise VNet. This virtual CPE is, in turn, connected to the Virtual WAN hub(s) using IPsec. The virtual CPE serves as an SD-WAN gateway into Azure. Branches that need to access their workloads in Azure will be able to access them via the v-CPE gateway.

Since the connectivity to Azure is via the v-CPE gateway (NVA), all traffic to and from Azure workload VNets to other SD-WAN branches go via the NVA. In this model, the user is responsible for managing and operating the SD-WAN NVA including high availability, scalability, and routing.

## Managed Hybrid WAN model



In this architecture model, enterprises can leverage a managed SD-WAN service offered by a Managed Service Provider (MSP) partner. This model is similar to the direct or indirect models described above. However, in this model, the SD-WAN design, orchestration, and operations are delivered by the SD-WAN Provider.

Azure Networking MSP partners can use [Azure Lighthouse](#) to implement the SD-WAN and Virtual WAN service in the enterprise customer's Azure subscription, as well as operate the end-to-end hybrid WAN on behalf of the customer. These MSPs may also be able to implement Azure ExpressRoute into the Virtual WAN and operate it as an end-to-end managed service.

## Additional information

- [Azure Virtual WAN FAQ](#)
- [Solving Remote Connectivity](#)

# Virtual WAN network topology

Article • 06/08/2023

Explore key design considerations and recommendations for virtual wide area networks (Virtual WAN) in Microsoft Azure.

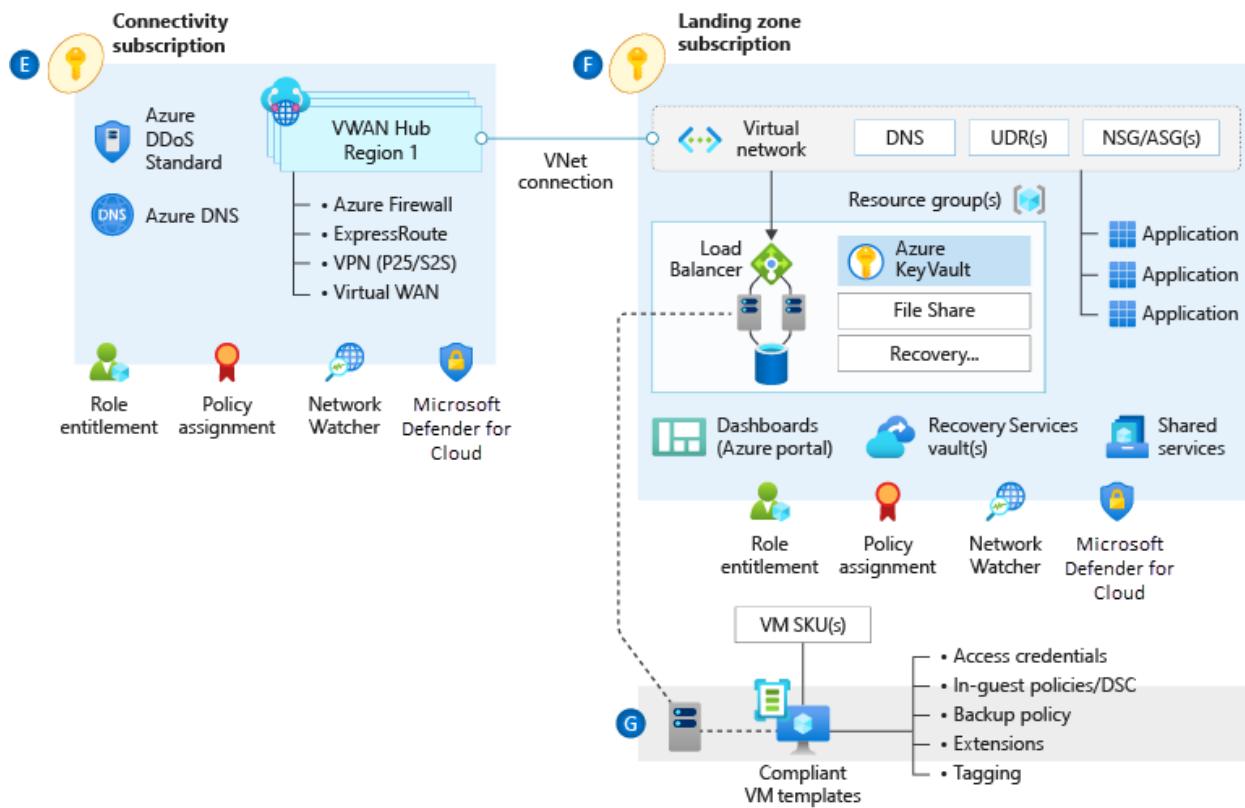
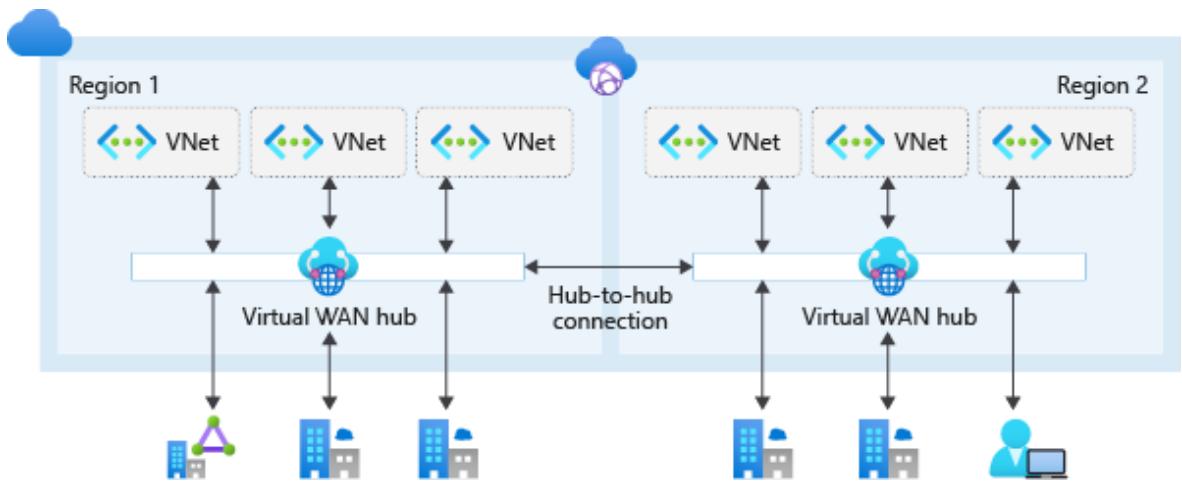


Figure 1: Virtual WAN network topology.

## Virtual WAN network design considerations

Azure Virtual WAN is a Microsoft-managed solution that provides end-to-end, global, and dynamic transit connectivity by default. Virtual WAN hubs eliminate the need to manually configure network connectivity. For example, you don't need to manage user-defined routes (UDR) or network virtual appliances (NVAs) to enable global transit connectivity.

- Azure Virtual WAN simplifies end-to-end network connectivity in Azure, and to Azure from on-premises, by creating a [hub-and-spoke network architecture](#). The architecture easily scales to support multiple Azure regions and on-premises locations (any-to-any connectivity) as shown in the following figure:



*Figure 2: Global transit network with Virtual WAN.*

- Azure Virtual WAN any-to-any transitive connectivity supports the following paths (within the same region and across regions):
  - Virtual network to virtual network
  - Virtual network to branch
  - Branch to virtual network
  - Branch to branch
- Azure Virtual WAN hubs are restricted to the deployment of Microsoft managed resources. The only resources that you can deploy within the WAN hubs are:
  - Virtual network gateways (point-to-site VPN, site-to-site VPN, and Azure ExpressRoute)
  - Azure Firewall via Firewall Manager
  - Route tables
  - Some [network virtual appliances \(NVA\)](#) for vendor-specific SD-WAN capabilities
- Virtual WAN is bound by [Azure subscription limits for Virtual WAN](#).
- Network-to-network transitive connectivity (within a region and across regions via hub-to-hub) is in general availability (GA).
- The Microsoft-managed routing function that's a part of every virtual hub enables the transit connectivity between virtual networks in Standard Virtual WAN. Each hub supports an aggregate throughput of up to 50 Gbps for VNet-to-VNet traffic.
- A single Azure Virtual WAN hub supports a specific maximum number of VM workloads across all directly attached VNets. For more information, see [Azure Virtual WAN limits](#).
- You can deploy multiple Azure Virtual WAN hubs in the same region to scale beyond the single hub limits.

- Virtual WAN integrates with various [SD-WAN providers](#).
- Many managed service providers offer [managed services](#) for Virtual WAN.
- User VPN (point-to-site) gateways in Virtual WAN scale up to 20-Gbps aggregated throughput and 100,000 client connections per virtual hub. For more information, see [Azure Virtual WAN limits](#).
- Site-to-site VPN gateways in Virtual WAN scale up to 20-Gbps aggregated throughput.
- You can connect ExpressRoute circuits to a Virtual WAN hub by using a Local, Standard, or Premium SKU.
- ExpressRoute Standard or Premium circuits, in locations supported by Azure ExpressRoute Global Reach, can connect to a Virtual WAN ExpressRoute gateway. And they have all the Virtual WAN transit capabilities (VPN-to-VPN, VPN, and ExpressRoute transit). ExpressRoute Standard or Premium circuits that are in locations not supported by Global Reach can connect to Azure resources, but can't use Virtual WAN transit capabilities.
- Azure Virtual WAN hub supports ExpressRoute Local if the spoke VNets connected to a Virtual WAN hub are in the same region as the Virtual WAN hub.
- Azure Firewall Manager supports deployment of Azure Firewall in the Virtual WAN hub, known as secured virtual hub. For more information, see the [Azure Firewall Manager overview](#) for secured virtual hubs and the latest [constraints](#).
- Virtual WAN hub-to-hub traffic, by way of Azure Firewall, isn't currently supported when the Azure Firewall deploys inside of the Virtual WAN hub itself (secured virtual hub). Depending on your requirements, you have workarounds. You can place the [Azure Firewall in a spoke virtual network](#), or use NSGs for traffic filtering.
- The Virtual WAN portal experience requires that all Virtual WAN resources deploy together into the same resource group.
- You can share an Azure DDoS Protection Standard plan across all VNets in a single Azure AD tenant to protect resources with public IP addresses. For more information, see [Azure DDoS Protection Standard](#).
  - Virtual WAN secure virtual hubs don't support Azure DDoS standard protection plans. For more information, see [Azure Firewall Manager known issues](#) and [Hub virtual network and secured virtual hub comparison](#).

- Azure DDoS Protection Standard plans only cover resources with public IP addresses.
- An Azure DDoS Protection Standard plan includes 100 public IP addresses. These public IP addresses span all protected VNets associated with the DDoS protection plan. Any other public IP addresses, beyond the 100 included with the plan, are charged separately. For more information on Azure DDoS Protection Standard protection pricing, see the [pricing page](#) or the [FAQ](#).
- Review the [supported resources of Azure DDoS Protection Standard plans](#).

## Virtual WAN network design recommendations

We recommend Virtual WAN for new large or global network deployments in Azure where you need global transit connectivity across Azure regions and on-premises locations. That way, you don't have to manually set up transitive routing for Azure networking.

The following figure shows a sample global enterprise deployment with datacenters spread across Europe and the United States. The deployment contains many branch offices within both regions. The environment is globally connected via Azure Virtual WAN and [ExpressRoute Global Reach](#).

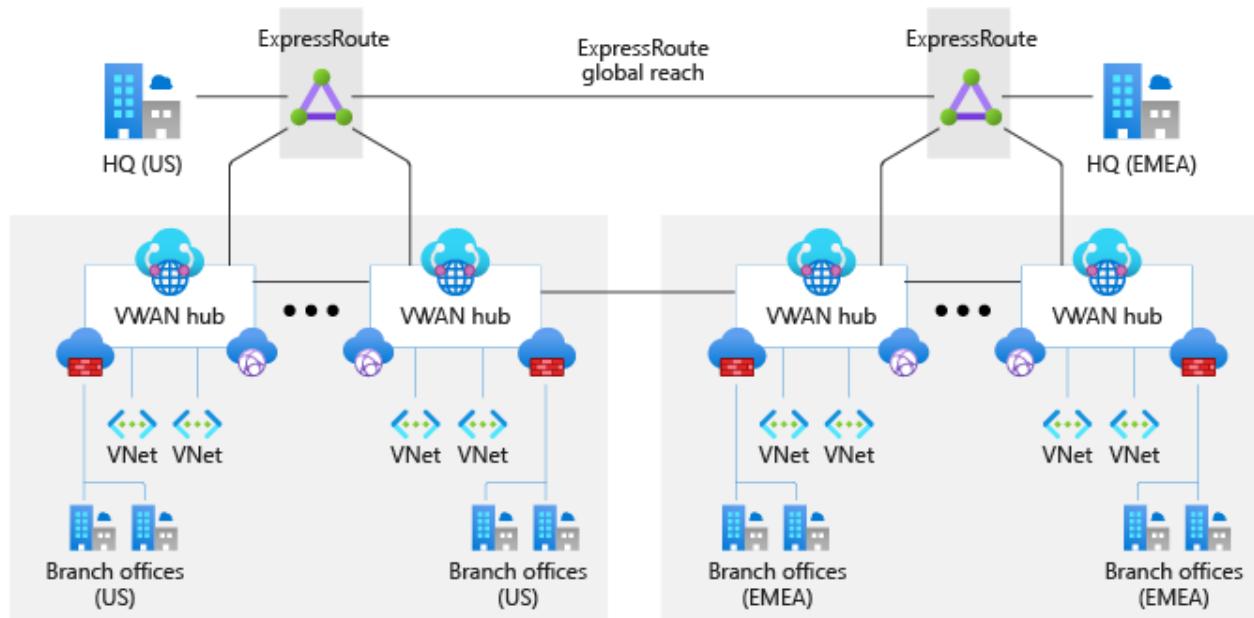


Figure 3: Sample network topology.

- Use a Virtual WAN hub per Azure region to connect multiple landing zones together across Azure regions by way of a common global Azure Virtual WAN.
- Deploy all Virtual WAN resources into a single resource group in the connectivity subscription, including when you're deploying across multiple regions.

- Use [virtual hub routing](#) features to further segment traffic between VNets and branches.
- Connect Virtual WAN hubs to on-premises datacenters by using ExpressRoute.
- Deploy required shared services, like DNS servers, in a dedicated spoke virtual network. Customer deployed shared resources can't be deployed inside the Virtual WAN hub itself.
- Connect branches and remote locations to the nearest Virtual WAN hub via Site-to-Site VPN, or enable branch connectivity to Virtual WAN via an SD-WAN partner solution.
- Connect users to the Virtual WAN hub via a Point-to-Site VPN.
- Follow the principle of "traffic in Azure stays in Azure" so that communication across resources in Azure occurs via the Microsoft backbone network, even when the resources are in different regions.
- For internet outbound protection and filtering, consider deploying Azure Firewall in the virtual hub.
- [Security provided by NVA firewalls](#). Customers can also deploy NVAs into a Virtual WAN hub that perform both SD-WAN connectivity and Next-Generation Firewall capabilities. Customers can connect on-premises devices to the NVA in the hub and also use the same appliance to inspect all North-South, East-West, and Internet-bound traffic.
- When you're deploying partner networking technologies and NVAs, follow the partner vendor's guidance to ensure there are no conflicting configurations with Azure networking.
- For brownfield scenarios where you're migrating from a hub-and-spoke network topology not based on Virtual WAN, see [Migrate to Azure Virtual WAN](#).
- Create Azure Virtual WAN and Azure Firewall resources within the connectivity subscription.
- Don't create more than 500 virtual network connections per Virtual WAN virtual hub.
  - If you need more than 500 virtual network connections per Virtual WAN virtual hub, you can deploy another Virtual WAN virtual hub. Deploy it in the same region as part of the same Virtual WAN and resource group.

- Plan your deployment carefully, and ensure that your network architecture is within the [Azure Virtual WAN limits](#).
- Use [insights in Azure Monitor for Virtual WAN \(preview\)](#) to monitor the end-to-end topology of your Virtual WAN and status and [key metrics](#).
- Deploy a single Azure DDoS standard protection plan in the connectivity subscription.
  - All landing zone and platform VNets should use this plan.

# Virtual WAN architecture optimized for department-specific requirements

Azure Virtual WAN

Azure ExpressRoute

Azure Virtual Network

A global manufacturing company provided the architecture that this article describes. The company's operational technology and information technology departments are highly integrated, demanding a single internal network. But the environments have drastically different security and performance requirements. Because of the sensitive nature of the company's operations, all traffic needs to be firewall protected, and an Intrusion Detection and Protection System (IDPS) solution needs to be in place. The information technology department has less demanding security requirements for the network, but that department wants to optimize for performance so users have low-latency access to IT applications.

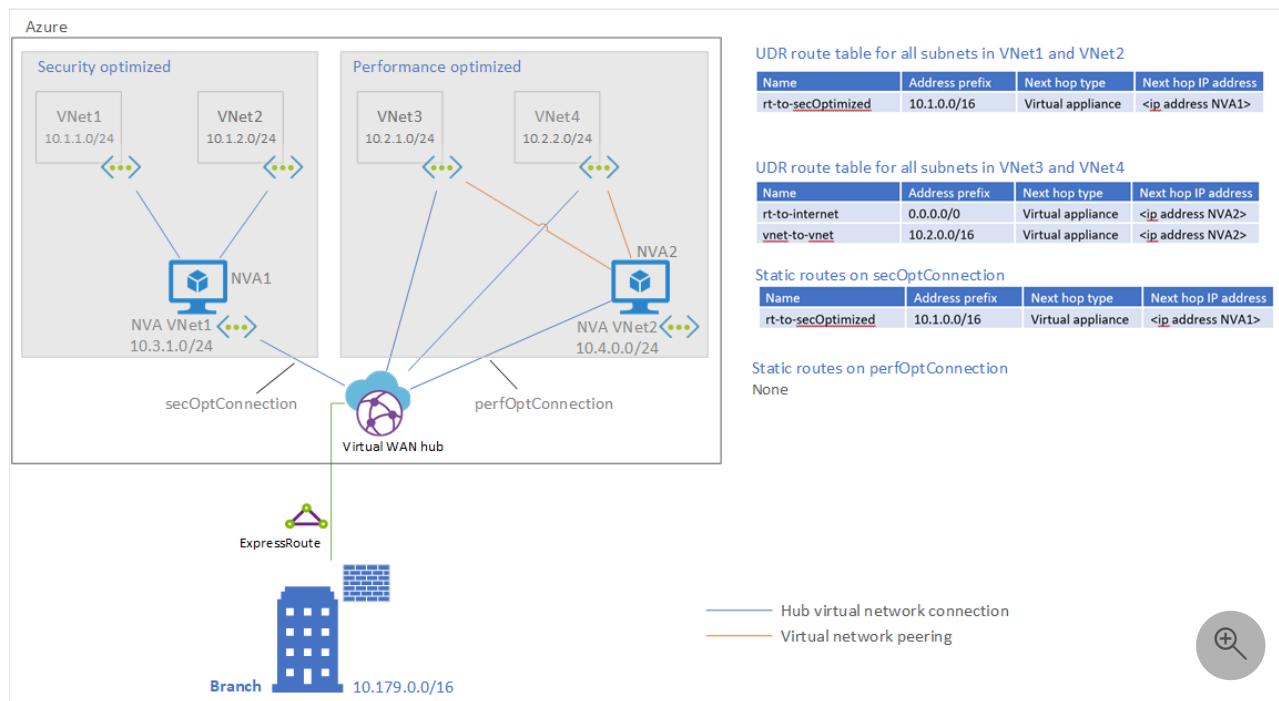
Decision makers at the company turned to Azure Virtual WAN to meet global needs for a single network with varying security and performance requirements. They also got a solution that's easy to manage, deploy, and scale. As they add regions, they can continue to grow seamlessly with a network that's highly optimized for their needs.

## Potential use cases

Typical use cases for this architecture include:

- A global organization that requires a centralized file solution for business-critical work.
- High-performing file workloads that require localized cached files.
- A flexible remote workforce for users both in and out of the office.

## Architecture



Download a [Visio file](#) of this architecture.

Here's a summary of the architecture:

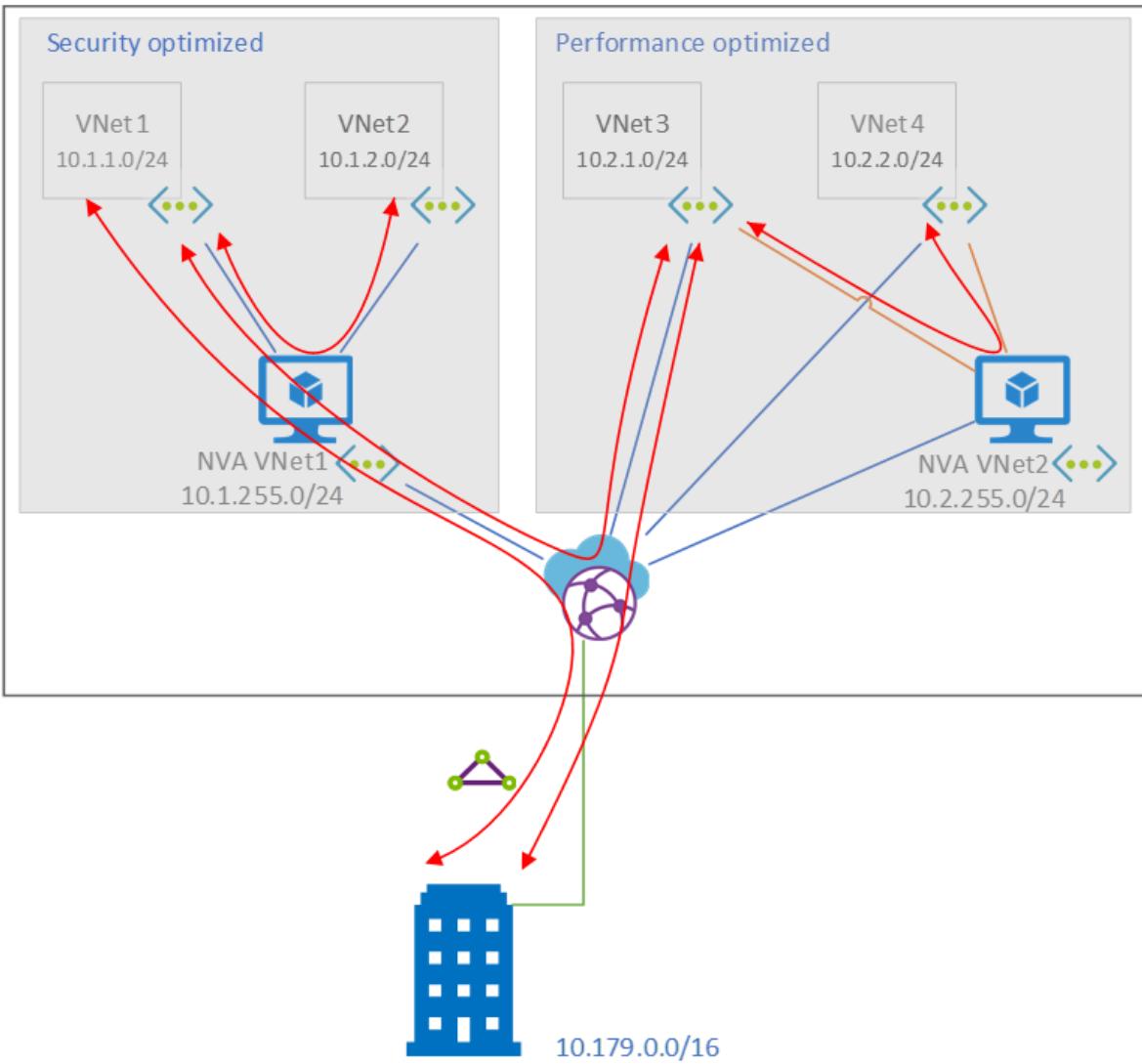
- Users access virtual networks from a branch.
- Azure ExpressRoute extends the on-premises networks into the Microsoft cloud over a private connection, with the help of a connectivity provider.
- A Virtual WAN hub routes traffic appropriately for security or performance. The hub contains various service endpoints to enable connectivity.
- User-defined routes force traffic to the NVAs when necessary.
- Each NVA inspects traffic that's flowing into a virtual network.
- Virtual network peering provides VNet-to-VNet inspection in the performance-optimized environment.

The company has multiple regions and continues to deploy regions to the model. The company deploys a security-optimized or performance-optimized environment only when needed. The environments route the following traffic through the network virtual appliance (NVA):

## Traffic pathways

Expand table

		Destinations					
		VNet1	VNet2	VNet3	VNet4	Branch	Internet
Security-optimized source	VNet1	Intra VNet	NVA1-VNet2	NVA1-hub-VNet3	NVA1-hub-VNet4	NVA1-hub-branch	NVA1-internet
Performance-optimized source	VNet3	hub-NVA1-VNet1	hub-NVA1-VNet2	Intra VNet	NVA2-VNet4	hub-branch	NVA2-internet
Branch source	Branch	hub-NVA1-VNet1	hub-NVA1-VNet2	hub-VNet3	hub-VNet4	Not applicable	Not applicable



As the preceding diagram shows, an NVA and routing architecture force all traffic pathways in the security-optimized environment to use the NVA between the virtual networks and the hub in a common layered architecture.

The performance-optimized environment has a more customized routing schema. This schema provides a firewall and traffic inspection where they're needed. It doesn't provide a firewall where it's not needed. VNet-to-VNet traffic in the performance-optimized space is forced through NVA2, but branch-to-VNet traffic can go directly across the hub. Likewise, anything headed to the secure environment doesn't need to go to NVA VNet2 because it's inspected at the edge of the secure environment by the NVA in NVA VNet1. The result is high-speed access to the branch. The architecture still provides VNet-to-VNet inspection in the performance-optimized environment. This isn't necessary for all customers but can be accomplished through the peerings that you can see in the architecture.

## Associations and propagations of the Virtual WAN hub

Configure routes for the Virtual WAN hub as follows:

[Expand table](#)

Name	Associated to	Propagating to
NVA VNet1	defaultRouteTable	defaultRouteTable
NVA VNet2	PerfOptimizedRouteTable	defaultRouteTable
VNet3	PerfOptimizedRouteTable	defaultRouteTable
VNet4	PerfOptimizedRouteTable	defaultRouteTable

## Routing requirements

- A custom route on the default route table in the Virtual WAN hub to route all traffic for VNet1 and VNet2 to secOptConnection.

[Expand table](#)

Route name	Destination type	Destination prefix	Next hop	Next hop IP
Security optimized route	CIDR	10.1.0.0/16	secOptConnection	<IP address of NVA1>

- A static route on the secOptConnection forwarding the traffic for VNet1 and VNet2 to the IP address of NVA1.

[Expand table](#)

Name	Address prefix	Next hop type	Next hop IP address
rt-to-secOptimized	10.1.0.0/16	Virtual appliance	<IP address of NVA1>

- A custom route table on the Virtual WAN hub that's named perfOptimizedRouteTable. This table is used to ensure the performance-optimized virtual networks can't communicate with each other over the hub and must use the peering to NVA VNet2.
- A UDR associated with all subnets in VNet1 and VNet2 to route all traffic back to NVA1.

[Expand table](#)

Name	Address prefix	Next hop type	Next hop IP address
rt-all	0.0.0.0/0	Virtual appliance	<IP address of NVA1>

- A UDR associated with all subnets in VNet3 and VNet4 to route VNet-to-VNet traffic and internet traffic to NVA2.

[\[+\] Expand table](#)

Name	Address prefix	Next hop type	Next hop IP address
rt-to-internet	0.0.0.0/0	Virtual appliance	<IP of address NVA2>
vnet-to-vnet	10.2.0.0/16	Virtual appliance	<IP address of NVA2>

#### ⚠ Note

You can replace NVA IP addresses with load balancer IP addresses in the routing if you're deploying a high-availability architecture with multiple NVAs behind the load balancer.

## Components

- [Azure Virtual WAN](#). Virtual WAN is a networking service that brings many networking, security, and routing functionalities together to provide a single operational interface. In this case, it simplifies and scales routing to the attached virtual networks and branches.
- [Azure ExpressRoute](#). ExpressRoute extends on-premises networks into the Microsoft cloud over a private connection.
- [Azure Virtual Network](#). Virtual Network is the fundamental building block for your private network in Azure. Virtual Network enables many types of Azure resources, like Azure virtual machines (VMs), to communicate with improved security with each other, the internet, and on-premises networks.
- [Virtual WAN hub](#). A virtual hub is a virtual network that Microsoft manages. The hub contains various service endpoints to enable connectivity.
- [Hub virtual network connections](#). The hub virtual network connection resource connects the hub seamlessly to your virtual networks.
- [Static routes](#). Static routes provide a mechanism for steering traffic through a next hop IP.
- [Hub route tables](#). You can create a virtual hub route and apply the route to the virtual hub route table.
- [Virtual network peering](#). By using virtual network peering, you can seamlessly connect two or more [virtual networks](#) in Azure.
- [User-defined routes](#). User-defined routes are static routes that override the default Azure system routes or add more routes to a subnet's route table. They're used here to force traffic to the NVAs when necessary.

- Network virtual appliances [2](#). Network virtual appliances are marketplace-offered network appliances. In this case, the company deployed Palo Alto's NVA, but any NVA firewall would work here.

## Alternatives

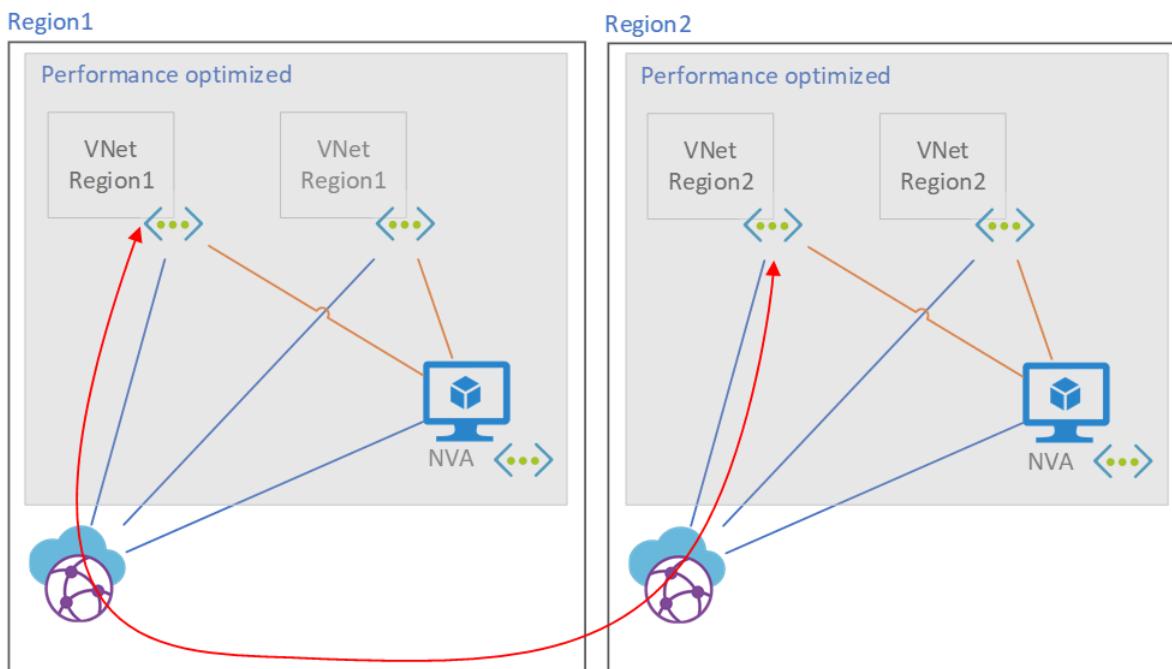
To deploy only a high-security NVA environment, you can follow this model: [Route traffic through an NVA](#).

To deploy a custom NVA model that supports both routing traffic to a dedicated firewall for the internet and routing branch traffic over an NVA, see [Route traffic through NVAs by using custom settings](#).

The previous alternative deploys a high-security environment behind an NVA and offers some capability to deploy a custom environment. But it differs from the use case described in this article in two ways. First, it shows the two models in isolation instead of in combination. Second, it doesn't support VNet-to-VNet traffic in the custom environment (what we call the *Performance-optimized environment* here).

## Considerations

In this deployment, routes that cross the Virtual WAN hub to a performance-optimized environment don't pass through the NVA in that environment. This presents a potential problem with cross-regional traffic that's illustrated here:



Traffic across regions between performance-optimized environments doesn't cross the NVA. This is a limitation of directly routing hub traffic to the virtual networks.

## Availability

Virtual WAN is a highly available networking service. You can set up more connectivity or paths from the branch to get multiple pathways to the Virtual WAN service. But you don't need anything additional within the VWAN service.

You should set up NVAs in a highly available architecture similar to the one described here: [Deploy highly available NVAs](#).

## Performance

This solution optimizes performance of the network where necessary. You can modify the routing according to your own requirements, enabling the traffic to the branch to cross the NVA and the traffic between virtual networks to flow freely or to use a single firewall for internet egress.

## Scalability

This architecture is scalable across regions. Consider your requirements when you set up routing labels for grouping routes and branch traffic forwarding between the virtual hubs.

## Security

With NVAs, you can use features like IDPS with Virtual WAN.

## Resiliency

For information about resiliency, see [Availability](#), earlier in this article.

## Cost optimization

Pricing for this architecture depends heavily on the NVAs that you deploy. For a 2-Gbps ER connection and a Virtual WAN hub that processes 10 TB per month, see this [pricing estimate](#).

## Contributors

*This article is maintained by Microsoft. It was originally written by the following contributors.*

Principal author:

- [John Poetzinger](#) | Senior Cloud Solution Architect

## Next steps

- [What is Azure Virtual WAN?](#)
- [What is Azure ExpressRoute?](#)
- [How to configure virtual hub routing - Azure Virtual WAN](#)
- [Firewall and Application Gateway for virtual networks](#)
- [Azure Virtual WAN and supporting remote work](#)

## Related resources

- [Hub-spoke network topology with Azure Virtual WAN](#)

- Choose between virtual network peering and VPN gateways
- Low-latency network connections for industry

# Low-latency network connections for industry

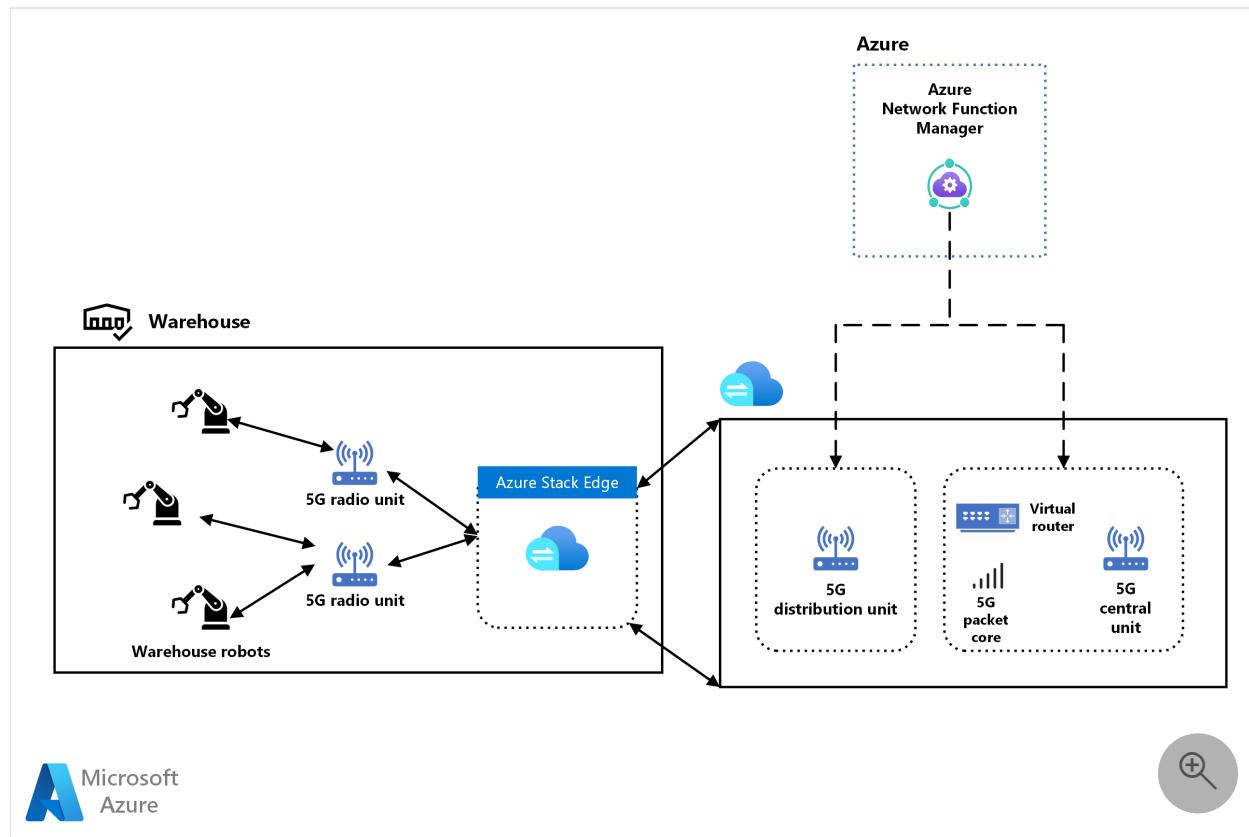
Azure Stack Edge

## 💡 Solution ideas

This article is a solution idea. If you'd like us to expand the content with more information, such as potential use cases, alternative services, implementation considerations, or pricing guidance, let us know by providing [GitHub feedback](#).

This solution describes how manufacturers can use high-performance, low-latency 5G Standalone networks to scale up industrial automation and productivity.

## Architecture



Download a [Visio file](#) of this architecture.

## Dataflow

1. Embedded 5G-enabled internet protocol (IP) modules connect warehouse robots to 5G Open Radio Access Network (ORAN) radio units. RAN is a common wireless network infrastructure for mobile networks.
2. The 5G radio units connect over a wired switching network to the 5G distribution unit software, which runs on private multi-access edge compute (MEC) on Azure Stack Edge.
3. The 5G distribution unit connects with the virtual router, 5G central unit, and 5G packet core, which run on a separate Azure private MEC instance on Azure Stack Edge.
4. The 5G packet core provides device authentication, an IP address, and connectivity based on a preconfigured profile.
5. Azure Network Function Manager controls both MEC instances.
6. Optimizing the 5G network and keeping traffic confined to Azure private MEC provides the low latency these connection scenarios require.

## Components

This solution uses the following Azure components:

- [Azure Stack Edge](#) is a portfolio of devices that bring compute, storage, and intelligence to the IoT Edge.
- [Azure Network Function Manager](#) enables the deployment of network functions to the IoT Edge using consistent Azure tools and interfaces.

## Scenario details

5G Standalone networks reliably connect machines to other machines or controllers. 5G-enabled Internet of Things (IoT) devices like robots can communicate and operate autonomously on the factory or warehouse floors.

The 5G Standalone network is deployed as a Non-Public Network (NPN), with all data remaining on premises. The NPN configuration offers security, privacy, and reliability. Azure deploys and manages the network and devices.

## Potential use cases

This solution is ideal for the manufacturing, agriculture, energy, facilities, telecommunications, and robotics industries. Use this approach for scenarios like:

- Picking shipments efficiently in a warehouse.

- Dispersing seeds from an autonomous seed sprayer machines on a farm, based on information from soil sensors.
- Conserving energy in commercial buildings by shutting off lights when no motion is detected in a room.

## Contributors

*This article is maintained by Microsoft. It was originally written by the following contributors.*

Principal author:

- [Nikhil Ravi](#) | Product Management Leader

## Next steps

- [Azure Private MEC](#)
- [Azure Industrial IoT](#)
- [Azure Network Function Manager simplifies 5G deployments \(Video\)](#)

## Related resources

- [Connect an on-premises network to Azure](#)
- [Condition monitoring for Industrial IoT](#)

# Video capture and analytics for retail

Azure IoT Edge

Azure IoT Hub

Azure Media Services

Azure Stack Edge

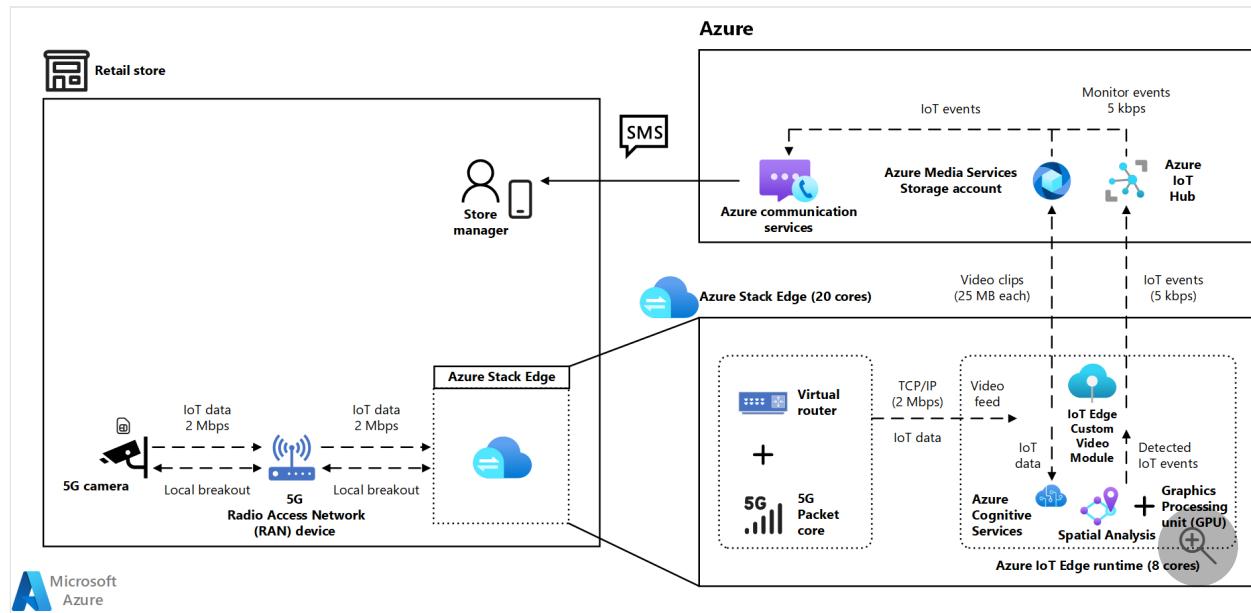
Azure App Service

## 💡 Solution ideas

This article is a solution idea. If you'd like us to expand the content with more information, such as potential use cases, alternative services, implementation considerations, or pricing guidance, let us know by providing [GitHub feedback](#).

This solution describes how retailers like grocery stores can monitor storefront events and take immediate actions to improve customer experience. In this solution, 5G-enabled internet protocol (IP) cameras capture real-time video of shelf inventory, curbside pickup, and cashier queues.

## Architecture



Download a [Visio file](#) of this architecture.

## Dataflow

1. 5G-enabled IP cameras capture video in real time, and send the video feed to a 5G Radio Access Network (RAN) device.

2. The 5G radios in the stores forward the data to the 5G packet core running on the Azure Stack Edge IoT Edge server.
3. The packet core authenticates the devices, applies Quality of Service (QoS) policies, and routes the video traffic to the target application.
4. The custom IoT Edge module also runs on the edge server, which provides the low latency necessary for transporting and processing the video feeds.
5. The custom module simplifies setting up a video streaming pipeline and pre-processing the video for spatial analysis.
6. The [Spatial Analysis](#) module on the edge server anonymously counts cars, goods on a shelf, or people in line. The module sends these event notifications to the Azure IoT Hub module in the cloud.
7. The IoT Hub module records the event notifications in a web app, and alerts store managers or stock keepers if certain thresholds are passed.
8. An Azure Media Services Storage account stores events for long-term trend analysis to help with resource planning.

## Components

This solution uses the following Azure components:

- [Azure Stack Edge](#) is a portfolio of devices that bring compute, storage, and intelligence to the IoT Edge. Azure Stack Edge acts as a cloud storage gateway that enables data transfers to Azure, while retaining local access to files.
- [Web Apps in Microsoft Azure App Service](#) creates and deploys mission-critical web applications that scale with your business.
- [Azure IoT Hub](#) is a cloud-based managed service for bidirectional communication between IoT devices and Azure.
- [Media Services Storage](#) uses Azure Storage to store large media files.
- [Azure Network Function Manager](#) enables the deployment of network functions to the IoT Edge using consistent Azure tools and interfaces.

## Scenario details

This solution describes how retailers like grocery stores can monitor storefront events and take immediate actions to improve customer experience. In this solution, 5G-enabled internet protocol (IP) cameras capture real-time video of shelf inventory, curbside pickup, and cashier queues. On-premises IoT Edge devices analyze the video

data in real time to detect the number of people in checkout queues, empty shelf space, or cars in the parking lot.

Metrics analysis can trigger anomaly events to alert the store manager or stock supervisors to take corrective actions. The solution stores summary video clips or events in the cloud for long-term trend analysis.

## Potential use cases

This solution is ideal for the retail, automotive, and facilities/real-estate industries. This approach includes the following scenarios:

- Monitor and maintain occupancy limits in an establishment.
- Stop unauthorized users from tailgating others into an office building.
- Prevent fraud at grocery store self-checkout stations.

## Contributors

*This article is maintained by Microsoft. It was originally written by the following contributors.*

Principal author:

- [Nikhil Ravi](#) | Product Management Leader

## Next steps

- [What is the Radio Access Network?](#)
- [Live Video Analytics on IoT Edge](#)
- [Azure Network Function Manager simplifies 5G deployments \(Video\)](#)
- [Introduction to Azure IoT Hub](#)
- [Introduction to Azure Stack](#)

## Related resources

- [IoT event routing](#)
- [Contactless IoT interfaces with Azure intelligent edge](#)

# IoT device connectivity for healthcare facilities

Azure Arc

Azure IoT Edge

Azure IoT Hub

Azure Sphere

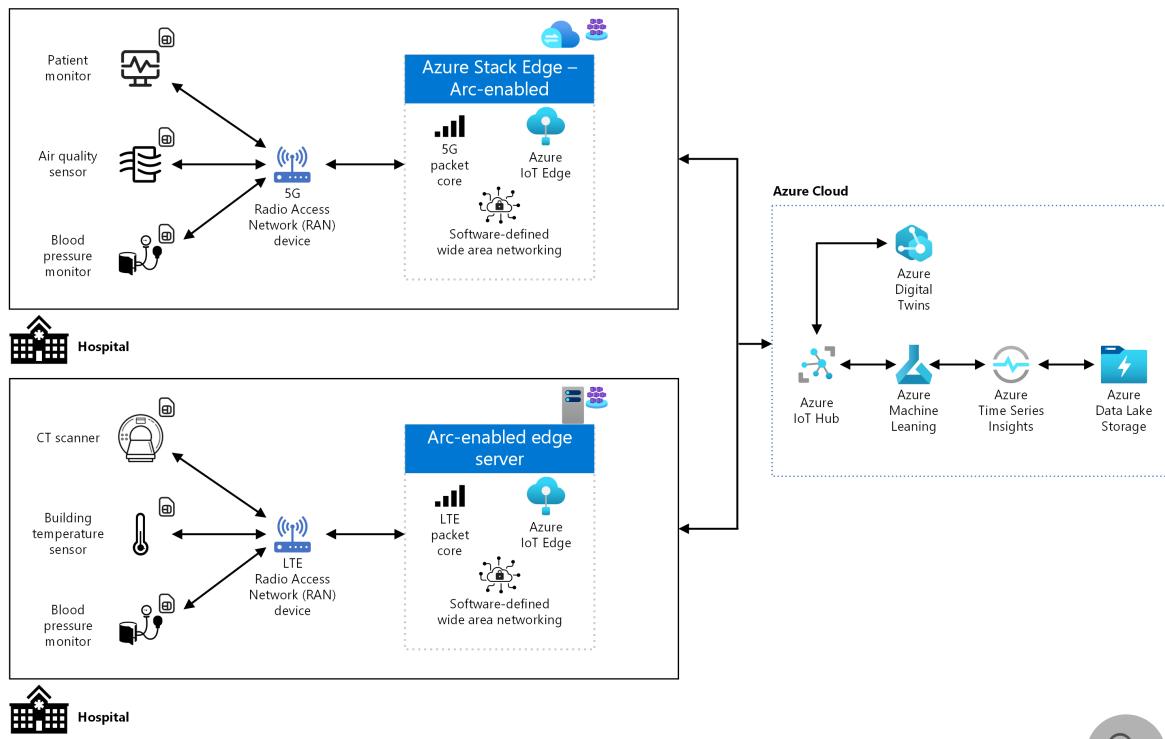
Azure Stack Edge

## 💡 Solution ideas

This article is a solution idea. If you'd like us to expand the content with more information, such as potential use cases, alternative services, implementation considerations, or pricing guidance, let us know by providing [GitHub feedback](#).

This solution describes how buildings and campuses can securely and reliably connect, and scale their on-premises Internet of Things (IoT) devices to the cloud.

## Architecture



Download a [Visio file](#) of this architecture.

## Workflow

1. Hospital buildings use various connected devices to monitor both patient health and facility performance.
  - Health-tracking devices include patient monitors, CT scanners, and blood pressure monitors.
  - Building safety and quality devices include air quality and building temperature sensors.
2. The patient health and building monitoring devices send data to LTE or 5G Radio Access Network (RAN) devices.
3. The 5G or LTE radios in the hospitals forward the data to the 5G or LTE packet cores running on the edge servers. The edge servers can be Azure Stack Edge or any Azure Arc-enabled servers.
4. On the edge servers, the IoT Edge runtime can preprocess data before sending it to Azure for further analysis.
5. In the cloud, Azure IoT Hub ingests data quickly and securely, and sends it to Azure Machine Learning.
6. Azure Machine Learning incorporates the new data to further optimize the model that controls the smart building settings.
7. Data from Azure IoT Hub also feeds into Azure Digital Twins, which provides a map of the hospitals' networked IoT devices as a virtual simulation.
8. Data also feeds into Azure Time Series Insights, which can analyze patient health over a period of time, or treatment efficacy over several hospitals. Time Series Insights also offers a visualization layer to aid in decision-making.
9. All the data is stored in Azure Data Lake Storage, which can store data of any format and size.

## Components

This solution uses the following Azure components:

- [Azure Stack Edge](#) is a portfolio of devices that bring compute, storage, and intelligence to the IoT Edge. Azure Stack Edge acts as a cloud storage gateway that enables data transfers to Azure, while retaining local access to files.
- [Azure Arc-enabled Kubernetes](#) connects Kubernetes clusters running inside or outside of Azure.

- [Azure Sphere](#) is a comprehensive IoT security solution that includes hardware, OS, and cloud components for IoT device security.
- [Azure IoT Edge](#) deploys cloud intelligence locally on IoT devices.
- [Azure IoT Hub](#) is a cloud-based managed service for bidirectional communication between IoT devices and Azure.
- [Azure Machine Learning](#) is an integrated data science solution for data scientists and developers to build, train, and deploy machine learning models.
- [Azure Digital Twins](#) is an IoT platform that creates digital representations of real-world things, places, processes, and people in the cloud.
- [Azure Time Series Insights](#) is an end-to-end IoT analytics platform to monitor, analyze, and visualize industrial IoT analytics data at scale.
- [Azure Data Lake Storage](#) is a scalable and secure data lake for high-performance analytics workloads.

## Scenario details

Cloud services can store and analyze the IoT data to diagnose anomalies and take corrective or preventive actions. Azure services can further analyze and store the data, and use machine learning to optimize building settings.

## Potential use cases

In this solution, a healthcare facility uses LTE or 5G-enabled IoT devices to track both patient health and building performance. The devices use built-in Azure Sphere certified chips to stream data to on-premises edge servers, which communicate with the Azure cloud. On-premises network administrators can view network health through the packet cores on the edge servers.

Other examples of this approach include:

- Predictive maintenance for machines in a coffeehouse.
- Safety and compliance monitoring for perishable food and drink temperatures in a food manufacturing plant.
- Detecting the optimal point for resource extraction in the energy sector, based on data collected by autonomous exploration vehicles.

## Contributors

*This article is maintained by Microsoft. It was originally written by the following contributors.*

Principal author:

- [Nikhil Ravi](#) | Production Management Leader

## Next steps

- [Azure Arc on Kubernetes on Azure Stack Edge](#)
- [Azure Private MEC](#)

## Related resources

- [Manage configurations for Azure Arc enabled servers](#)
- [Deploy AI and ML computing on-premises and to the edge](#)

# Oracle on Azure architecture design

Article • 12/16/2022

Microsoft and Oracle have partnered to enable customers to deploy Oracle applications in the cloud. You can run your Oracle Database and enterprise applications on Oracle Linux, Windows Server, and other supported operating systems in Azure. In addition to Oracle databases, Azure also supports:

- WebLogic Server integrated with Azure services
- Applications on Oracle Linux and WebLogic Server
- Options for high availability and for disaster recovery
- Options for backing up Oracle workloads

The interoperability of Microsoft and Oracle's cloud services enables you to migrate and run mission-critical enterprise workloads across Microsoft Azure and Oracle Cloud Infrastructure (OCI).

Azure provides a wide range of services to support Oracle on Azure. Following are some of the key services:

- [Accelerate your cloud adoption with Microsoft and Oracle](#). Run your Oracle Database and enterprise applications on Azure and Oracle Cloud.
- [Java on Azure](#). Run Java EE applications with Oracle WebLogic Server on Azure Kubernetes Service (AKS) with solutions validated by Microsoft and Oracle.
- [Linux virtual machines in Azure](#). Use preconfigured solutions from Oracle and host Java application servers with Oracle WebLogic on Azure virtual machines (VMs).

## Introduction to Oracle on Azure

If you're new to Azure, the best place to start learning about Azure is Microsoft Learn. This free online platform provides interactive training for Microsoft products and more.

If you have an SAP workload that depends on an Oracle database, the Learn modules in the following table can help you understand what Azure has to offer for Oracle databases and SAP.

- [Explore Azure for SAP databases](#). This module explores SAP databases on Azure and best practices for Azure for SAP workloads, including recommendations from Oracle.

- [Implement high availability for SAP workloads in Azure](#). This module explores high availability and disaster recovery support of Azure for SAP workloads, including use of Oracle Data Guard for high availability of Oracle databases that support SAP workloads.
- [Perform backups and restores for SAP workloads on Azure](#). This module explores backup and restoration of Azure VMs and examines the steps and considerations in backing up and restoring SAP workloads on Azure, including the Oracle databases that support them.

[Search Learn for current offerings about Oracle](#)

## Path to production

The following sections can help you on the path to production for Oracle on Azure:

- [Database migration and deployment](#)
- [Backup and recovery of databases and workloads](#)
- [WebLogic Server](#)

## Database migration and deployment

The following articles describe how to run an Oracle database on Azure and connect to an Oracle database that's running in on OCI.

- [Oracle database migration to Azure](#). This solution idea describes how to migrate an Oracle database to Azure by using Oracle Active Data Guard and Azure Load Balancer. This solution allows you to gradually migrate your application tier in multiple steps.
- [Oracle database migration: Cross-cloud connectivity](#). This example scenario describes creation of an interconnection between Azure and OCI to allow applications that are hosted on Azure to communicate with an Oracle database that's hosted on OCI.
- [Design and implement an Oracle database in Azure](#). This article describes how to size an Oracle workload to run in Azure and decide on the best architecture solution for optimal performance.

## Backup and recovery of databases and workloads

The articles in this section describe methods of backing up and recovering Oracle databases by using Azure resources.

- [Oracle Database in Azure Linux VM backup strategies](#). This article describes strategies for backing up Oracle databases that run on Azure.
- [Back up and recover an Oracle Database on an Azure Linux VM using Azure Files](#). This article demonstrates backing up an Oracle database that's running on a VM by using Oracle RMAN and Azure Files.
- [Back up and recover an Oracle Database on an Azure Linux VM using Azure Backup](#). This article demonstrates using Azure Backup to create snapshots of the VM disks, which include the database files and fast recovery area. Azure Backup can take full-disk snapshots, which are stored in Recovery Services Vault, that are suitable as backups.

## WebLogic Server

The articles in this section can help you decide on a solution for running Oracle WebLogic Server on Azure and help you prepare for migration.

- [What are solutions for running Oracle WebLogic Server on Azure Virtual Machines?](#) This article describes solutions for running Oracle WebLogic Server (WLS) on Azure VMs.
- [What are solutions for running Oracle WebLogic Server on the Azure Kubernetes Service?](#) This article describes solutions for running Oracle WebLogic Server (WLS) on the Azure Kubernetes Service (AKS).
- [Migrate WebLogic Server applications to Azure Virtual Machines](#). This guide describes what you should be aware of when you want to migrate an existing WebLogic application to run on Azure VMs.

## Best practices

The articles in this section can help you identify and select the services and configurations that will best support your solutions for Oracle on Azure.

- [SAP deployment on Azure using an Oracle database](#). This reference architecture shows a set of proven practices for running SAP NetWeaver with Oracle Database in Azure, with high availability.

- [Connectivity to Oracle Cloud Infrastructure](#). This article describes methods for integrating an Azure landing zone architecture with Oracle Cloud Infrastructure (OCI).

## Oracle on Azure architectures

The articles in this section describe architectures for deploying Oracle applications on Azure and integrating services on Azure with services on OCI.

- [Architectures to deploy Oracle applications on Azure](#). This article describes recommended architectures for deploying Oracle E-Business Suite, JD Edwards EnterpriseOne, and PeopleSoft in cross-cloud configurations or entirely on Azure.
- [Oracle application solutions integrating Microsoft Azure and Oracle Cloud Infrastructure](#). This article describes how to partition a multi-tier application to run the database tier on Oracle Cloud Infrastructure (OCI) and the application and other tiers on Microsoft Azure.
- [Reference architectures for Oracle Database Enterprise Edition on Azure](#). This article provides detailed information about deploying Oracle Database Enterprise Edition on Azure and using Oracle Data Guard for disaster recovery.

## Stay current with Oracle on Azure

To stay informed about Oracle on Azure, check Azure updates and the Microsoft Azure blog.

[Check Azure updates for news about Oracle on Azure](#)

[Check Microsoft Azure Blog for posts about Oracle on Azure](#)

## Additional resources

The following articles provide additional support for implementing Oracle on Azure:

- [Overview of Oracle Applications and solutions on Azure](#). This article introduces capabilities to run Oracle solutions by using Azure infrastructure.
- [Oracle VM images and their deployment on Microsoft Azure](#). This article provides information about Oracle solutions based on virtual machine images published by Oracle in the Azure Marketplace.

- Oracle application solutions integrating Microsoft Azure and Oracle Cloud Infrastructure. Microsoft and Oracle provide low-latency, high-throughput, cross-cloud connectivity between Azure and OCI, allowing you to partition a multi-tier application across both cloud services.

## Example solutions

Following are some additional solution ideas that might be helpful:

- Run Oracle databases on Azure. This solution idea illustrates a canonical architecture to achieve high availability for your Oracle Database Enterprise Edition in Azure by using Azure Load Balancers or Application Gateways.
- Oracle database migration: Lift and shift. If you're properly licensed to use Oracle software, you're allowed to migrate Oracle databases to Azure Virtual Machines (VMs).

# Connectivity to Oracle Cloud Infrastructure

Article • 12/01/2022

This section provides different connectivity approaches to integrate an Azure landing zone architecture to Oracle Cloud Infrastructure (OCI).

## Design considerations:

- Using ExpressRoute and FastConnect, customers can connect a virtual network in Azure with a virtual cloud network in OCI, if the private IP address space doesn't overlap. Once you establish connectivity, resources in the Azure virtual network can communicate with resources in the OCI virtual cloud network as if they were both in the same network.
- Azure ExpressRoute [FastPath](#) is designed to improve the data path performance between two networks, both on-premises and Azure, and for this scenario, between OCI and Azure. When enabled, FastPath sends network traffic directly to virtual machines in the virtual network, bypassing the ExpressRoute gateway.
  - FastPath is available on all ExpressRoute circuits.
  - FastPath still requires a virtual network gateway to be created for route exchange purposes. The virtual network gateway must use either the Ultra Performance SKU or the ErGw3AZ SKU for the ExpressRoute gateway to enable route management.
- There are features that are currently [not supported](#) in ExpressRoute FastPath, such as Azure Virtual WAN hubs or VNet peering.
- While you can use [ExpressRoute Global Reach](#) to enable communication from on-premises to OCI via ExpressRoute circuits, it might incur more bandwidth costs that you can calculate by using the [Azure pricing calculator](#). It's important to consider any extra costs when you migrate large amounts of data from on-premises to Oracle by using ExpressRoute circuits.
- In Azure regions that support [Availability Zones](#), placing your Azure workloads in one zone or the other can have a small effect on latency. Design your application to balance availability and performances requirements.
- Interconnectivity between Azure and OCI is only available for [specific regions](#).

- For more in-depth documentation about interconnectivity between Azure and OCI, see [Oracle application solutions to integrate Microsoft Azure and Oracle Cloud Infrastructure](#) or see [Access to Microsoft Azure in OCI](#).

### Design recommendations:

- Create the ExpressRoute circuit that will be used to interconnect Azure with OCI in the **connectivity** subscription.
- You can interconnect an Azure network architecture based on the traditional hub and spoke architecture or Azure Virtual WAN-based network topologies. It can be done by connecting the ExpressRoute circuit that will be used to interconnect Azure to OCI to the hub VNet or Virtual WAN hub as shown in the following diagram.

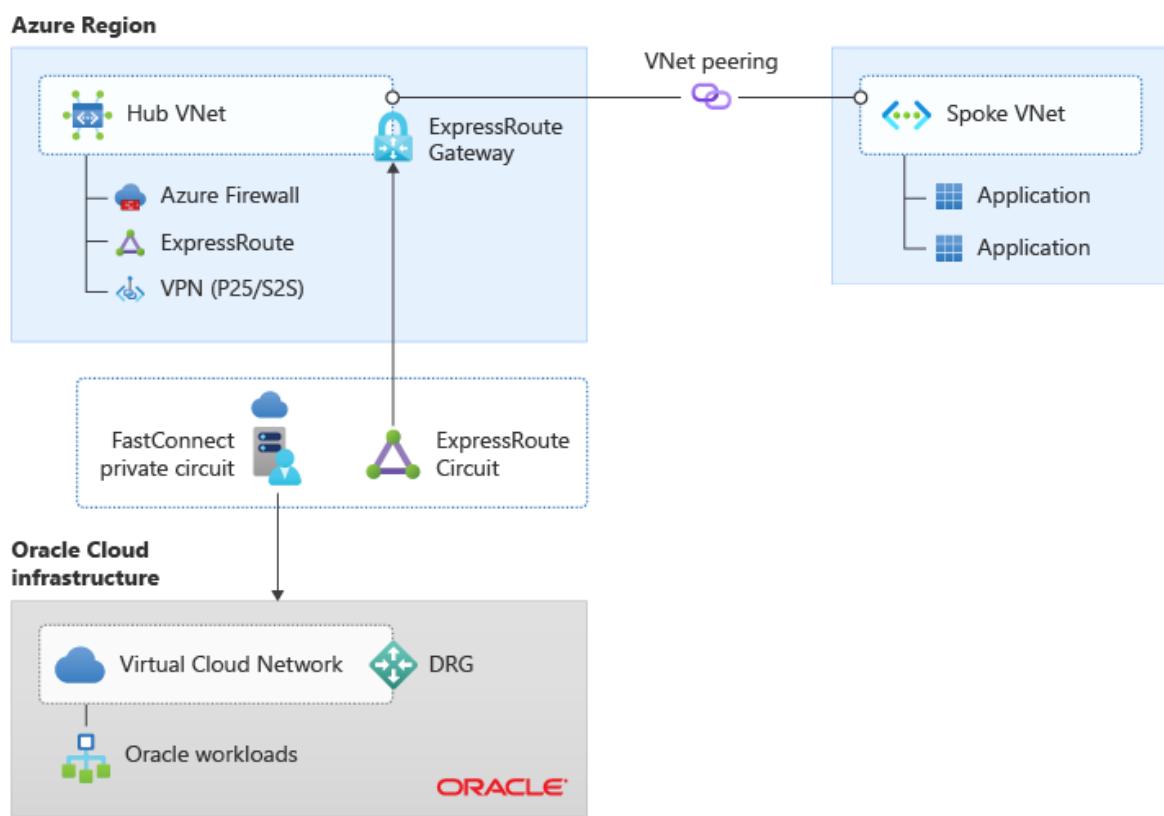


Figure 1: Interconnectivity between Azure and OCI via ExpressRoute.

- If your application requires the lowest possible latency between Azure and OCI, consider deploying your application in a single VNet with an ExpressRoute gateway and FastPath enabled.

## Azure Region

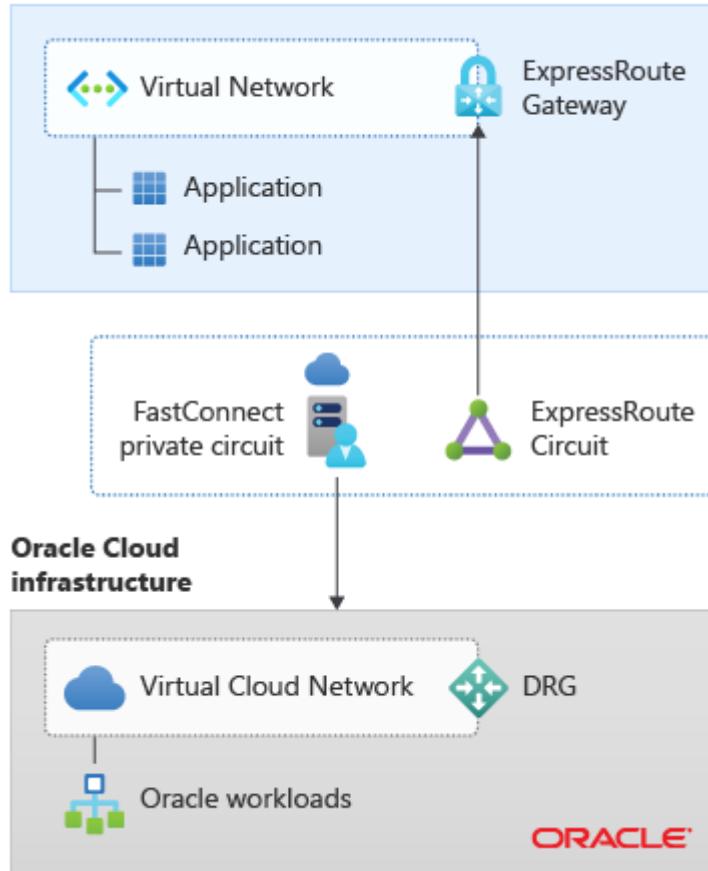


Figure 2: Interconnectivity between Azure and OCI with a single VNet.

- When you deploy Azure resources across Availability Zones, perform latency tests from Azure VMs located in different Availability Zones to OCI resources to understand which of the three Availability Zones provides the lowest latency to the OCI resources.
- To operate Oracle resources hosted in OCI by using Azure resources and technologies, you could:
  - **From Azure:** Deploy a jump box in a spoke VNet. The jump box provides access to the virtual cloud network in OCI as shown in the following picture:

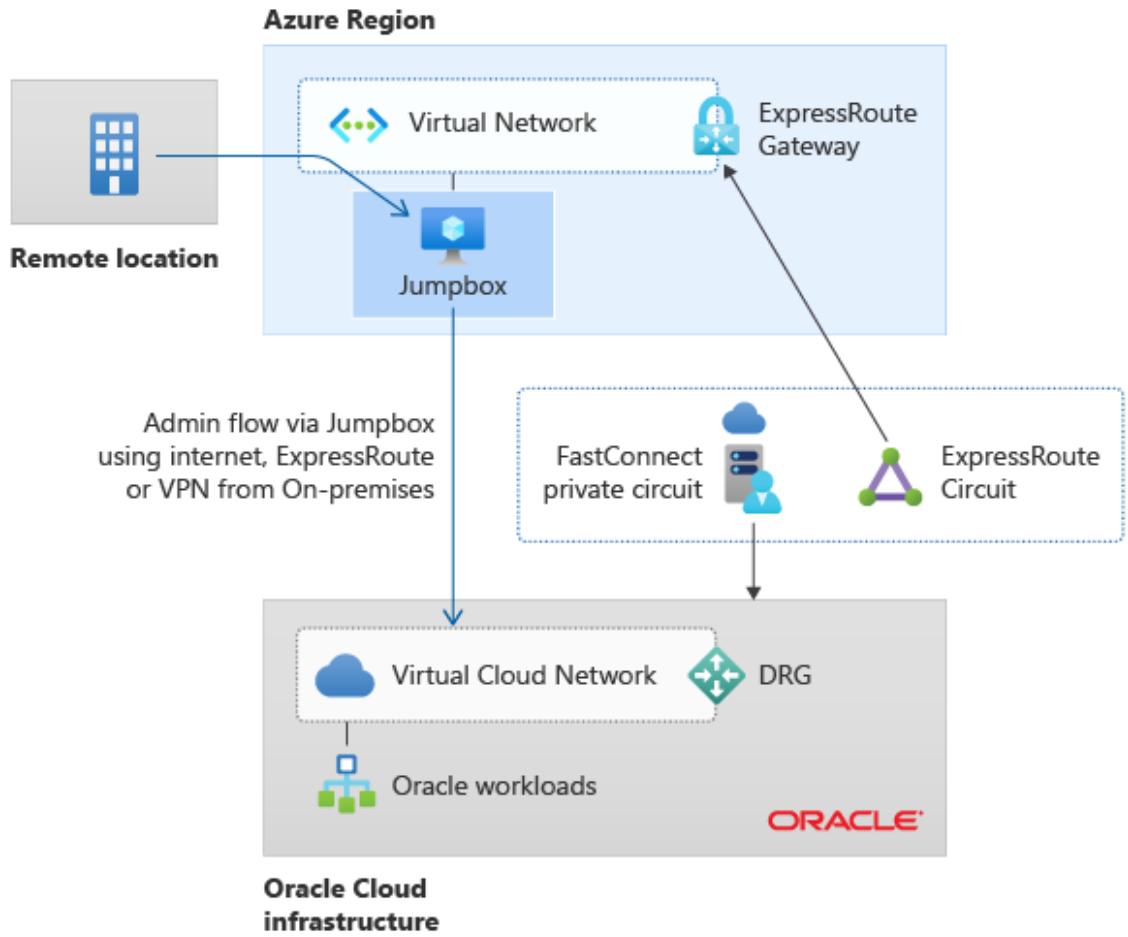


Figure 3: Managing OCI resources from Azure via a jump box.

- **From on-premises:** Use ExpressRoute Global Reach to bind an existing ExpressRoute circuit that connects on-premises to Azure, to an OCI ExpressRoute circuit that interconnects Azure to OCI. In this way, the Microsoft Enterprise Edge (MSEE) router becomes the central routing point between both ExpressRoute circuits.

## Azure Region

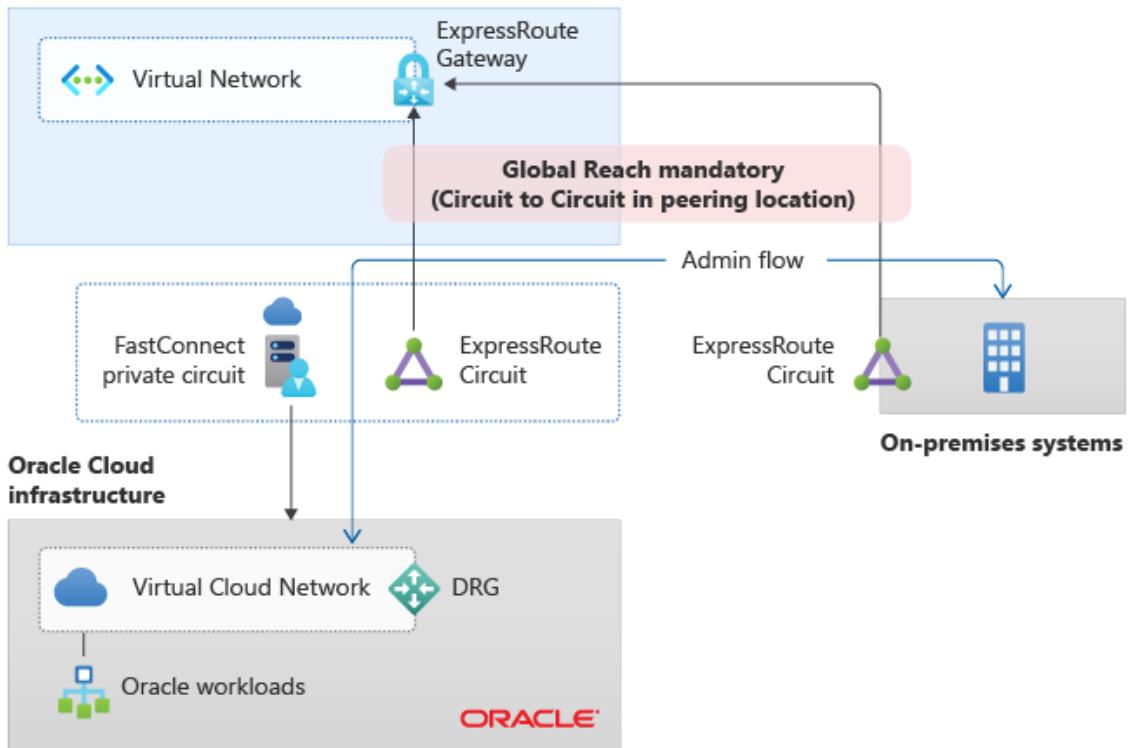


Figure 4: Managing OCI resources from on-premises via ExpressRoute Global Reach.

## Next steps

For information on connectivity to other cloud providers, see [Connectivity to other cloud providers](#).

# Overview of Oracle Applications and solutions on Azure

Article • 06/20/2023

Applies to:  Linux VMs

In this article, you learn about running Oracle solutions using the Azure infrastructure.

## Oracle databases on Azure infrastructure

Oracle supports running its Database 12.1 and higher Standard and Enterprise editions in Azure on VM images based on Oracle Linux. You can run Oracle databases on Azure infrastructure using Oracle Database on Oracle Linux images available in the Azure Marketplace.

- Oracle Database 12.2, and 18.3 Enterprise Edition
- Oracle Database 12.2, and 18.3 Standard Edition
- Oracle Database 19.3

You can also take one of the following approaches:

- Set up Oracle Database on a non-Oracle Linux image available in Azure.
- Build a solution on a custom image you create from scratch in Azure.
- Upload a custom image from your on-premises environment.

You can also choose to configure your solution with multiple attached disks. You can improve database performance by installing Oracle Automated Storage Management (ASM). For the best performance for production workloads of Oracle Database on Azure, be sure to properly size the VM image and select the right storage options based on throughput, IOPS & latency. For instructions on how to quickly get an Oracle Database up and running in Azure using the Oracle published VM image, see [Create an Oracle Database in an Azure VM](#).

## Deploy Oracle VM images on Microsoft Azure

This section covers information about Oracle solutions based on virtual machine (VM) images published by Oracle in the Azure Marketplace. To get a list of currently available Oracle images, run the following command using Azure CLI or Azure Cloud Shell

```
az vm image list --publisher oracle --output table -all
```

The images are bring-your-own-license. You're charged only for the costs of compute, storage, and networking incurred running a VM. You can also choose to build your solutions on a custom image that you create from scratch in Azure or upload a custom image from your on-premises environment.

 **Important**

You require a proper license to use Oracle software and a current support agreement with Oracle. Oracle has guaranteed license mobility from on-premises to Azure. For more information about license mobility, see the [Oracle and Microsoft Strategic Partnership FAQ](#).

## Applications on Oracle Linux and WebLogic server

Run enterprise applications on WebLogic server in Azure on supported Oracle Linux images. For more information, see the WebLogic documentation, [Oracle WebLogic Server on Azure Solution Overview](#).

## WebLogic Server with Azure service integrations

Oracle and Microsoft are collaborating to bring WebLogic Server to the Azure Marketplace in the form of Azure Application offering. For more information about these offers, see [What are solutions for running Oracle WebLogic Server](#).

## Oracle WebLogic Server VM images

**Clustering is supported on Enterprise Edition only.** You're licensed to use WebLogic clustering only when you use the Enterprise Edition of Oracle WebLogic Server. Don't use clustering with Oracle WebLogic Server Standard Edition. **UDP multicast isn't supported.** Azure supports UDP unicasting, but not multicasting or broadcasting. Oracle WebLogic Server can rely on Azure UDP unicast capabilities. For best results relying on UDP unicast, we recommend that the WebLogic cluster size is kept static, or kept with no more than 10 managed servers. **Oracle WebLogic Server expects public and private ports to be the same for T3 access.** For example, when using Enterprise JavaBeans (EJB). Consider a multi-tier scenario where a service layer application is running on an Oracle WebLogic Server cluster consisting of two or more VMs, in a virtual network named

SLWLS. The client tier is in a different subnet in the same virtual network, running a simple Java program trying to call EJB in the service layer. Because you must load balance the service layer, a public load-balanced endpoint needs to be created for the VMs in the Oracle WebLogic Server cluster. If the private port specified is different from the public port an error occurs. For example, if you use `7006:7008`, the following error occurs because for any remote T3 access, Oracle WebLogic Server expects the load balancer port and the WebLogic managed server port to be the same.

```
[java] javax.naming.CommunicationException [Root exception is  
java.net.ConnectException: t3://example.cloudapp.net:7006:  
  
Bootstrap to: example.cloudapp.net/138.91.142.178:7006' over: 't3' got an error or  
timed out]
```

In the preceding case, the client is accessing port 7006, which is the load balancer port, and the managed server is listening on 7008, which is the private port. This restriction is applicable only for T3 access, not HTTP.

To avoid this issue, use one of the following workarounds:

- Use the same private and public port numbers for load balanced endpoints dedicated to T3 access.
- Include the following JVM parameter when starting Oracle WebLogic Server:  
`configCopy Dweblogic.rjvm.enableprotocolswitch=true`
- Dynamic clustering and load balancing limitations. Suppose you want to use a dynamic cluster in Oracle WebLogic Server and expose it through a single, public load-balanced endpoint in Azure. This approach can be done as long as you use a fixed port number for each of the managed servers, not dynamically assigned from a range, and don't start more managed servers than there are machines the administrator is tracking. There should be no more than one managed server per VM. If your configuration results in more Oracle WebLogic Servers being started than there are VMs, it isn't possible for more than one of those instances of Oracle WebLogic Servers to bind to a given port number. That is, if multiple Oracle WebLogic Server instances share the same virtual machine, the others on that VM fail. If you configure the admin server to automatically assign unique port numbers to its managed servers, then load balancing isn't possible because Azure doesn't support mapping from a single public port to multiple private ports, as would be required for this configuration.
- Multiple instances of Oracle WebLogic Server on a VM. Depending on your deployment requirements, you might consider running multiple instances of

Oracle WebLogic Server on the same VM, if the VM is large enough. For example, on a midsize VM, which contains two cores, you could choose to run two instances of Oracle WebLogic Server. However, we still recommend that you avoid introducing single points of failure into your architecture. Running multiple instances of Oracle WebLogic Server on just one VM would be such a single point.

Using at least two VMs could be a better approach. Each VM can run multiple instances of Oracle WebLogic Server. Each instance of Oracle WebLogic Server could still be part of the same cluster. However, it's currently not possible to use Azure to load-balance endpoints that are exposed by such Oracle WebLogic Server deployments within the same VM. Azure Load Balancer requires the load-balanced servers to be distributed among unique VMs.

## High availability and disaster recovery options

When using Oracle solutions in Azure, you're responsible for implementing a high availability and disaster recovery solution to avoid any downtime. You can also implement high availability and disaster recovery for Oracle Database Enterprise Edition by using Data Guard, Active Data Guard, or Oracle GoldenGate. The approach requires two databases on two separate VMs, which should be in the same virtual network to ensure they can access each other over the private persistent IP address.

We recommend placing the VMs in the same availability set to allow Azure to place them into separate fault domains and upgrade domains. If you want to have geo-redundancy, set up the two databases to replicate between two different regions and connect the two instances with a VPN Gateway. To walk through the basic setup procedure on Azure, see [Implement Oracle Data Guard on an Azure Linux virtual machine](#).

With Oracle Data Guard, you can achieve high availability with a primary database in one VM, a secondary (standby) database in another VM, and one-way replication set up between them. The result is read access to the copy of the database. With Oracle GoldenGate, you can configure bi-directional replication between the two databases. To learn how to set up a high-availability solution for your databases using these tools, see [Active Data Guard and GoldenGate](#). If you need read-write access to the copy of the database, you can use Oracle Active Data Guard. To walk through the basic setup procedure on Azure, see [Implement Oracle Golden Gate on an Azure Linux VM](#).

In addition to having a high availability and disaster recovery solution architected in Azure, you should have a backup strategy in place to restore your database.

# Backup Oracle workloads

Different [backup strategies](#) are available for Oracle on Azure VMs, the following backups are other options:

- Using [Azure files](#)
- Using [Azure backup](#)
- Using [Oracle RMAN Streaming data](#) backup

# Deploy Oracle applications on Azure

Use Terraform templates to set up Azure infrastructure and install Oracle applications.

For more information, see [Terraform on Azure](#).

Oracle has certified the following applications to run in Azure when connecting to an Oracle database by using the Azure with Oracle Cloud interconnect solution:

- E-Business Suite
- JD Edwards EnterpriseOne
- PeopleSoft
- Oracle Retail applications
- Oracle Hyperion Financial Management

You can deploy custom applications in Azure that connect with OCI and other Azure services.

# Support for JD Edwards

According to Oracle Support, JD Edwards EnterpriseOne versions 9.2 and above are supported on any public cloud offering that meets their specific Minimum Technical Requirements (MTR). You need to create custom images that meet their MTR specifications for operating system and software application compatibility. For more information, see [Doc ID 2178595.1](#).

# Licensing

Deployment of Oracle solutions in Azure is based on a bring-your-own-license model. This model assumes that you have licenses to use Oracle software and that you have a current support agreement in place with Oracle. Microsoft Azure is an authorized cloud environment for running Oracle Database. The Oracle Core Factor table isn't applicable when licensing Oracle databases in the cloud. Instead, when using VMs with Hyper-

Threading Technology enabled for Enterprise Edition databases, count two vCPUs as equivalent to one Oracle Processor license if hyperthreading is enabled, as stated in the policy document. The policy details can be found at [Licensing Oracle Software in the Cloud Computing Environment](#).

Oracle databases generally require higher memory and I/O. For this reason, we recommend [Memory Optimized VMs](#) for these workloads. To optimize your workloads further, we recommend [Constrained Core vCPUs](#) for Oracle Database workloads that require high memory, storage, and I/O bandwidth, but not a high core count. When you migrate Oracle software and workloads from on-premises to Microsoft Azure, Oracle provides license mobility as stated in [Oracle and Microsoft Strategic Partnership FAQ](#).

## Next steps

You now have an overview of current Oracle databases and solutions based on VM images in Microsoft Azure. Your next step is to deploy your first Oracle database on Azure.

- [Create an Oracle database on Azure](#)

# Design and implement an Oracle database in Azure

Article • 04/13/2023

Applies to:  Linux VMs

Azure is home for all Oracle workloads, including workloads that need to continue to run optimally in Azure with Oracle. If you have the [Oracle Diagnostic Pack](#) or the [Automatic Workload Repository](#) (AWR), you can gather data about your workloads. Use this data to assess the Oracle workload, size the resource needs, and migrate the workload to Azure. The various metrics provided by Oracle in these reports can provide an understanding of application performance and platform usage.

This article helps you to prepare an Oracle workload to run in Azure and explore the best architecture solutions to provide optimal cloud performance. The data provided by Oracle in the Statspack and even more so in its descendent, the AWR, assists you in developing clear expectations. These expectations include the limits of physical tuning through architecture, the advantages of logical tuning of database code, and the overall database design.

## Differences between the two environments

When you're migrating on-premises applications to Azure, keep in mind a few important differences between the two environments.

One important difference is that in an Azure implementation, resources such as VMs, disks, and virtual networks are shared among other clients. In addition, resources can be throttled based on the requirements. Instead of focusing on avoiding failing, Azure focuses more on surviving the failure. The first approach tries to increase *mean time between failures* (MTBF) and the second tries to decrease *mean time to recovery* (MTTR).

The following table lists some of the differences between an on-premises implementation and an Azure implementation of an Oracle database.

On-premises implementation	Azure implementation
<b>Networking</b>	LAN/WAN
<b>Security group</b>	IP/port restriction tools
<b>Resilience</b>	MTBF
	<a href="#">Network security group (NSG)</a>
	MTTR

	<b>On-premises implementation</b>	<b>Azure implementation</b>
<b>Planned maintenance</b>	Patching/upgrades	<a href="#">Availability sets</a> with patching/upgrades managed by Azure
<b>Resource</b>	Dedicated	Shared with other clients
<b>Regions</b>	Datacenters	<a href="#">Region pairs</a>
<b>Storage</b>	SAN/physical disks	<a href="#">Azure-managed storage</a>
<b>Scale</b>	Vertical scale	Horizontal scale

## Requirements

Consider the following requirements before you start your migration:

- Determine the real CPU usage. Oracle licenses by core, which means that sizing your vCPU needs can be essential to help you reduce costs.
- Determine the database size, backup storage, and growth rate.
- Determine the I/O requirements, which you can estimate based on Oracle Statspack and the AWR reports. You can also estimate the requirements from storage monitoring tools available from the operating system.

## Configuration options

It's a good idea to generate an AWR report and obtain some metrics from it to help you make decisions about configuration. Then, there are four potential areas that you can tune to improve performance in an Azure environment:

- Virtual machine size
- Network throughput
- Disk types and configurations
- Disk cache settings

## Generate an AWR report

If you have an existing an Oracle Enterprise Edition database and are planning to migrate to Azure, you have several options. If you have the [Diagnostics Pack](#) for your Oracle instances, you can run the Oracle AWR report to get the metrics, such as IOPS, Mbps, and GiBs. For those databases without the Diagnostics Pack license, or for an Oracle Standard Edition database, you can collect the same important metrics with a

Statspack report after you collect manual snapshots. The main differences between these two reporting methods are that AWR is automatically collected, and that it provides more information about the database than does Statspack.

Consider running your AWR report during both regular and peak workloads, so you can compare. To collect the more accurate workload, consider an extended window report of one week, as opposed to one day. AWR provides averages as part of its calculations in the report. By default, the AWR repository retains eight days of data and takes snapshots at hourly intervals.

For a datacenter migration, you should gather reports for sizing on the production systems. Estimate remaining database copies used for user testing, test, and development by percentages. For example, estimate 50 percent of production sizing.

To run an AWR report from the command line, use the following command:

Bash

```
sqlplus / as sysdba  
@$ORACLE_HOME/rdbms/admin/awrrpt.sql;
```

## Key metrics

The report prompts you for the following information:

- Report type: HTML or TEXT. The HTML type provides more information.
- The number of days of snapshots to display. For example, for one-hour intervals, a one-week report produces 168 snapshot IDs.
- The beginning `SnapshotID` for the report window.
- The ending `SnapshotID` for the report window.
- The name of the report that the AWR script creates.

If you're running the AWR report on a Real Application Cluster (RAC), the command-line report is the `awrgrpt.sql` file, instead of `awrrpt.sql`. The `g` report creates a report for all nodes in the RAC database in a single report. This report eliminates the need to run one report on each RAC node.

You can obtain the following metrics from the AWR report:

- Database name, instance name, and host name
- Database version for supportability by Oracle
- CPU/Cores
- SGA/PGA, and advisors to let you know if undersized

- Total memory in GB
- CPU percentage busy
- DB CPUs
- IOPs (read/write)
- MBPs (read/write)
- Network throughput
- Network latency rate (low/high)
- Top wait events
- Parameter settings for database
- Whether the database is RAC, Exadata, or using advanced features or configurations

## Virtual machine size

Here are some steps you can take to configure virtual machine size for optimal performance.

### Estimate VM size based on CPU, memory, and I/O usage from the AWR report

Look at the top five timed foreground events that indicate where the system bottlenecks are. For example, in the following diagram, the log file sync is at the top. It indicates the number of waits that is required before the log writer writes the log buffer to the redo log file. These results indicate that better performing storage or disks are required. In addition, the diagram also shows the number of CPU cores and the amount of memory.

## Top 5 Timed Foreground Events

Event	Waits	Time(s)	Avg wait (ms)	% DB time	Wait Class
log file sync	16,997,881	163,293	10	72.04	Commit
buffer busy waits	4,014,884	18,767	5	8.28	Concurrency
DB CPU		15,413		6.80	
enq: TX - index contention	2,001,818	11,531	6	5.09	Concurrency
library cache: mutex X	2,024,372	6,986	3	3.08	Concurrency

## Host CPU (CPUs: 16 Cores: 16 Sockets: 2)

Load Average Begin	Load Average End	%User	%System	%WIO	%Idle
4.65	11.03	25.8	5.1	6.4	68.4

## Instance CPU

%Total CPU	%Busy CPU	%DB time waiting for CPU (Resource Manager)
31.7	100.3	0.0

## Memory Statistics

	Begin	End
Host Mem (MB):	112,804.7	112,804.7
SGA use (MB):	61,440.0	61,440.0
PGA use (MB):	1,051.6	1,559.6
% Host Mem used for SGA+PGA:	55.40	55.85

The following diagram shows the total I/O of read and write. There were 59 GB read and 247.3 GB written during the time of the report.

## IOStat by Function/Filetype summary

- 'Data' columns suffixed with M,G,T,P are in multiples of 1024 other columns suffixed with K,M,G,T,P are in multiples of 1000
- Ordered by (Data Read + Write) desc for each function

Function/File Name	Reads: Data	Reqs per sec	Data per sec	Writes: Data	Reqs per sec	Data per sec	Waits: Count	Avg Tm(ms)
Others	42G	25.51	13.1724	87.5G	63.98	27.4202	171.1K	6.55
Others (Flashback Log)	32M	0.03	0.009795	47.5G	46.92	14.8805	153.4K	7.10
Others (Log File)	41.7G	20.11	13.0716	0M	0.00	0M	127	13.66
Others (Archive Log)	0M	0.00	0M	39.8G	12.47	12.4677	0	
Others (Control File)	297M	5.37	0.09018	234M	4.58	0.01632	17.6K	1.70
Others (Data File)	0M	0.00	0M	1M	0.00	0.00306	10	1.80
DBWR	0M	0.00	0M	120.8G	802.32	37.8629	6	1.67
DBWR (Data File)	0M	0.00	0M	120.8G	802.32	37.8629	0	
DBWR (Control File)	0M	0.00	0M	0M	0.00	0M	6	1.67
LGWR	2M	0.05	0.00612	39G	587.91	12.2329	180	1.93
LGWR (Log File)	0M	0.01	0M	39G	587.83	12.2316	36	1.92
LGWR (Control File)	2M	0.04	0.00612	4M	0.08	0.001224	144	1.94
Buffer Cache Reads	17.1G	638.76	5.36845	0M	0.00	0M	2086.6K	1.96
Buffer Cache Reads (Data File)	17.1G	638.76	5.36845	0M	0.00	0M	2086.6K	1.96
Direct Writes	0M	0.00	0M	1M	0.02	0.00306	0	
Direct Writes (Data File)	0M	0.00	0M	1M	0.02	0.00306	0	
Direct Reads	1M	0.02	0.00306	0M	0.00	0M	0	
Direct Reads (Data File)	1M	0.02	0.00306	0M	0.00	0M	0	
Streams AQ	0M	0.00	0M	0M	0.00	0M	1	2.00
Streams AQ (Data File)	0M	0.00	0M	0M	0.00	0M	1	2.00
<b>TOTAL:</b>	<b>59.2G</b>	<b>664.34</b>	<b>18.5417</b>	<b>247.3G</b>	<b>1454.23</b>	<b>77.5164</b>	<b>2257.8K</b>	<b>2.31</b>

## Choose a VM

Based on the information that you collected from the AWR report, the next step is to choose a VM of a similar size that meets your requirements. For more information about available VMs, see [Memory optimized virtual machine sizes](#).

## Fine-tune the VM sizing with a similar VM series based on the ACU

After you choose the VM, pay attention to the Azure compute unit (ACU) for the VM. You might choose a different VM based on the ACU value that better suits your requirements. For more information, see [Azure compute unit](#).

M-series*							
ACU: 160-180							
Size	vCPU	Memory: GiB	Local SSD: GiB	Max data disks	Max cached and local disk throughput: IOPS / MBps (cache size in GiB)	Max uncached disk throughput: IOPS / MBps	Max NICs / Network bandwidth
Standard_M64ms	64	1792	2048	32	80,000 / 800 (6348)	40,000 / 1,000	32 / extremely high
Standard_M128s**	128	2048	4096	64	160,000 / 1,600 (12,696)	80,000 / 2,000	32 / extremely high

## Network throughput

The following diagram shows the relation between throughput and IOPS:



The total network throughput is estimated based on the following information:

- SQL\*Net traffic
- MBps times the number of servers (outbound stream, such as Oracle Data Guard)
- Other factors, such as application replication

background timeouts	20,862	6.39	0.00
branch node splits	614	0.19	0.00
buffer is not pinned count	238,355,647	72,965.81	14.03
buffer is pinned count	353,113,420	108,095.64	20.78
bytes received via SQL*Net from client	1,758,217,872	538,228.42	103.49
bytes sent via SQL*Net to client	1,582,672,212	484,490.11	93.16
calls to get snapshot scn: kcmgss	50,773,529	50,230.03	5.01
calls to kcmgas	17,354,528	5,312.60	1.02
calls to kcmgcs	3,293,854	1,008.32	0.19
cell physical IO interconnect bytes	329,064,153,600	100,733,636.76	19,369.14
change write time	112,450	34.42	0.01
cleanout - number of ktugct calls	4,186,965	1,281.72	0.25
cleanouts and rollbacks - consistent read gets	4,231	1.30	0.00

Based on your network bandwidth requirements, there are various gateway types for you to choose from. These types include basic, VpnGw, and Azure ExpressRoute. For more information, see [VPN Gateway pricing](#).

## Recommendations

- Network latency is higher compared to an on-premises deployment. Reducing network round trips can greatly improve performance.
- To reduce round-trips, consolidate applications that have high transactions or *chatty* apps on the same virtual machine.
- Use virtual machines with [accelerated networking](#) for better network performance.
- For certain Linux distributions, consider enabling [TRIM/UNMAP support](#).
- Install [Oracle Enterprise Manager](#) on a separate virtual machine.
- Huge pages aren't enabled on Linux by default. Consider enabling huge pages, and set `use_large_pages = ONLY` on the Oracle DB. This approach might help increase performance. For more information, see [USE\\_LARGE\\_PAGES](#).

## Disk types and configurations

Here are some tips as you consider disks.

- **Default OS disks:** These disk types offer persistent data and caching. They're optimized for operating system access at startup, and aren't designed for either transactional or data warehouse (analytical) workloads.
- **Managed disks:** Azure manages the storage accounts that you use for your VM disks. You specify the disk type and the size of the disk that you need. The type is most often Premium (SSD) for Oracle workloads. Azure creates and manages the disk for you. A premium SSD-managed disk is only available for memory-optimized and designed VM series. After you choose a particular VM size, the

menu shows only the available premium storage SKUs that are based on that VM size.

Home > Disks >

## Create a managed disk

Basics    Encryption    Networking    Advanced    Tags    Review + create

Select the disk type and size needed for your workload. Azure disks are designed for 99.999% availability. Azure managed disks encrypt your data at rest, by default, using Storage Service Encryption. [Learn more about disks](#).

**Project details**

Select the subscription to manage deployed resources and costs. Use resource groups like folders to organize and manage all your resources.

Subscription \* ⓘ Contoso Subscription

Resource group \* ⓘ (New) myResourceGroup

Create new

**Disk details**

Disk name \* ⓘ myDisk

Region \* ⓘ (US) East US 2

Availability zone None

Source type None

Size \* ⓘ 1024 GiB  
Premium SSD LRS  
Change size

Review + create    < Previous    Next : Encryption >   

After you configure your storage on a VM, you might want to load test the disks before you create a database. Knowing the I/O rate in terms of both latency and throughput can help you determine if the VMs support the expected throughput with latency targets. There are several tools for application load testing, such as Oracle Orion, Sysbench, SLOB, and Fio.

Run the load test again after you deploy an Oracle database. Start your regular and peak workloads, and the results show you the baseline of your environment. Be realistic in the workload test. It doesn't make sense to run a workload that is nothing like what you run on the VM in reality.

Because Oracle can be an I/O intensive database, it's important to size the storage based on the IOPS rate rather than the storage size. For example, if the required IOPS value is 5,000, but you only need 200 GB, you might still get the P30 class premium disk even though it comes with more than 200 GB of storage.

You can get the IOPS rate from the AWR report. The redo log, physical reads, and writes rate determine the IOPS rate. Always verify that the VM series you choose has the ability to handle the I/O demand of the workload. If the VM has a lower I/O limit than the storage, the VM sets the limit maximum.

	Per Second	Per Transaction	Per Exec	Per Call
DB Time(s):	69.4	0.0	0.00	0.01
DB CPU(s):	4.7	0.0	0.00	0.00
Redo size:	12,261,192.2	2,357.6		
Logical reads:	198,807.9	38.2		
Block changes:	76,974.9	14.8		
Physical reads:	687.4	0.1		
Physical writes:	4,846.9	0.9		
User calls:	10,406.6	2.0		
Parses:	5.5	0.0		
Hard parses:	0.1	0.0		
W/A MB processed:	0.1	0.0		
Logons:	0.2	0.0		
Executes:	38,500.2	7.4		
Rollbacks:	0.0	0.0		
Transactions:	5,200.7			

For example, the redo size is 12,200,000 bytes per second, which is equal to 11.63 MBPs. The IOPS value is  $12,200,000 / 2,358 = 5,174$ .

After you have a clear picture of the I/O requirements, you can choose a combination of drives that are best suited to meet those requirements.

## Disk type recommendations

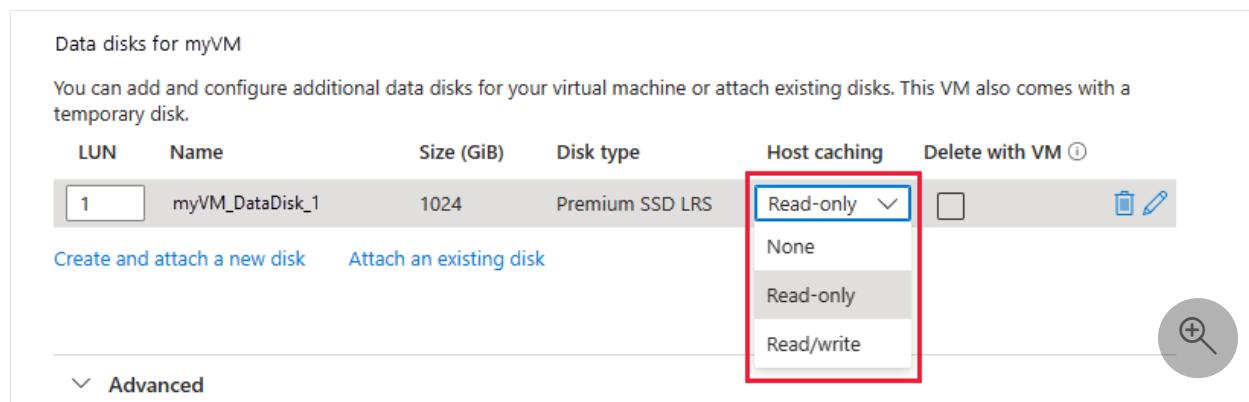
- For data tablespace, spread the I/O workload across several disks by using managed storage or Oracle Automatic Storage Management (ASM).
- Use Oracle advanced compression to reduce I/O for both data and indexes.
- Separate redo logs, temp, and undo tablespaces on separate data disks.
- Don't put any application files on default operating system disks. These disks aren't optimized for fast VM boot times, and they might not provide good performance for your application.
- When you're using M-Series VMs on premium storage, enable [write accelerator](#) on the redo logs disk.
- Consider moving redo logs with high latency to the ultra disk.

# Disk cache settings

Although you have three options for host caching, only read-only caching is recommended for a database workload on an Oracle database. Read/write can introduce significant vulnerabilities to a data file, because the goal of a database write is to record it to the data file, not to cache the information. With read-only, all requests are cached for future reads. All writes continue to be written to disk.

## Disk cache recommendations

To maximize throughput, start with read-only for host caching whenever possible. For premium storage, keep in mind that you must disable the barriers when you mount the file system with the read-only options. Update the `/etc/fstab` file with the universally unique identifier to the disks.



The screenshot shows the 'Data disks for myVM' section. It lists a single disk entry: LUN 1, Name 'myVM\_DataDisk\_1', Size 1024 GiB, Disk type 'Premium SSD LRS'. The 'Host caching' dropdown menu is open, showing four options: 'Read-only' (selected), 'None', 'Read-only' (disabled), and 'Read/write'. A red box highlights the 'Read-only' option in the dropdown. Below the table, there are buttons for 'Create and attach a new disk' and 'Attach an existing disk'. At the bottom left is an 'Advanced' link, and at the bottom right is a search icon.

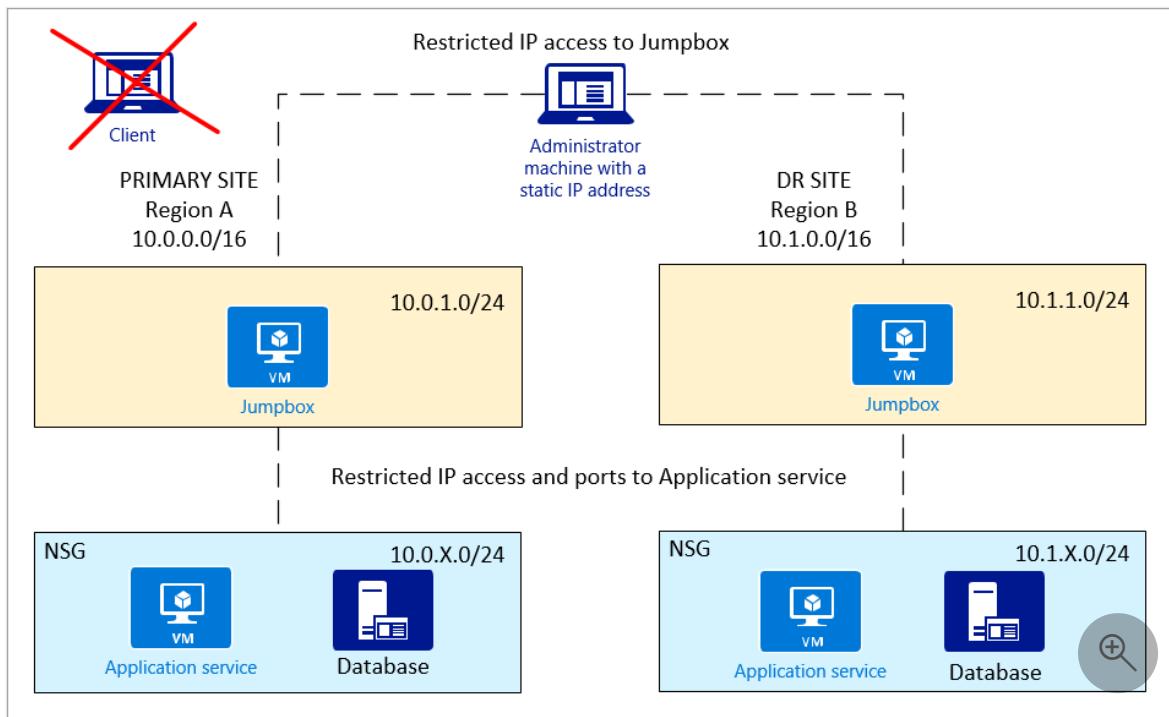
- For operating system disks, use premium SSD with read-write host caching.
- For data disks that contain the following, use premium SSD with read-only host caching: Oracle data files, temp files, control files, block change tracking files, BFILEs, files for external tables, and flashback logs.
- For data disks that contain Oracle online redo log files, use premium SSD or UltraDisk with no host caching, the **None** option. Oracle redo log files that are archived and Oracle Recovery Manager backup sets, can also reside with the online redo log files. Host caching is limited to 4095 GiB, so don't allocate a premium SSD larger than P50 with host caching. If you need more than 4 TiB of storage, stripe several premium SSDs with RAID-0. Use Linux LVM2 or Oracle Automatic Storage Management.

If workloads vary greatly between the day and evening, and the I/O workload can support it, P1-P20 premium SSD with bursting might provide the performance required during night-time batch loads or limited I/O demands.

# Security

After you set up and configure your Azure environment, you need to secure your network. Here are some recommendations:

- **NSG policy:** You can define your NSG by a subnet or a network interface card. It's simpler to control access at the subnet level, both for security and for force-routing application firewalls.
- **Jumpbox:** For more secure access, administrators shouldn't directly connect to the application service or database. Use a jumpbox between the administrator machine and Azure resources.



The administrator machine should only offer IP-restricted access to the jumpbox. The jumpbox should have access to the application and database.

- **Private network (subnets):** It's a good idea to have the application service and database on separate subnets, so that NSG policy can set better control.

## Resources

- [Configure Oracle ASM](#)
- [Configure Oracle Data Guard](#)
- [Configure Oracle GoldenGate](#)
- [Oracle backup and recovery](#)

## Next steps

- Create a complete Linux virtual machine with the Azure CLI
- Explore VM deployment Azure CLI samples ↗

# Backup strategies for Oracle Database on an Azure Linux VM

Article • 07/05/2023

Applies to:  Linux VMs

Database backups help protect the database against data loss that's due to storage component failure and datacenter failure. They can also be a means of recovery from human error and a way to clone a database for development or testing purposes.

In Azure, all storage is highly redundant. The loss of one or more disks won't lead to a database outage. Backups are most often used to protect against human error, to facilitate cloning operations, or to preserve data for regulatory purposes.

Backups also help protect against regional outages when a disaster recovery technology like DataGuard is not in use. In this case, the backups must be stored in different Azure regions via geo-redundant replication, so they're available outside the primary database region.

## Azure Storage

The [Azure Storage services](#) are Microsoft's cloud solution for modern data-storage scenarios. Azure Storage offers services that you can use to mount external storage to an Azure Linux virtual machine (VM), which is suitable as backup media for Oracle Database instances. A backup tool such as Oracle Recovery Manager (RMAN) is required to initiate a backup or restore operation, and to copy the backup to or from Azure Storage.

Azure Storage services offer the following benefits:

- **Durable and highly available.** Redundancy helps keep data safe during transient hardware failures. All storage is triple mirrored by default. You can also opt to replicate data across datacenters or geographical regions for more protection from local catastrophes or natural disasters. Data replicated in this way remains highly available in the event of an unexpected outage.
- **Secure:** Azure Storage encrypts all data written to a storage account. Azure Storage gives you detailed control over who has access to your data.
- **Scalable:** Azure Storage is massively scalable to meet the data storage and performance needs of today's applications.

- **Managed:** Azure handles hardware maintenance, updates, and critical issues for you.
- **Accessible:** Data in Azure Storage is accessible from anywhere in the world over HTTP or HTTPS. Microsoft provides client libraries for Azure Storage in a variety of languages, including .NET, Java, Node.js, Python, PHP, Ruby, and Go. Microsoft also provides a mature REST API.

Azure Storage supports scripting in Azure PowerShell or the Azure CLI. The Azure portal and Azure Storage Explorer offer visual solutions for working with your data.

The Azure Storage platform includes the following data services that are suitable to use as backup media for Oracle Database:

- **Azure Blob Storage:** An object store for text and binary data. It also includes support for big data analytics through Azure Data Lake Storage Gen2.
- **Azure Files:** Managed file shares for cloud or on-premises deployments.
- **Azure Disk Storage:** Block-level storage volumes for Azure VMs.

## Cross-regional storage mounting

The ability to access backup storage across regions is an important aspect of business continuity and disaster recovery (BCDR). It's also useful for cloning databases from backups into different geographical regions. Azure cloud storage provides five levels of [redundancy](#):

- [Locally redundant storage \(LRS\)](#): Your data is replicated three times within a single physical location in the primary region.
- [Zone-redundant storage \(ZRS\)](#): Your data is replicated synchronously across three availability zones in the primary region. LRS helps protect your data in the primary region and helps protect each availability zone.
- [Geo-redundant storage \(GRS\)](#): Your data is replicated asynchronously to a secondary region. LRS helps protect your data in the primary and secondary regions.
- [Geo-zone-redundant storage \(GZRS\)](#): Your data is copied synchronously across three Azure availability zones in the primary region via ZRS. Your data is then copied asynchronously to a single physical location in the secondary region. In all locations, LRS helps protect the data.
- [Read-access geo-redundant storage \(RA-GRS\)](#) and [read-access geo-zone-redundant storage \(RA-GZRS\)](#): You have read-only access to data replicated to the secondary region at all times.

## Blob and file storage

When you're using Azure Files with either the Server Message Block (SMB) protocol or the Network File System (NFS) 4.1 protocol to mount as backup storage, Azure Files doesn't support RA-GRS or RA-GZRS.

If the backup storage requirement is greater than 5 tebibytes (TiB), Azure Files requires you to enable the [large file shares](#) feature. This feature doesn't support GRS or GZRS redundancy. It supports only LRS.

Azure Blob Storage mounted via the NFS 3.0 protocol currently supports only LRS and ZRS redundancy. Azure Blob Storage configured with any redundancy option can be mounted via Blobfuse.

## Recovery Services vault

A Recovery Services vault is a management entity that stores recovery points created over time. It provides an interface to perform backup-related operations. These operations include taking on-demand backups, performing restores, and creating backup policies.

Azure Backup automatically handles storage for the vault. You do need to specify how that storage is replicated at the time of creation. You can't change replication after items are protected in the vault. For regional redundancy, choose the geo-redundant setting.

If you intend to restore to a secondary [Azure paired region](#), enable the [Cross Region Restore](#) feature. When you enable Cross Region Restore, the backup storage is moved from GRS to RA-GRS.

## Azure Blob Storage

Azure Blob Storage is a cloud-based service for storing large amounts of unstructured data and is suitable for Oracle Database backups. You can mount Azure Blob Storage to Azure Linux VMs by using Blobfuse (Linux FUSE) or the NFS v3.0 protocol.

## Blobfuse

[Blobfuse](#) is an open-source project that provides a virtual file system backed by Azure Blob Storage. It uses the libfuse open-source library to communicate with the Linux FUSE kernel module. It implements file-system operations by using the Azure Blob Storage REST APIs.

Blobfuse is currently available for Ubuntu and Centos/RedHat distributions. It's also available for Kubernetes via the [CSI driver](#).

Blobfuse is ubiquitous across Azure regions and works with all storage account types, including general-purpose v1/v2 and Azure Data Lake Storage Gen2. But it doesn't perform as well as alternative protocols. For suitability as the database backup medium, we recommend using the SMB or [NFS](#) protocol to mount Azure Blob Storage.

## NFS v3.0

Azure support for the NFS v3.0 protocol is available. [NFS support](#) enables Windows and Linux clients to mount an Azure Blob Storage container to an Azure VM.

To ensure network security, the storage account that you use for NFS mounting must be contained within a virtual network. Azure Active Directory (Azure AD) security and access control lists (ACLs) are not yet supported in accounts that have NFS 3.0 protocol support enabled on them.

## Azure Files

[Azure Files](#) is a cloud-based, fully managed distributed file system. You can mount it to on-premises or cloud-based Windows, Linux, or macOS clients.

Azure Files offers fully managed cross-platform file shares in the cloud that are accessible via the SMB and NFS protocols. Azure Files doesn't currently support multiple-protocol access, so a share can only be either an NFS share or an SMB share. We recommend determining which protocol best suits your needs before you create Azure file shares.

You can also help protect Azure file shares by using Azure Backup for a Recovery Services vault. This approach provides another layer of protection to the Oracle RMAN backups.

## Azure Files with NFS v4.1

You can mount Azure file shares in Linux distributions by using the NFS v4.1 protocol. There are limitations to supported features. For more information, see [Support for Azure Storage features](#).

Azure NFS file shares are supported in all the same regions that support premium file storage.

For the most up-to-date list, see the [Premium Files Storage](#) entry on the [page for Azure products available by region](#).

## Azure Files with SMB 3.0

You can mount Azure file shares in Linux distributions by using the SMB kernel client. The Common Internet File System (CIFS) protocol, available on Linux distributions, is a dialect of SMB. When you mount an Azure file share on Linux VMs by using SMB, it's mounted as a CIFS-type file system, and the CIFS package must be installed.

The ability to mount Azure file shares via SMB is generally available in all Azure regions. It shows the same performance characteristics as NFS v3.0 and v4.1 protocols, so we currently recommend it as the method to provide backup storage media to Azure Linux VMs.

Two supported versions of SMB are available: SMB 2.1 and SMB 3.0. We recommend SMB 3.0, because it supports encryption in transit. However, Linux kernel versions have differing support for SMB 2.1 and 3.0. To ensure that your application supports SMB 3.0, see [Mount an SMB Azure file share on Linux](#).

Because Azure Files is a multiuser file-share service, you should tune certain characteristics to make it more suitable as backup storage media. We recommend turning off caching and setting the user and group IDs for created files.

## Azure NetApp Files

The [Azure NetApp Files](#) service is a complete storage solution for Oracle Database in Azure VMs. Built on metered file storage, it supports any workload type and is highly available by default. Together with the Oracle Direct NFS driver, Azure NetApp Files provides a highly optimized storage layer for Oracle Database.

Azure NetApp Files provides efficient storage-based snapshots on the underlying storage system that uses a redirect-on-write mechanism. Although snapshots are fast to take and restore, they serve only as a first line of defense. They can account for most of the required restore operations of any organization, which are often part of recovery from human error.

However, snapshots are not a complete backup. To cover all backup and restore requirements, you must create [external snapshot replicas](#) or other [backup vaults](#) in a remote geography to help protect against regional outages. [Read more about how Azure NetApp Files snapshots work](#).

To ensure the creation of database-consistent snapshots, the backup process must be orchestrated between the database and the storage. The Azure Application Consistent Snapshot (AzAcSnap) command-line tool enables data protection for third-party databases. It handles all the orchestration required to put those databases into an application-consistent state before taking a storage snapshot. After that, it returns the databases to an operational state. Oracle Database instances are supported with AzAcSnap since [version 5.1](#).

To learn more about using Azure NetApp Files for Oracle Database on Azure, see [Solution architectures using Azure NetApp Files](#).

## Azure Backup service

[Azure Backup](#) is a fully managed platform as a service (PaaS) solution for backing up your data and recovering it from the Microsoft Azure cloud. Azure Backup can back up and restore on-premises clients, Azure VMs, and Azure file shares. It can also back up SQL Server, Oracle, MySQL, PostgreSQL, and SAP HANA databases on Azure VMs.

Azure Backup provides independent and isolated backups to guard against accidental destruction of original data. Backups are stored in a [Recovery Services vault](#) with built-in management of recovery points.

Azure Backup uses the Azure cloud to deliver high availability with no maintenance or monitoring overhead. It doesn't limit the amount of inbound or outbound data that you transfer, and it doesn't charge for the data that you transfer. Data is secured in transit and at rest.

Azure Backup offers multiple types of replication to keep your backup data highly available:

- [LRS](#) replicates your data three times (that is, it creates three copies of your data) in a storage scale unit in a datacenter.
- [GRS](#) is the default and recommended replication option. GRS replicates your data to a secondary region, hundreds of miles away from the primary location of the source data.

A vault created with GRS redundancy includes the option to configure the [Cross Region Restore](#) feature. You can use this feature to restore data in a secondary Azure paired region.

The Azure Backup service provides a [framework](#) to achieve application consistency during backups of Windows and Linux VMs for various applications like Oracle, MySQL, Mongo DB, SAP HANA, and PostgreSQL: application-consistent snapshots. This

framework involves invoking pre-scripts (to quiesce the applications) before taking a snapshot of disks. It calls post-scripts (commands to unfreeze the applications) after the snapshot is completed, to return the applications to the normal mode.

Although you can find sample pre-scripts and post-scripts on GitHub, you're responsible for creating and maintaining these scripts. In the case of Oracle, the database must be in archive log mode to allow online backups. You must create and maintain the appropriate database beginning and ending backup commands that are run in the pre-scripts and post-scripts.

Azure Backup provides an [enhanced pre-script and post-script framework](#) in which it provides packaged pre-scripts and post-scripts for selected applications. You just name the application, and then Azure Backup automatically invokes the relevant pre-scripts and post-scripts. Microsoft manages the packaged pre-scripts and post-scripts, so you can be assured of the support, ownership, and validity of these scripts.

Currently, the supported applications for the enhanced framework are Oracle 12.1 or later and MySQL. The snapshot is a full copy of the storage and not an incremental or copy-on-write snapshot, so it's an effective medium to restore your database from.

## VLDB considerations

Backup strategies for very large databases (VLDBs) require careful consideration because of their size. Using RMAN to back up to Azure Blob Storage or Azure Files might not provide the required throughput to back up a VLDB in the target time frame.

You can use RMAN incremental backup to reduce backup sizes. This approach might allow Azure Storage to be used as the backup medium for VLDBs. However, it might not be effective for VLDBs that have high volumes of changes.

We recommend using Azure services that provide snapshot capabilities, such as Azure Backup or Azure NetApp Files, for VLDBs. Application-consistent snapshots, where the databases are automatically placed in and out of backup mode, take only seconds to create regardless of the size of the database.

Your backup strategy might be also tied to the overall storage solution that the organization uses for Oracle Database. Database workloads that have extreme I/O throughput often use Azure NetApp Files or third-party Azure Marketplace solutions such as Silk to underpin the database storage throughput and IOPS requirements. These solutions also provide application-consistent snapshots for fast database backup and restore operations.

## Next steps

- Create an Oracle Database instance on an Azure VM
- Back up Oracle Database to Azure Files
- Back up Oracle Database by using the Azure Backup service

# Back up and recover Oracle Database on an Azure Linux VM by using Azure Files

Article • 07/05/2023

Applies to:  Linux VMs

This article demonstrates the use of Azure Files as a medium to back up and restore an Oracle database running on an Azure virtual machine (VM). The steps in this article have been tested against Oracle 12.1 and later.

In this article, you use Oracle Recovery Manager (RMAN) to back up the database to an Azure file share mounted to a VM via the Server Message Block (SMB) protocol. Using Azure Files for backup media is cost-effective and performant. However, for very large databases, Azure Backup provides a better solution.

## Prerequisites

- Use the Bash environment in [Azure Cloud Shell](#). For more information, see [Quickstart for Bash in Azure Cloud Shell](#).

 [Launch Cloud Shell](#) 

- If you prefer to run CLI reference commands locally, [install](#) the Azure CLI. If you're running on Windows or macOS, consider running Azure CLI in a Docker container. For more information, see [How to run the Azure CLI in a Docker container](#).
  - If you're using a local installation, sign in to the Azure CLI by using the `az login` command. To finish the authentication process, follow the steps displayed in your terminal. For other sign-in options, see [Sign in with the Azure CLI](#).
  - When you're prompted, install the Azure CLI extension on first use. For more information about extensions, see [Use extensions with the Azure CLI](#).
  - Run `az version` to find the version and dependent libraries that are installed. To upgrade to the latest version, run `az upgrade`.
- To perform the backup and recovery process, you must first create a Linux VM that has an installed instance of Oracle Database. We recommend using Oracle 12.x or later.
- Create an Oracle Database instance by following the steps in [Create an Oracle Database instance in an Azure VM](#).

# Prepare the database environment

1. To create a Secure Shell (SSH) session with the VM, use the following command. Replace <publicIpAddress> with the public address value for your VM.

```
Bash
```

```
ssh azureuser@<publicIpAddress>
```

2. Switch to the root user:

```
Bash
```

```
sudo su -
```

3. Add the `oracle` user to the `/etc/sudoers` file:

```
Bash
```

```
echo "oracle    ALL=(ALL)      NOPASSWD: ALL" >> /etc/sudoers
```

4. This step assumes that you have an Oracle instance (`test`) that's running on a VM named `vmoracle19c`.

Switch to the `oracle` user:

```
Bash
```

```
sudo su - oracle
```

5. Before you connect, set the environment variable `ORACLE_SID`:

```
Bash
```

```
export ORACLE_SID=test;
```

You should also add the `ORACLE_SID` variable to the `oracle` user's `.bashrc` file for future sign-ins by using the following command:

```
Bash
```

```
echo "export ORACLE_SID=test" >> ~oracle/.bashrc
```

6. Start the Oracle listener if it isn't already running:

```
Bash
```

```
lsnrctl start
```

The output should look similar to the following example:

```
Bash
```

```
LSNRCTL for Linux: Version 19.0.0.0.0 - Production on 18-SEP-2020  
03:23:49  
  
Copyright (c) 1991, 2019, Oracle. All rights reserved.  
  
Starting /u01/app/oracle/product/19.0.0/dbhome_1/bin/tnslsnr: please  
wait...  
  
TNSLSNR for Linux: Version 19.0.0.0 - Production  
System parameter file is  
/u01/app/oracle/product/19.0.0/dbhome_1/network/admin/listener.ora  
Log messages written to  
/u01/app/oracle/diag/tnslsnr/vmoracle19c/listener/alert/log.xml  
Listening on: (DESCRIPTION=(ADDRESS=(PROTOCOL=tcp)  
(HOST=vmoracle19c.eastus.cloudapp.azure.com)(PORT=1521)))  
Listening on: (DESCRIPTION=(ADDRESS=(PROTOCOL=ipc)(KEY=EXTPROC1521)))  
  
Connecting to (DESCRIPTION=(ADDRESS=(PROTOCOL=TCP)  
(HOST=vmoracle19c.eastus.cloudapp.azure.com)(PORT=1521)))  
STATUS of the LISTENER  
-----  
Alias LISTENER  
Version TNSLSNR for Linux: Version 19.0.0.0.0 -  
Production  
Start Date 18-SEP-2020 03:23:49  
Uptime 0 days 0 hr. 0 min. 0 sec  
Trace Level off  
Security ON: Local OS Authentication  
SNMP OFF  
Listener Parameter File  
/u01/app/oracle/product/19.0.0/dbhome_1/network/admin/listener.ora  
Listener Log File  
/u01/app/oracle/diag/tnslsnr/vmoracle19c/listener/alert/log.xml  
Listening Endpoints Summary...  
(DESCRIPTION=(ADDRESS=(PROTOCOL=tcp)  
(HOST=vmoracle19c.eastus.cloudapp.azure.com)(PORT=1521)))  
(DESCRIPTION=(ADDRESS=(PROTOCOL=ipc)(KEY=EXTPROC1521)))  
The listener supports no services  
The command completed successfully
```

7. Create the location for the fast recovery area:

Bash

```
mkdir /u02/fast_recovery_area
```

8. Connect to the database:

Bash

```
sqlplus / as sysdba
```

9. Start the database if it isn't already running:

Bash

```
SQL> startup
```

10. Set database environment variables for the fast recovery area:

Bash

```
SQL> alter system set db_recovery_file_dest_size=4096M scope=both;
SQL> alter system set db_recovery_file_dest='/u02/fast_recovery_area'
scope=both;
```

11. Make sure the database is in ARCHIVELOG mode to enable online backups.

Check the log archive status:

Bash

```
SQL> SELECT log_mode FROM v$database;
```

```
LOG_MODE
```

```
-----
```

```
NOARCHIVELOG
```

If the log archive is in NOARCHIVELOG mode, run the following commands in SQL

Plus:

Bash

```
SQL> SHUTDOWN IMMEDIATE;
SQL> STARTUP MOUNT;
SQL> ALTER DATABASE ARCHIVELOG;
```

```
SQL> ALTER DATABASE OPEN;
SQL> ALTER SYSTEM SWITCH LOGFILE;
```

12. Create a table to test the backup and restore operations:

Bash

```
SQL> create user scott identified by tiger quota 100M on users;
SQL> grant create session, create table to scott;
SQL> connect scott/tiger
SQL> create table scott_table(col1 number, col2 varchar2(50));
SQL> insert into scott_table VALUES(1, 'Line 1');
SQL> commit;
SQL> quit
```

## Back up to Azure Files

To back up to Azure Files, complete these steps:

1. [Set up Azure Files.](#)
2. [Mount the Azure file share to your VM.](#)
3. [Back up the database.](#)
4. [Restore and recover the database.](#)

## Set up Azure Files

In this section, you back up the Oracle database to Azure Files by using Oracle RMAN. Azure file shares are fully managed file shares that stay in the cloud. You can access them by using either the SMB protocol or the Network File System (NFS) protocol.

The following procedures cover creating a file share that uses the SMB protocol to mount to your VM. For information about how to mount by using NFS, see [Create an NFS share](#).

When you're mounting the Azure file share, use the `cache=none` option to disable caching of file share data. To ensure that the `oracle` user owns the files created in the share, set the `uid=oracle` and `gid=oinstall` options.

Portal

Set up your storage account:

1. In the Azure portal, select **+ Create a resource**, and then search for and select **Storage Account**.

The screenshot shows the Microsoft Azure (Preview) portal's 'New' blade. At the top, there's a search bar with the text 'Storage account'. Below the search bar, the 'Azure Marketplace' section is visible, with tabs for 'See all' and 'Popular'. Under the 'Popular' tab, several services are listed with their icons and names: Windows Server 2016 Datacenter, Ubuntu Server 18.04 LTS, AI + Machine Learning, Web App, Blockchain, Compute, Containers, SQL Database, Databases, Developer Tools, Function App, DevOps, Identity, Integration, Azure Cosmos DB, Internet of Things, IT & Management Tools, Kubernetes Service, Media, Migration, Mixed Reality, DevOps Starter, Monitoring & Diagnostics, Networking, Storage account, Security, Virtual machines, Time Series Insights env..., Stream Analytics jobs, Storage accounts, Data factories, Function App, IoT Hub, Load balancers, Network security groups, Virtual networks, and Azure Active Directory.

2. On the **Create storage account** pane:
  - a. For **Resource group**, select your existing resource group, **rg-oracle**.
  - b. For **Storage account name**, enter **oracbkup1**.
  - c. Ensure that **Location** is set to the same region as all your other resources in the resource group.
  - d. Set **Performance** to **Standard**.
  - e. For **Account kind**, select **StorageV2 (general purpose v2)**.
  - f. For **Replication**, select **Locally-redundant storage (LRS)**.

## Create storage account

[Basics](#) [Networking](#) [Data protection](#) [Advanced](#) [Tags](#) [Review + create](#)

Azure Storage is a Microsoft-managed service providing cloud storage that is highly available, secure, durable, scalable, and redundant. Azure Storage includes Azure Blobs (objects), Azure Data Lake Storage Gen2, Azure Files, Azure Queues, and Azure Tables. The cost of your storage account depends on the usage and the options you choose below.

[Learn more about Azure storage accounts](#)

### Project details

Select the subscription to manage deployed resources and costs. Use resource groups like folders to organize and manage all your resources.

Subscription *	<input type="text"/>
Resource group *	<input type="text"/> rg-oracle <a href="#">Create new</a>

### Instance details

The default deployment model is Resource Manager, which supports the latest Azure features. You may choose to deploy using the classic deployment model instead. [Choose classic deployment model](#)

Storage account name * ⓘ	<input type="text"/> orabkup1
Location *	<input type="text"/> (US) East US
Performance ⓘ	<input checked="" type="radio"/> Standard <input type="radio"/> Premium
Account kind ⓘ	<input type="text"/> StorageV2 (general purpose v2)
Replication ⓘ	<input type="text"/> Locally-redundant storage (LRS)

[Review + create](#)[< Previous](#)[Next : Networking >](#)

### 3. Select the Advanced tab. Under Azure Files, set Large file shares to Enabled.

Select **Review + Create**, and then select **Create**.

## Create storage account

[Basics](#) [Networking](#) [Data protection](#) [Advanced](#) [Tags](#) [Review + create](#)

### Security

Secure transfer required ⓘ	<input type="radio"/> Disabled <input checked="" type="radio"/> Enabled
Minimum TLS version ⓘ	<input type="text"/> Version 1.2
Infrastructure encryption ⓘ	<input type="radio"/> Disabled <input checked="" type="radio"/> Enabled

ⓘ Sign up is currently required to enable infrastructure encryption on a per-subscription basis. [Sign up for infrastructure encryption](#)

### Blob storage

Allow Blob public access ⓘ	<input type="radio"/> Disabled <input checked="" type="radio"/> Enabled
Blob access tier (default) ⓘ	<input type="radio"/> Cool <input checked="" type="radio"/> Hot
NFS v3 ⓘ	<input type="radio"/> Disabled <input checked="" type="radio"/> Enabled

ⓘ The current combination of storage account kind, performance, replication, and location does not support the NFS v3 feature. [Learn more about NFS v3](#)

### Data Lake Storage Gen2

Hierarchical namespace ⓘ	<input checked="" type="radio"/> Disabled <input type="radio"/> Enabled
--------------------------	---

### Azure Files

Large file shares ⓘ	<input type="radio"/> Disabled <input checked="" type="radio"/> Enabled
---------------------	---

ⓘ Large file share storage accounts do not have the ability to convert to geo-redundant storage offerings and upgrade is permanent.

[Review + create](#)[< Previous](#)[Next : Tags >](#)

- When the storage account is created, go to the resource and select **File shares**.

- Select **+ File share**, and then on the **New file share** panel:

- For **Name**, enter **orabkup1**.
- Set **Quota** to **10240 gibibytes (GiB)**.

The quota reflects an upper boundary that the file share can grow to. Because you're using standard storage in this example, resources are pay-as-you-go and not provisioned. So, setting the quota to 10 tebibytes (TiB) won't incur costs beyond what you use. If your backup strategy requires more storage, set the quota to an appropriate level to hold all backups.

- Under **Tiers**, select **Transaction optimized**.
- Select **Create**.

Dashboard > Resource groups > rg-oracle > orabkup1

## orabkup1 | File shares

Storage account

File share settings

Active Directory: Not configured   Soft delete: Disabled   Share capacity: 100 TiB

Search file shares by prefix (case-sensitive)

Name	Modified	Tier
You don't have any file shares yet. Click '+ File share' to get started.		

**Tiers**

- Premium
- Transaction optimized
- Hot
- Cool

**Create** **Discard**

6. When the file share is created, select **orabkup1** on the **File share settings** pane.
7. Select the **Connect** tab to open the **Connect** panel, and then select the **Linux** tab. Copy the provided commands to mount the file share by using the SMB protocol.

Dashboard > Resource groups > rg-oracle > orabkup1 >

## orabkup1

File share

Connect

Secure transfer required is enabled on the storage account. SMB clients must support 3.0 encryption to connect. Click here to learn more about connecting Azure files.

Windows   **Linux**   macOS

Mount point

orabkup1

To connect to this file share from a Linux computer, run this command:

```
sudo mkdir /mnt/orabkup1
if [ ! -d "/etc/smbcredentials" ]; then
    sudo mkdir /etc/smbcredentials
fi
if [ ! -f "/etc/smbcredentials/orabkup1.cred" ]; then
    sudo bash -c 'echo "username=orabkup1" >> /etc/smbcredentials/orabkup1.cred'
```

In order to mount an Azure file share outside of the Azure region it is hosted in, such as on-premises or in a different Azure region, the OS must support the encryption functionality of SMB 3.0.

[Learn more about Azure File Storage with Linux](#)

## Mount the Azure file share to your VM

1. Create the mount point:

Bash

```
sudo mkdir /mnt/orabackup
```

2. Set up credentials:

Bash

```
if [ ! -d "/etc/smbcredentials" ]; then
    sudo mkdir /etc/smbcredentials
fi
```

3. Run the following command. Substitute <Your Storage Account Key1> with the storage account key that you retrieved earlier.

Bash

```
if [ ! -f "/etc/smbcredentials/orabackup1.cred" ]; then
    sudo bash -c 'echo "username=orabackup1" >>
/etc/smbcredentials/orabackup1.cred'
    sudo bash -c 'echo "password=<Your Storage Account Key1>" >>
/etc/smbcredentials/orabackup1.cred'
fi
```

4. Change permissions on the credentials file:

Bash

```
sudo chmod 600 /etc/smbcredentials/orabackup1.cred
```

5. Add the mount to the */etc/fstab* file:

Bash

```
sudo bash -c 'echo "//orabackup1.file.core.windows.net/orabackup
/mnt/orabackup cifs
nofail,vers=3.0,credentials=/etc/smbcredentials/orabackup1.cred,dir_mod
e=0777,file_mode=0777,serverino,cache=none,uid=oracle,gid=oinstall" >>
/etc/fstab'
```

6. Run the commands to mount the Azure file share by using the SMB protocol:

Bash

```
sudo mount -t cifs //orabackup1.file.core.windows.net/orabackup
/mnt/orabackup -o
```

```
vers=3.0,credentials=/etc/smbcredentials/orabackup1.cred,dir_mode=0777,  
file_mode=0777,serverino,cache=none,uid=oracle,gid=oinstall
```

If you get an error similar to the following example, the Common Internet File System (CIFS) package might not be installed on your Linux host:

Output

```
mount: wrong fs type, bad option, bad superblock on  
//orabackup1.file.core.windows.net/orabackup
```

To check if the CIFS package is installed, run the following command:

Bash

```
sudo rpm -qa|grep cifs-utils
```

If the command returns no output, install the CIFS package by using the following command. Then rerun the `mount` command to mount the Azure file share.

Bash

```
sudo yum install cifs-utils
```

7. Check that the file share is mounted properly by using the following command:

Bash

```
df -h
```

The output should look similar to this example:

Output

```
$ df -h  
Filesystem          Size  Used Avail Use%  
Mounted on  
devtmpfs            3.3G   0    3.3G  0%  
/dev  
tmpfs              3.3G   0    3.3G  0%  
/dev/shm  
tmpfs              3.3G  17M  3.3G  1%  
/run  
tmpfs              3.3G   0    3.3G  0%  
/sys/fs/cgroup  
/dev/sda2           30G  9.1G  19G  34% /
```

/dev/sdc1	59G	2.7G	53G	5%
/u02	497M	199M	298M	41%
/dev/sda1	495M	9.7M	486M	2%
/boot	671M	0	671M	0%
/dev/sda15	14G	2.1G	11G	16%
/boot/efi	671M	0	671M	0%
tmpfs	671M	0	671M	0%
/run/user/54321	671M	0	671M	0%
/dev/sdb1	10T	0	10T	0%
/mnt/resource				
tmpfs				
/run/user/54322				
//orabackup1.file.core.windows.net/orabackup				
/mnt/orabackup				

## Back up the database

In this section, you use Oracle RMAN to take a full backup of the database and archive logs. You then write the backup as a backup set to the Azure file share that you mounted earlier.

1. Configure RMAN to back up to the Azure Files mount point:

```
Bash

rman target /
RMAN> configure snapshot controlfile name to
' /mnt/orabkup/snapcf_ev.f';
RMAN> configure channel 1 device type disk format
' /mnt/orabkup/%d/Full_%d_%U_%T_%s';
RMAN> configure channel 2 device type disk format
' /mnt/orabkup/%d/Full_%d_%U_%T_%s';
```

2. In this example, you're limiting the size of RMAN backup pieces to 4 GiB. However, the RMAN backup `maxpiecesize` value can go up to 4 TiB, which is the file size limit for Azure standard file shares and premium file shares. For more information, see [Azure Files scalability and performance targets](#).

```
Bash

RMAN> configure channel device type disk maxpiecesize 4000G;
```

3. Confirm the configuration change details:

```
Bash
```

```
RMAN> show all;
```

4. Run the backup. The following command takes a full database backup, including archive log files, as a backup set in compressed format:

```
Bash
```

```
RMAN> backup as compressed backupset database plus archivelog;
```

You've now backed up the database online by using Oracle RMAN, with the backup located in Azure Files. Because you're storing the backups in Azure Files, you can access them from other VMs if you need to clone the database.

Although using RMAN and Azure Files for database backup has many advantages, backup and restore time is linked to the size of the database. For very large databases, these operations can take considerable time.

An alternative is to use application-consistent VM backups through Azure Backup. This mechanism uses snapshot technology to perform fast backups irrespective of database size. Integration with a Recovery Services vault provides cloud storage of your Oracle Database backups, so you can access them from other VMs and other Azure regions.

## Restore and recover the database

1. Shut down the Oracle instance:

```
Bash
```

```
sqlplus / as sysdba
SQL> shutdown abort
ORACLE instance shut down.
```

2. Remove the database datafiles:

```
Bash
```

```
cd /u02/oradata/TEST
rm -f *.dbf
```

3. The following commands use RMAN to restore the missing datafiles and recover the database:

```
Bash
```

```
rman target /  
RMAN> startup mount;  
RMAN> restore database;  
RMAN> recover database;  
RMAN> alter database open;
```

4. Check that the database content is fully recovered:

Bash

```
RMAN> SELECT * FROM scott.scott_table;
```

The backup and recovery of the Oracle Database 19c database on an Azure Linux VM are now finished.

## Delete the VM

When you no longer need the VM, you can use the following command to remove the resource group, the VM, and all related resources:

Azure CLI

```
az group delete --name rg-oracle
```

## Next steps

[Create highly available VMs](#)

[Explore Azure CLI samples for VM deployment ↗](#)

# Back up and recover Oracle Database on an Azure Linux VM by using Azure Backup

Article • 07/05/2023

Applies to:  Linux VMs

This article demonstrates the use of Azure Backup to take disk snapshots of virtual machine (VM) disks, which include the Oracle Database files and the Oracle fast recovery area. By using Azure Backup, you can take full disk snapshots that are suitable as backups and are stored in a [Recovery Services vault](#).

Azure Backup also provides application-consistent backups, which ensure that more fixes aren't required to restore the data. Application-consistent backups work with both file system and Oracle Automatic Storage Management (ASM) databases.

Restoring application-consistent data reduces restoration time, so you can quickly return to a running state. Oracle Database recovery is still necessary after restore. You facilitate the recovery by using Oracle archived redo log files that are captured and stored in a separate Azure file share.

This article walks you through the following tasks:

- ✓ Back up the database with application-consistent backup.
- ✓ Restore and recover the database from a recovery point.
- ✓ Restore the VM from a recovery point.

## Prerequisites

- Use the Bash environment in [Azure Cloud Shell](#). For more information, see [Quickstart for Bash in Azure Cloud Shell](#).

 [Launch Cloud Shell](#) 

- If you prefer to run CLI reference commands locally, [install](#) the Azure CLI. If you're running on Windows or macOS, consider running Azure CLI in a Docker container. For more information, see [How to run the Azure CLI in a Docker container](#).
  - If you're using a local installation, sign in to the Azure CLI by using the [az login](#) command. To finish the authentication process, follow the steps displayed in

your terminal. For other sign-in options, see [Sign in with the Azure CLI](#).

- When you're prompted, install the Azure CLI extension on first use. For more information about extensions, see [Use extensions with the Azure CLI](#).
- Run `az version` to find the version and dependent libraries that are installed. To upgrade to the latest version, run `az upgrade`.
- To perform the backup and recovery process, you must first create a Linux VM that has an installed instance of Oracle Database 12.1 or later.
- Create an Oracle Database instance by following the steps in [Create an Oracle Database instance in an Azure VM](#).

## Prepare the environment

To prepare the environment, complete these steps:

1. [Connect to the VM](#).
2. [Set up Azure Files storage](#).
3. [Prepare the databases](#).

## Connect to the VM

1. To create a Secure Shell (SSH) session with the VM, use the following command.

Replace `<publicIpAddress>` with the public address value for your VM.

```
Bash
```

```
ssh azureuser@<publicIpAddress>
```

2. Switch to the root user:

```
Bash
```

```
sudo su -
```

3. Add the `oracle` user to the `/etc/sudoers` file:

```
Bash
```

```
echo "oracle    ALL=(ALL)      NOPASSWD: ALL" >> /etc/sudoers
```

# Set up Azure Files storage for the Oracle archived redo log files

The Oracle Database instance's archived redo log files play a crucial role in database recovery. They store the committed transactions needed to roll forward from a database snapshot taken in the past.

When the database is in `ARCHIVELOG` mode, it archives the contents of online redo log files when they become full and switch. Together with a backup, they're required to achieve point-in-time recovery when the database is lost.

Oracle provides the capability to archive redo log files to different locations. The industry best practice is that at least one of those destinations should be on remote storage, so it's separate from the host storage and protected with independent snapshots. Azure Files meets those requirements.

An Azure file share is storage that you can attach to a Linux or Windows VM as a regular file-system component, by using the Server Message Block (SMB) or Network File System (NFS) protocol. To set up an Azure file share on Linux (by using the SMB 3.0 protocol) for use as archive log storage, see [Mount an SMB Azure file share on Linux](#). When you complete the setup, return to this guide and complete all remaining steps.

## Prepare the databases

This part of the process assumes that you followed [Create an Oracle Database instance in an Azure VM](#). As a result:

- You have an Oracle instance named `oratest1` that's running on a VM named `vmoracle19c`.
- You're using the standard Oracle `oraenv` script with its dependency on the standard Oracle configuration file `/etc/oratab` to set up environment variables in a shell session.

Perform the following steps for each database on the VM:

1. Switch to the `oracle` user:

```
Bash
```

```
sudo su - oracle
```

- Set the environment variable `ORACLE_SID` by running the `oraenv` script. It will prompt you to enter the `ORACLE_SID` name.

```
Bash
```

```
. oraenv
```

- Add the Azure file share as another destination for database archive log files.

This step assumes that you configured and mounted an Azure file share on the Linux VM. For each database installed on the VM, make a subdirectory that's named after your database security identifier (SID).

In this example, the mount point name is `/backup` and the SID is `oratest1`. So you create the subdirectory `/backup/oratest1` and change ownership to the `oracle` user. Substitute `/backup/SID` for your mount point name and database SID.

```
Bash
```

```
sudo mkdir /backup/oratest1  
sudo chown oracle:oinstall /backup/oratest1
```

- Connect to the database:

```
Bash
```

```
sqlplus / as sysdba
```

- Start the database if it's not already running:

```
Bash
```

```
SQL> startup
```

- Set the first archive log destination of the database to the file-share directory that you created earlier:

```
Bash
```

```
SQL> alter system set log_archive_dest_1='LOCATION=/backup/oratest1'  
scope=both;
```

- Define the recovery point objective (RPO) for the database.

To achieve a consistent RPO, consider the frequency at which the online redo log files will be archived. These factors control the frequency:

- The size of the online redo log files. As an online log file becomes full, it's switched and archived. The larger the online log file, the longer it takes to fill up. The added time decreases the frequency of archive generation.
- The setting of the `ARCHIVE_LAG_TARGET` parameter controls the maximum number of seconds permitted before the current online log file must be switched and archived.

To minimize the frequency of switching and archiving, along with the accompanying checkpoint operation, Oracle online redo log files generally have a large size (for example, 1,024M, 4,096M, or 8,192M). In a busy database environment, logs are still likely to switch and archive every few seconds or minutes. In a less active database, they might go hours or days before the most recent transactions are archived, which would dramatically decrease archival frequency.

We recommend that you set `ARCHIVE_LAG_TARGET` to ensure a consistent RPO. A setting of 5 minutes (300 seconds) is a prudent value for `ARCHIVE_LAG_TARGET`. It ensures that any database recovery operation can recover to within 5 minutes of the time of failure.

To set `ARCHIVE_LAG_TARGET`, run this command:

```
Bash
```

```
SQL> alter system set archive_lag_target=300 scope=both;
```

To better understand how to deploy highly available Oracle Database instances in Azure with zero RPO, see [Reference architectures for Oracle Database](#).

8. Make sure the database is in archive log mode to enable online backups.

Check the log archive status first:

```
Bash
```

```
SQL> SELECT log_mode FROM v$database;  
  
LOG_MODE  
-----  
NOARCHIVELOG
```

If it's in `NOARCHIVELOG` mode, run the following commands:

Bash

```
SQL> SHUTDOWN IMMEDIATE;
SQL> STARTUP MOUNT;
SQL> ALTER DATABASE ARCHIVELOG;
SQL> ALTER DATABASE OPEN;
SQL> ALTER SYSTEM SWITCH LOGFILE;
```

9. Create a table to test the backup and restore operations:

Bash

```
SQL> create user scott identified by tiger quota 100M on users;
SQL> grant create session, create table to scott;
SQL> connect scott/tiger
SQL> create table scott_table(col1 number, col2 varchar2(50));
SQL> insert into scott_table VALUES(1, 'Line 1');
SQL> commit;
SQL> quit
```

## Back up your data by using Azure Backup

The Azure Backup service provides solutions to back up your data and recover it from the Microsoft Azure cloud. Azure Backup provides independent and isolated backups to guard against accidental destruction of original data. Backups are stored in a Recovery Services vault with built-in management of recovery points, so you can restore as needed.

In this section, you use Azure Backup to take application-consistent snapshots of your running VM and Oracle Database instances. The databases are placed into backup mode, which allows a transactionally consistent online backup to occur while Azure Backup takes a snapshot of the VM disks. The snapshot is a full copy of the storage and not an incremental or copy-on-write snapshot. It's an effective medium to restore your database from.

The advantage of using Azure Backup application-consistent snapshots is that they're fast to take, no matter how large your database is. You can use a snapshot for restore operations as soon as you take it, without having to wait for it to be transferred to the Recovery Services vault.

To use Azure Backup to back up the database, complete these steps:

1. [Understand the Azure Backup framework.](#)

2. Prepare the environment for an application-consistent backup.
3. Set up application-consistent backups.
4. Trigger an application-consistent backup of the VM.

## Understand the Azure Backup framework

The Azure Backup service provides a [framework](#) to achieve application consistency during backups of Windows and Linux VMs for various applications. This framework involves invoking a pre-script to quiesce the applications before taking a snapshot of disks. It calls a post-script to unfreeze the applications after the snapshot is completed.

Microsoft has enhanced the framework so that the Azure Backup service provides packaged pre-scripts and post-scripts for selected applications. These pre-scripts and post-scripts are already loaded on the Linux image, so there's nothing for you to install. You just name the application, and then Azure Backup automatically invokes the relevant scripts. Microsoft manages the packaged pre-scripts and post-scripts, so you can be assured of the support, ownership, and validity of them.

Currently, the supported applications for the enhanced framework are Oracle 12.x or later and MySQL. For details, see [Support matrix for managed Azure VM backups](#).

You can author your own scripts for Azure Backup to use with pre-12.x databases. Example scripts are available on [GitHub](#).

Each time you do a backup, the enhanced framework runs the pre-scripts and post-scripts on all Oracle Database instances installed on the VM. The `configuration_path` parameter in the `workload.conf` file points to the location of the Oracle `/etc/oratab` file (or a user-defined file that follows the oratab syntax). For details, see [Set up application-consistent backups](#).

Azure Backup runs the pre-scripts and post-scripts for each database listed in the file that `configuration_path` points to. Exceptions are lines that begin with `#` (treated as comment) or `+ASM` (an Oracle ASM instance).

The Azure Backup enhanced framework takes online backups of Oracle Database instances that operate in `ARCHIVELOG` mode. The pre-scripts and post-scripts use the `ALTER DATABASE BEGIN` and `END BACKUP` commands to achieve application consistency.

For the database backup to be consistent, databases in `NOARCHIVELOG` mode must be shut down cleanly before the snapshot starts.

# Prepare the environment for an application-consistent backup

Oracle Database employs job role separation to provide separation of duties by using least privilege. It associates separate operating system (OS) groups with separate database administrative roles. Users can then have different database privileges granted to them, depending on their membership in OS groups.

The `SYSBACKUP` database role (generic name `OSBACKUPDBA`) provides limited privileges to perform backup operations in the database. Azure Backup requires it.

During Oracle installation, we recommend that you use `backupdba` as the OS group name to associate with the `SYSBACKUP` role. But you can use any name, so you need to determine the name of the OS group that represents the Oracle `SYSBACKUP` role first.

1. Switch to the `oracle` user:

```
Bash  
sudo su - oracle
```

2. Set the Oracle environment:

```
Bash  
export ORACLE_SID=oratest1  
export ORAENV_ASK=NO  
. oraenv
```

3. Determine the name of the OS group that represents the Oracle `SYSBACKUP` role:

```
Bash  
grep "define SS_BKP" $ORACLE_HOME/rdbms/lib/config.c
```

The output looks similar to the following example:

```
Output  
#define SS_BKP_GRP "backupdba"
```

In the output, the value enclosed within double quotation marks is the name of the Linux OS group to which the Oracle `SYSBACKUP` role is externally authenticated. In

this example, it's `backupdba`. Note down the actual value.

4. Verify that the OS group exists by running the following command. Substitute `<group name>` with the value that the previous command returned (without the quotation marks).

Bash

```
grep <group name> /etc/group
```

The output looks similar to the following example:

Output

```
backupdba:x:54324:oracle
```

### Important

If the output doesn't match the Oracle OS group value that you retrieved in step 3, use the following command to create the OS group that represents the Oracle `SYSBACKUP` role. Substitute `<group name>` with the group name that you retrieved in step 3.

Bash

```
sudo groupadd <group name>
```

5. Create a new backup user named `azbackup` that belongs to the OS group that you verified or created in the previous steps. Substitute `<group name>` with the name of the verified group. The user is also added to the `oinstall` group to enable it to open ASM disks.

Bash

```
sudo useradd -g <group name> -G oinstall azbackup
```

6. Set up external authentication for the new backup user.

The backup user `azbackup` needs to be able to access the database by using external authentication, so it isn't challenged by a password. To enable this access,

you must create a database user that authenticates externally through `azbackup`.

The database uses a prefix for the user name, which you need to find.

Perform the following steps for each database installed on the VM:

- Log in to the database by using SQL Plus, and check the default settings for external authentication:

Bash

```
sqlplus / as sysdba
SQL> show parameter os_authent_prefix
SQL> show parameter remote_os_authent
```

The output should look like this example, which shows `ops$` as the database username prefix:

Output

NAME	TYPE	VALUE
os_authent_prefix	string	ops\$
remote_os_authent	boolean	FALSE

- Create a database user named `ops$azbackup` for external authentication to the `azbackup` user, and grant `SYSBACKUP` privileges:

Bash

```
SQL> CREATE USER ops$azbackup IDENTIFIED EXTERNALLY;
SQL> GRANT CREATE SESSION, ALTER SESSION, SYSBACKUP TO ops$azbackup;
```

- If you receive the error `ORA-46953: The password file is not in the 12.2 format` when you run the `GRANT` statement, follow these steps to migrate the `orapwd` file to 12.2 format. Perform these steps for every Oracle Database instance on the VM.

- Exit SQL Plus.
- Move the password file with the old format to a new name.
- Migrate the password file.
- Remove the old file.
- Run the following commands:

Bash

```
mv $ORACLE_HOME/dbs/orapworatest1 $ORACLE_HOME/dbs/orapworatest1.tmp  
orapwd file=$ORACLE_HOME/dbs/orapworatest1  
input_file=$ORACLE_HOME/dbs/orapworatest1.tmp  
rm $ORACLE_HOME/dbs/orapworatest1.tmp
```

f. Rerun the `GRANT` operation in SQL Plus.

8. Create a stored procedure to log backup messages to the database alert log. Use the following code for each database installed on the VM:

Bash

```
sqlplus / as sysdba  
SQL> GRANT EXECUTE ON DBMS_SYSTEM TO SYSBACKUP;  
SQL> CREATE PROCEDURE sysbackup.azmessage(in_msg IN VARCHAR2)  
AS  
    v_timestamp      VARCHAR2(32);  
BEGIN  
    SELECT TO_CHAR(SYSDATE, 'YYYY-MM-DD HH24:MI:SS')  
    INTO v_timestamp FROM DUAL;  
    DBMS_OUTPUT.PUT_LINE(v_timestamp || ' - ' || in_msg);  
    SYS.DBMS_SYSTEM.KSDWRT(SYS.DBMS_SYSTEM.ALERT_FILE, in_msg);  
END azmessage;  
/  
SQL> SHOW ERRORS  
SQL> QUIT
```

## Set up application-consistent backups

1. Switch to the root user:

Bash

```
sudo su -
```

2. Check for the `/etc/azure` folder. If it isn't present, create the working directory for the application-consistent backup:

Bash

```
if [ ! -d "/etc/azure" ]; then  
    mkdir /etc/azure  
fi
```

3. Check for the `workload.conf` file within the folder. If it isn't present, create it in the `/etc/azure` directory and give it the following contents. The comments must begin with `[workload]`. If the file is already present, just edit the fields so that they match the following contents. Otherwise, the following command creates the file and populates the contents:

```
Bash
```

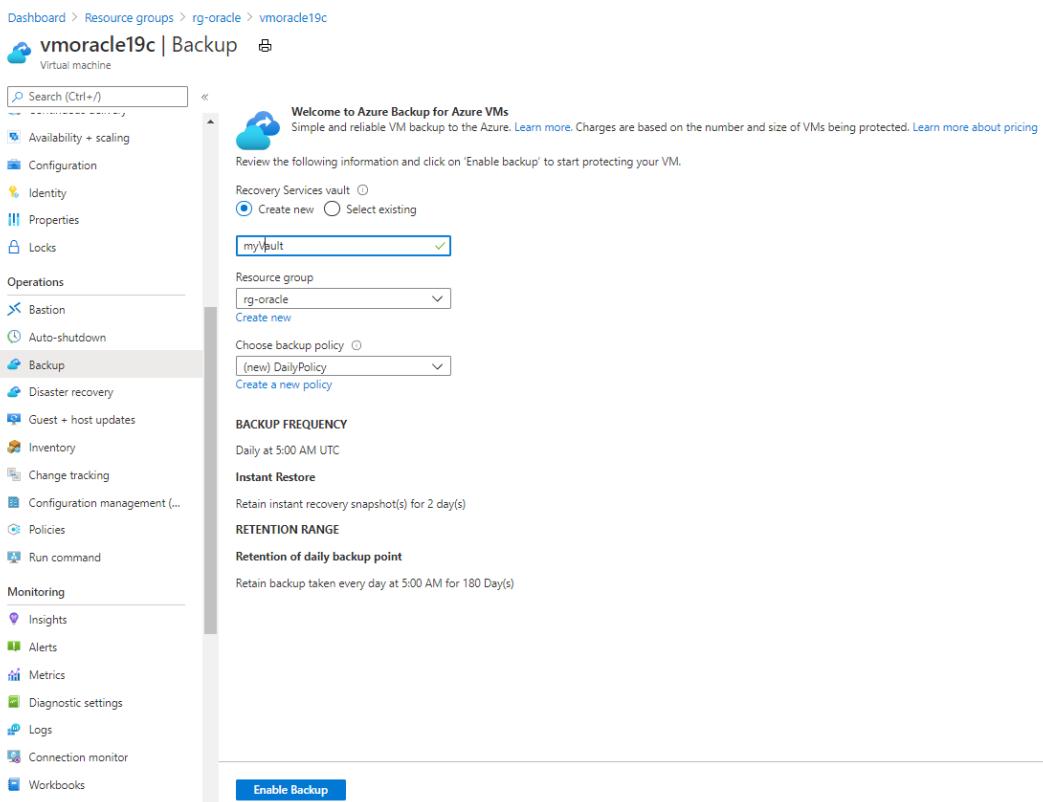
```
echo "[workload]
workload_name = oracle
configuration_path = /etc/oratab
timeout = 90
linux_user = azbackup" > /etc/azure/workload.conf
```

The `workload.conf` file uses the following format:

- The `workload_name` parameter indicates the database workload type. In this case, setting the parameter to `Oracle` allows Azure Backup to run the correct pre-scripts and post-scripts (consistency commands) for Oracle Database instances.
- The `timeout` parameter indicates the maximum time, in seconds, that each database must complete storage snapshots.
- The `linux_user` parameter indicates the Linux user account that Azure Backup uses to run database quiesce operations. You created this user, `azbackup`, previously.
- The `configuration_path` parameter indicates the absolute path name for a text file on the VM. Each line lists a database instance running on the VM. This is typically the `/etc/oratab` file that Oracle generates during database installation, but it can be any file with any name that you choose. It must follow these format rules:
  - The file is a text file. Each field is delimited with the colon character (`:`).
  - The first field in each line is the name for an `ORACLE_SID` instance.
  - The second field in each line is the absolute path name for `ORACLE_HOME` for that `ORACLE_SID` instance.
  - All text after the first two fields is ignored.
  - If the line starts with a pound sign (#), the entire line is ignored as a comment.
  - If the first field has the value `+ASM`, denoting an Oracle ASM instance, it's ignored.

## Trigger an application-consistent backup of the VM

1. In the Azure portal, go to your **rg-oracle** resource group and select your **vmoracle19c** virtual machine.
2. On the **Backup** pane:
  - a. Under **Recovery Services vault**, select **Create new**.
  - b. For the name of the vault, use **myVault**.
  - c. For **Resource group**, select **rg-oracle**.
  - d. For **Choose backup policy**, use **(new) DailyPolicy**. If you want to change the backup frequency or retention range, select **Create a new policy** instead.



3. Select **Enable Backup**.

The backup process doesn't start until the scheduled time expires. To set up an immediate backup, complete the next step.

4. From the resource group pane, select your newly created Recovery Services vault named **myVault**. You might need to refresh the page to see it.
5. On the **myVault - Backup items** pane, under **BACkUP ITEM COUNT**, select the backup item count.

Dashboard > Resource groups > rg-oracle > myVault

## myVault | Backup items

Recovery Services vault

Search (Ctrl+ /) Refresh

Overview Activity log Access control (IAM) Tags Diagnose and solve problems Settings Identity Private endpoint connections Properties Locks Getting started Backup Site Recovery Protected items Backup items Replicated items Manage Backup policies

BACKUP MANAGEMENT TYPE	BACKUP ITEM COUNT
Azure Virtual Machine	1
SAP HANA in Azure VM	0
SQL in Azure VM	0
Azure Storage (Azure Files)	0
DPM	0
Azure Backup Server	0
Azure Backup Agent	0

6. On the **Backup Items (Azure Virtual Machine)** pane, select the ellipsis (...) button, and then select **Backup now**.

Dashboard > Resource groups > rg-oracle > myVault >

## Backup Items (Azure Virtual Machine)

myVault

Refresh Add Filter

Fetching data from service completed.

Filter items ...

Name	Resource Group	Backup Pre-Check	Last Backup Status	Latest restore point
vmoracle19c	rg-oracle	Passed	Warning(Initial backup pending)	

Pin to dashboard Backup now Restore VM File Recovery Stop backup Delete backup data Undelete

7. Accept the default **Retain Backup Till** value, and then select **OK**. Wait for the backup process to finish.

8. To view the status of the backup job, select **Backup Jobs**.

The screenshot shows the 'myVault' backup jobs page. The left sidebar has sections for Manage, Monitoring, and Automation. Under Monitoring, 'Backup Jobs' is selected. The main area shows a table of backup jobs:

Workload name	Operation	Status	Type	Start time	Duration
vmoracle19c	Backup	In progress	Azure virtual machine	10/20/2020, 10:55:43 ...	01:40:00
vmoracle19c	Configure backup	Completed	Azure virtual machine	10/20/2020, 10:35:40 ...	00:00:30

Select the backup job to see details about its status.

The screenshot shows the 'Backup' details page for 'vmoracle19c'. It includes sections for Job Details and Sub Tasks.

VM Name	Recovery Point Expiry Time in UTC	Activity ID
vmoracle19c	11/19/2020 12:59:59 PM	e6539e15-1e2a-4e91-98d3-fc9b13fb2049-2020-10-19T23:55:23Z-lbz

Sub Tasks:

Name	Status
Take Snapshot	Completed
Transfer data to vault	In progress

Although it takes only seconds to execute the snapshot, it can take some time to transfer it to the vault. The backup job is not completed until the transfer is finished.

9. For an application-consistent backup, address any errors in the log file at `/var/log/azure/Microsoft.Azure.RecoveryServices.VMSnapshotLinux/extension.log`.

## Restore the VM

Restoring an entire VM means that you restore the VM and its attached disks to a new VM from a selected restore point. This action also restores all databases that run on the VM. Afterward, you need to recover each database.

To restore an entire VM, complete these steps:

1. Stop and delete the VM.
2. Recover the VM.
3. Set the public IP address.
4. Recover the database.

There are two main choices when you're restoring a VM:

- Restore the VM from which the backups were originally taken.
- Restore (clone) a new VM without affecting the VM from which the backups were originally taken.

The first steps in this exercise (stopping, deleting, and then recovering the VM) simulate the first use case.

## Stop and delete the VM

Portal

1. In the Azure portal, go to the **vmoracle19c** virtual machine, and then select **Stop**.

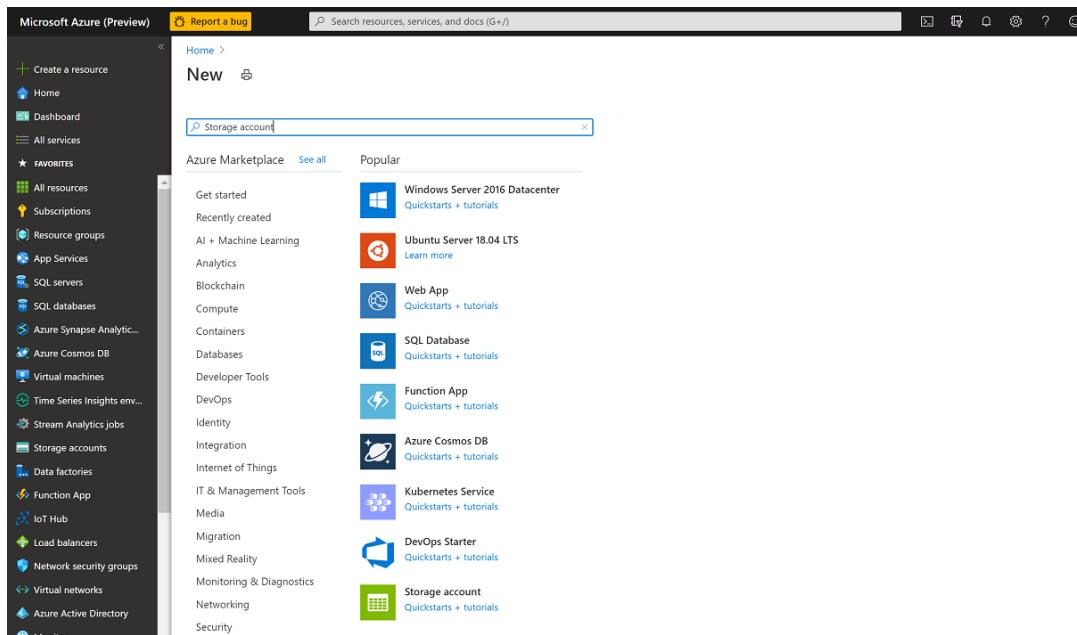
2. When the virtual machine is no longer running, select **Delete** and then **Yes**.

## Recover the VM

Portal

1. Create a storage account for staging in the Azure portal:

- a. In the Azure portal, select **+ Create a resource**, and then search for and select **Storage account**.



b. On the **Create storage account** pane:

- i. For **Resource group**, select your existing resource group, **rg-oracle**.
- ii. For **Storage account name**, enter **oracrestore**.
- iii. Ensure that **Location** is set to the same region as all your other resources in the resource group.
- iv. Set **Performance** to **Standard**.
- v. For **Account kind**, select **StorageV2 (general purpose v2)**.
- vi. For **Replication**, select **Locally-redundant storage (LRS)**.

Dashboard > Resource groups > rg-oracle > New >

### Create storage account

[Basics](#) [Networking](#) [Data protection](#) [Advanced](#) [Tags](#) [Review + create](#)

Azure Storage is a Microsoft-managed service providing cloud storage that is highly available, secure, durable, scalable, and redundant. Azure Storage includes Azure Blobs (objects), Azure Data Lake Storage Gen2, Azure Files, Azure Queues, and Azure Tables. The cost of your storage account depends on the usage and the options you choose below. [Learn more about Azure storage accounts](#)

**Project details**  
Select the subscription to manage deployed resources and costs. Use resource groups like folders to organize and manage all your resources.

**Subscription \***

**Resource group \***  [Create new](#)

**Instance details**  
The default deployment model is Resource Manager, which supports the latest Azure features. You may choose to deploy using the classic deployment model instead. [Choose classic deployment model](#)

<b>Storage account name *</b> <input type="text" value="orarestore"/>	<input type="button" value=""/>
<b>Location *</b> <input type="button" value="(US) East US"/>	<input type="radio"/> Standard <input type="radio"/> Premium
<b>Performance</b> <input type="radio"/> <input checked="" type="radio"/> Premium	<input type="button" value=""/>
<b>Account kind</b> <input type="radio"/> <input checked="" type="radio"/> StorageV2 (general purpose v2)	<input type="button" value=""/>
<b>Replication</b> <input type="radio"/> <input checked="" type="radio"/> Locally-redundant storage (LRS)	<input type="button" value=""/>

[Review + create](#) [< Previous](#) [Next : Networking >](#)

c. Select **Review + Create**, and then select **Create**.

2. In the Azure portal, search for the **myVault** Recovery Services vault and select it.

myVault

Services

- Key vaults
- Backup vaults
- Recovery Services vaults

Resources

- myVault Recovery Services vault

Marketplace

No results were found.

Documentation

- Overview of Backup vaults - Azure Backup | Microsoft Docs
- Quickstart to create an Azure Recovery Services vault ...
- Use Key Vault references - Azure App Service | Microsoft Docs
- Azure Key Vault VM Extension for Linux - Azure Virtual ...

Resource Groups

No results were found.

Didn't find what you were looking for?

- Try searching in Activity Log
- Try searching in Azure Active Directory

Searching 2 of 16 subscriptions.

CustomEntityLookup20191114114315

Function App

3. On the **Overview** pane, select **Backup items**. Then select **Azure Virtual Machine**, which should have a nonzero number for **BACKUP ITEM COUNT**.

Home > myVault

## myVault | Backup items

Recovery Services vault

Search (Ctrl+ /) Refresh

BACKUP MANAGEMENT TYPE	BACKUP ITEM COUNT
Azure Virtual Machine	1
SAP HANA in Azure VM	0
SQL in Azure VM	0
Azure Storage (Azure Files)	0
DPM	0
Azure Backup Server	0
Azure Backup Agent	0

Overview

Activity log

Access control (IAM)

Tags

Diagnose and solve problems

Settings

Identity

Private endpoint connections

Properties

Locks

Getting started

Backup

Site Recovery

Protected items

Backup items

Replicated items

Manage

Backup policies

4. On the **Backups Items (Azure Virtual Machine)** pane, select the **vmoracle19c** VM.

Home > myVault >

## Backup Items (Azure Virtual Machine)

myVault

Fetching data from service completed.

Name	Resource Group	Backup Pre-Check	Last Backup Status	Latest restore point	...
vmoracle19c	rg-oracle	Passed	Success	12/15/2020, 11:06:29 PM	...

5. On the **vmoracle19c** pane, choose a restore point that has a consistency type of **Application Consistent**. Select the ellipsis (...), and then select **Restore VM**.

Home > myVault > Backup Items (Azure Virtual Machine) >

## vmoracle19c

Backup Item

[Backup now](#) [Restore VM](#) [File Recovery](#) [Stop backup](#) [Resume backup](#) [Delete backup data](#) [Restore to Secondary Region](#) [Undelete](#)

Alerts and Jobs	Backup status	Summary
<a href="#">View all Alerts</a> (last 24 hours)	Backup Pre-Check <span style="color: green;">Passed</span>	Recovery services vault <b>myVault</b>
<a href="#">View all Jobs</a> (last 24 hours)	Last backup status <span style="color: green;">Success 12/15/2020, 11:06:26 PM</span>	Backup policy <b>DailyPolicy</b>
		Oldest restore point <b>12/15/2020, 10:16:18 PM (1 hour(s) ago)</b>

**Restore points (2)**

This list is filtered for last 30 days of restore points. To recover from restore point older than 30 days, [click here](#).

Time	Consistency	Recovery Type	...
12/15/2020, 11:06:29 PM	Application Consistent	Snapshot	<a href="#">Restore VM</a>
12/15/2020, 10:16:18 PM	File-system Consistent	Snapshot	<a href="#">File Recovery</a>

6. On the **Restore Virtual Machine** pane:

- Select **Create New**.
- For **Restore Type**, select **Create new virtual machine**.
- For **Virtual machine name**, enter **vmoracle19c**.
- For **Virtual network**, select **vmoracle19cVNET**.

The subnet is automatically populated based on your selection for the virtual network.

- For **Staging Location**, the process of restoring a VM requires an Azure storage account in the same resource group and region. You can choose a storage account or a restore task that you set up earlier.

## Restore Virtual Machine

vmoracle19c

Restore allows you to restore VM/disks from a selected Restore Point.

Restore point \*

12/15/2020, 11:06:29 PM

Select

### Restore Configuration

Create new  Replace existing

**i** To create an alternate configuration when restoring your VM (from the following menus), use PowerShell cmdlets.

Restore Type \* ⓘ

Create new virtual machine

Virtual machine name \* ⓘ

vmoracle19c

Resource group \* ⓘ

rg-oracle2

Virtual network \* ⓘ

vmoracle19cVNET (rg-oracle)

Subnet \* ⓘ

vmoracle19cSubnet

Staging Location \* ⓘ

orabackup1 (StandardLRS)

[Can't find your storage account ?](#)**Restore**

7. To restore the VM, select the **Restore** button.

8. To view the status of the restore process, select **Jobs**, and then select **Backup Jobs**.

Workload name	Operation	Status	Type	Start time	Duration
vmoracle19c	Restore	In progress	Azure virtual machine	12/16/2020, 12:16:45 AM	00:00:15
vmoracle19c	Backup	Completed	Azure virtual machine	12/15/2020, 11:06:26 PM	00:17:37
vmoracle19c	Backup	Cancelled	Azure virtual machine	12/15/2020, 10:44:45 PM	00:06:55
vmoracle19c	Backup	In progress	Azure virtual machine	12/15/2020, 10:16:15 PM	02:00:46
vmoracle19c	Configure backup	Completed	Azure virtual machine	12/15/2020, 10:15:06 PM	00:00:30
vmoracle19c	Delete backup data	Completed	Azure virtual machine	12/15/2020, 10:05:56 PM	00:01:51
vmoracle19c	Delete backup data	Completed	Azure virtual machine	12/15/2020, 10:02:37 PM	00:01:51
vmoracle19c	Disable backup	Completed	Azure virtual machine	12/15/2020, 10:00:41 PM	00:00:10
vmoracle19c	Configure backup	Completed	Azure virtual machine	12/15/2020, 9:40:55 PM	00:00:40

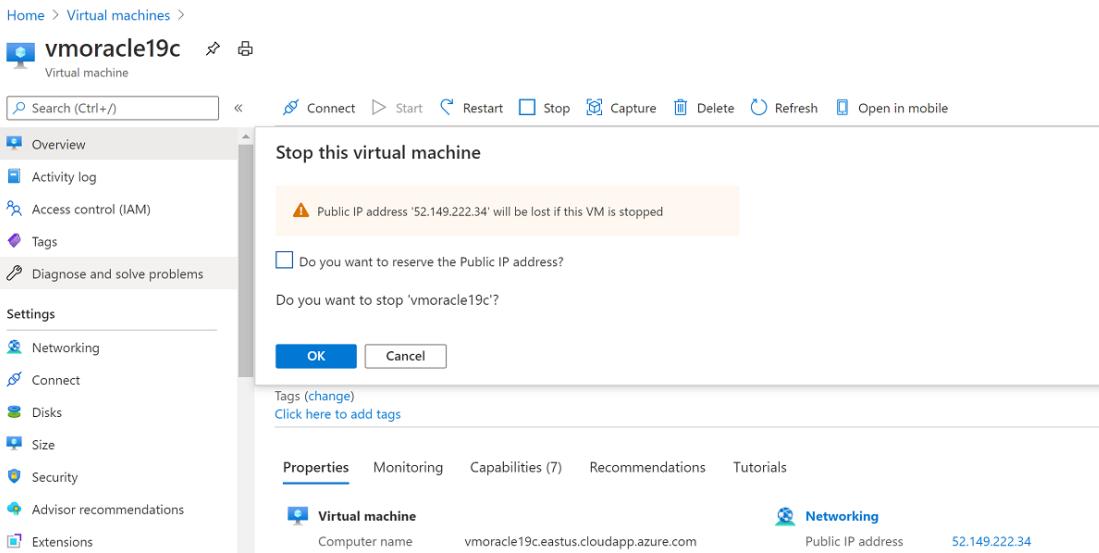
Select the **In Progress** restore operation to show details about the status of the restore process.

## Set the public IP address

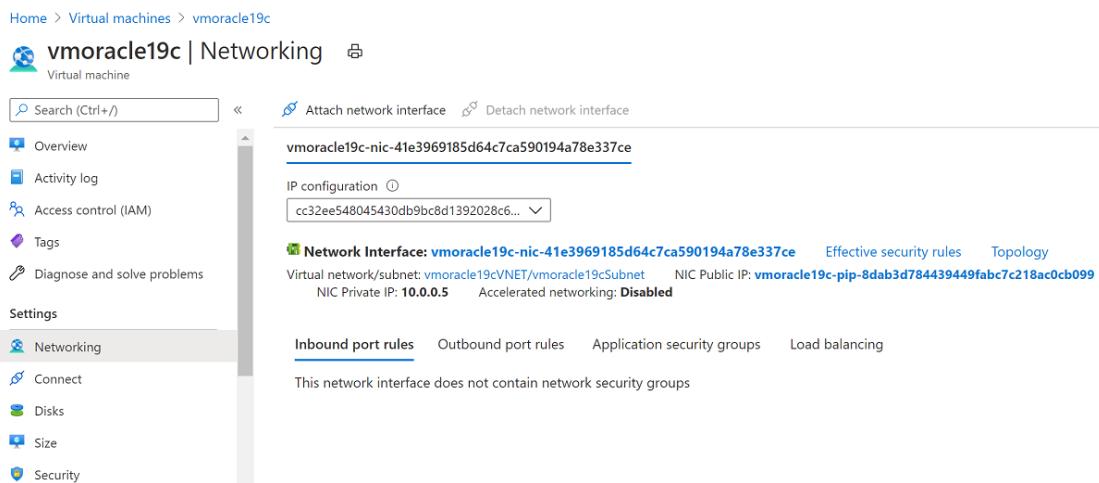
After the VM is restored, you should reassign the original IP address to the new VM.

Virtual machine		Networking	
Computer name	vmoracle19c.eastus.cloudapp.azure.com	Public IP address	52.149.222.34
Operating system	Linux (oracle 7.7)	Public IP address (IPv6)	-
Publisher	N/A	Private IP address	10.0.0.5
Offer	N/A	Private IP address (IPv6)	-
Plan	N/A	Virtual network/subnet	vmoracle19cVNET/vmoracle19cSubnet
VM generation	V1	DNS name	Configure
Agent status	Ready		

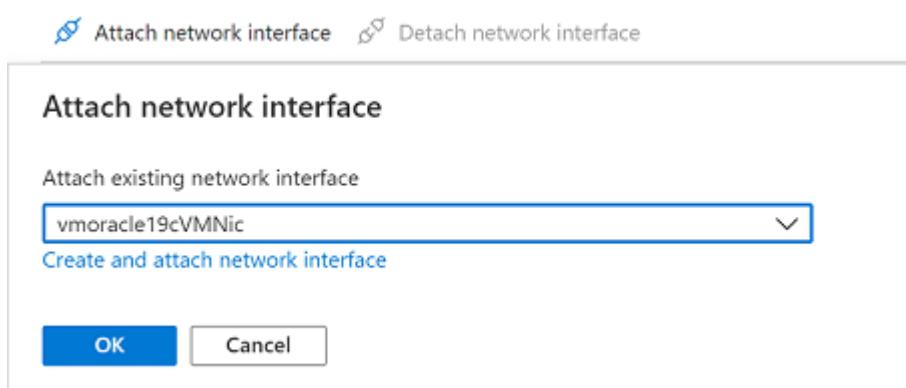
1. In the Azure portal, go to your virtual machine named **vmoracle19c**. It has been assigned a new public IP and NIC similar to **vmoracle19c-nic-XXXXXXXXXXXX**, but it doesn't have a DNS address. When the original VM was deleted, its public IP and NIC were retained. The next steps reattach them to the new VM.
2. Stop the VM.



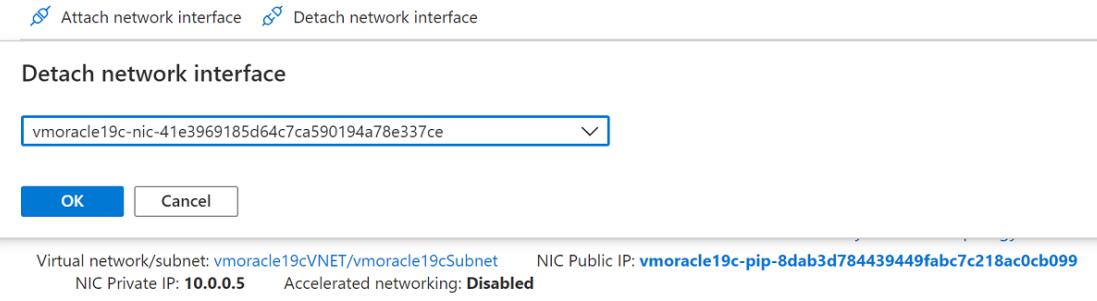
### 3. Go to Networking.



4. Select **Attach network interface**. Select the original NIC **vmoracle19cVMNic**, which the original public IP address is still associated with. Then select **OK**.



5. Detach the NIC that you created with the VM restore operation, because it's configured as the primary interface. Select **Detach network interface**, select the NIC that's similar to **vmoracle19c-nic-XXXXXXXXXXXX**, and then select **OK**.



Your re-created VM now has the original NIC, which is associated with the original IP address and network security group rules.

Priority	Name	Port	Protocol	Source	Destination	Action
1000	default-allow-ssh	22	TCP	Any	Any	<span style="color: green;">A</span>
1001	allow-oracle	1521	TCP	Any	Any	<span style="color: green;">A</span>
1002	allow-oracle-EM	5502	TCP	Any	Any	<span style="color: green;">A</span>
65000	AllowVnetInBound	Any	Any	VirtualNetwork	VirtualNetwork	<span style="color: green;">A</span>
65001	AllowAzureLoadBalancerInBound	Any	Any	AzureLoadBalancer	Any	<span style="color: green;">A</span>
65500	DenyAllInBound	Any	Any	Any	Any	<span style="color: red;">D</span>

6. Go back to the Overview pane and select Start.

## Recover the database

To recover a database after a complete VM restore:

1. Reconnect to the VM:

Bash

```
ssh azureuser@<publicIpAddress>
```

When the whole VM has been restored, it's important to recover the databases on the VM by performing the following steps on each database.

2. You might find that the instance is running, because the autostart attempted to start the database on VM startup. However, the database requires recovery and is

likely to be at the mount stage only. Run a preparatory shutdown before starting the mount stage:

```
Bash

sudo su - oracle
sqlplus / as sysdba
SQL> shutdown immediate
SQL> startup mount
```

### 3. Perform database recovery.

It's important to specify the `USING BACKUP CONTROLFILE` syntax to inform the `RECOVER AUTOMATIC DATABASE` command that recovery should not stop at the Oracle system change number (SCN) recorded in the restored database control file.

The restored database control file was a snapshot, along with the rest of the database. The SCN stored within it is from the point in time of the snapshot. There might be transactions recorded after this point, and you want to recover to the point of the last transaction committed to the database.

```
Bash

SQL> recover automatic database using backup controlfile until cancel;
```

### 4. When the last available archive log file has been applied, enter `CANCEL` to end recovery.

When recovery finishes successfully, the message `Media recovery complete` appears.

However, when you're using the `BACKUP CONTROLFILE` clause, the recover command ignores online log files. It's possible that changes in the current online redo log are required to complete point-in-time recovery. In this situation, you might see messages similar to these examples:

```
Output

SQL> recover automatic database until cancel using backup controlfile;
ORA-00279: change 2172930 generated at 04/08/2021 12:27:06 needed for
thread 1
ORA-00289: suggestion :
/u02/fast_recovery_area/ORATEST1/archivelog/2021_04_08/o1_mf_1_13_%u_.a
rc
ORA-00280: change 2172930 for thread 1 is in sequence #13
ORA-00278: log file
```

```
'/u02/fast_recovery_area/ORATEST1/archivelog/2021_04_08/o1_mf_1_13_%u_.  
arc' no  
longer needed for this recovery  
ORA-00308: cannot open archived log  
'/u02/fast_recovery_area/ORATEST1/archivelog/2021_04_08/o1_mf_1_13_%u_.  
arc'  
ORA-27037: unable to obtain file status  
Linux-x86_64 Error: 2: No such file or directory  
Additional information: 7  
  
Specify log: {<RET>=suggested | filename | AUTO | CANCEL}
```

### *i* Important

If the current online redo log is lost or corrupted and you can't use it, you might cancel recovery at this point.

To correct this situation, you can identify which online log wasn't archived and supply the fully qualified file name to the prompt.

## 5. Open the database.

The `RESETLOGS` option is required when the `RECOVER` command uses the `USING BACKUP CONTROLFILE` option. `RESETLOGS` creates a new incarnation of the database by resetting the redo history back to the beginning, because there's no way to determine how much of the previous database incarnation was skipped in the recovery.

Bash

```
SQL> alter database open resetlogs;
```

## 6. Check that the database content was recovered:

Bash

```
SQL> select * from scott.scott_table;
```

The backup and recovery of Oracle Database on an Azure Linux VM are now finished.

You can find more information about Oracle commands and concepts in the Oracle documentation, including:

- [Performing Oracle user-managed backups of the entire database ↗](#)
- [Performing complete user-managed database recovery ↗](#)

- Oracle STARTUP command ↗
- Oracle RECOVER command ↗
- Oracle ALTER DATABASE command ↗
- Oracle LOG\_ARCHIVE\_DEST\_n parameter ↗
- Oracle ARCHIVE\_LAG\_TARGET parameter ↗

## Delete the VM

When you no longer need the VM, you can use the following commands to remove the resource group, the VM, and all related resources:

1. Disable soft delete of backups in the vault:

Azure CLI

```
az backup vault backup-properties set --name myVault --resource-group rg-oracle --soft-delete-feature-state disable
```

2. Stop protection for the VM and delete backups:

Azure CLI

```
az backup protection disable --resource-group rg-oracle --vault-name myVault --container-name vmoracle19c --item-name vmoracle19c --delete-backup-data true --yes
```

3. Remove the resource group, including all resources:

Azure CLI

```
az group delete --name rg-oracle
```

## Next steps

[Create highly available VMs](#)

[Explore Azure CLI samples for VM deployment ↗](#)

# Disaster recovery for an Oracle Database 12c database in an Azure environment

Article • 08/23/2021

Applies to:  Linux VMs

## Assumptions

- You have an understanding of Oracle Data Guard design and Azure environments.

## Goals

- Design the topology and configuration that meet your disaster recovery (DR) requirements.

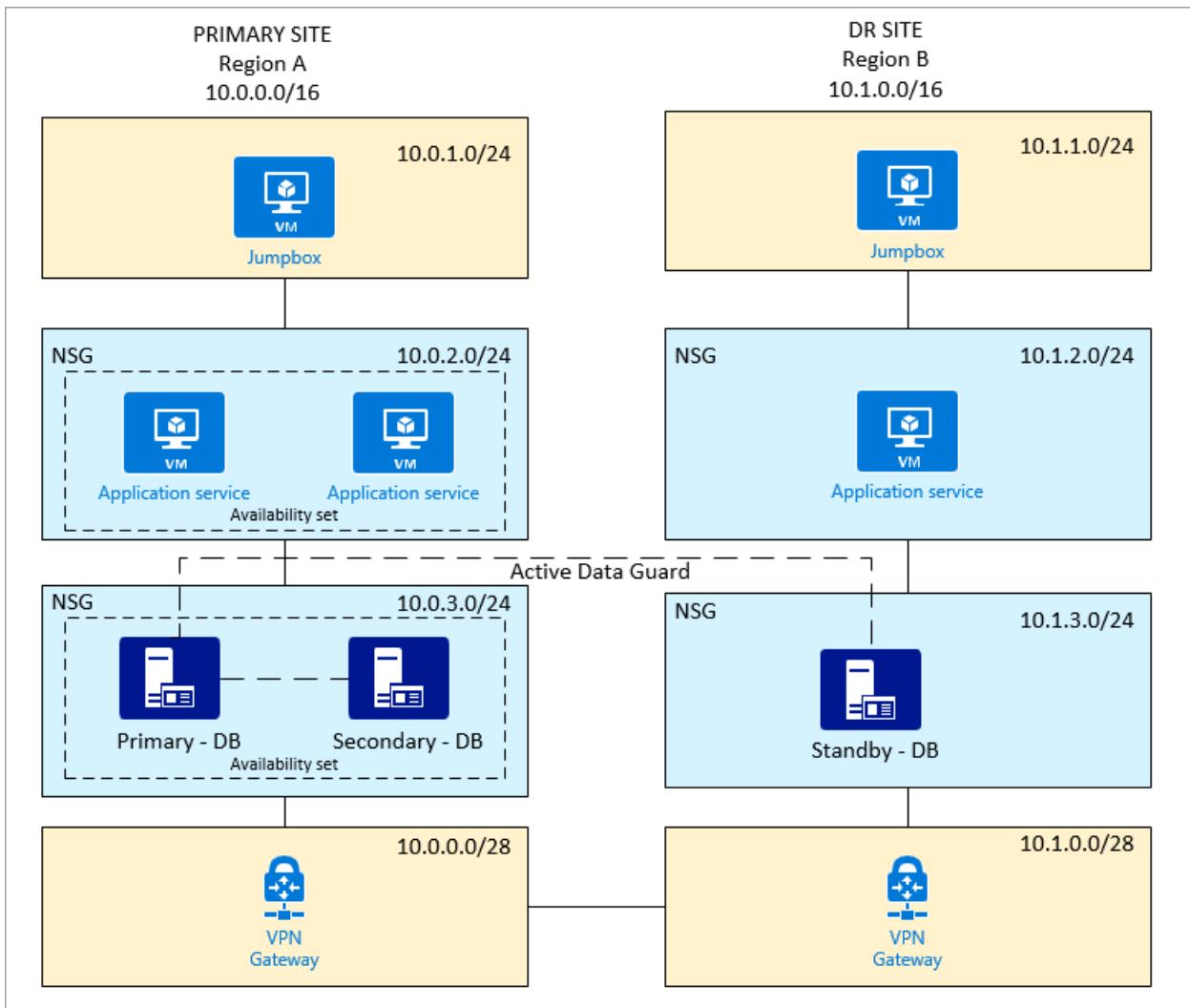
## Scenario 1: Primary and DR sites on Azure

A customer has an Oracle database set up on the primary site. A DR site is in a different region. The customer uses Oracle Data Guard for quick recovery between these sites. The primary site also has a secondary database for reporting and other uses.

## Topology

Here is a summary of the Azure setup:

- Two sites (a primary site and a DR site)
- Two virtual networks
- Two Oracle databases with Data Guard (primary and standby)
- Two Oracle databases with Golden Gate or Data Guard (primary site only)
- Two application services, one primary and one on the DR site
- An *availability set*, which is used for database and application service on the primary site
- One jumpbox on each site, which restricts access to the private network and only allows sign-in by an administrator
- A jumpbox, application service, database, and VPN gateway on separate subnets
- NSG enforced on application and database subnets



## Scenario 2: Primary site on-premises and DR site on Azure

A customer has an on-premises Oracle database setup (primary site). A DR site is on Azure. Oracle Data Guard is used for quick recovery between these sites. The primary site also has a secondary database for reporting and other uses.

There are two approaches for this setup.

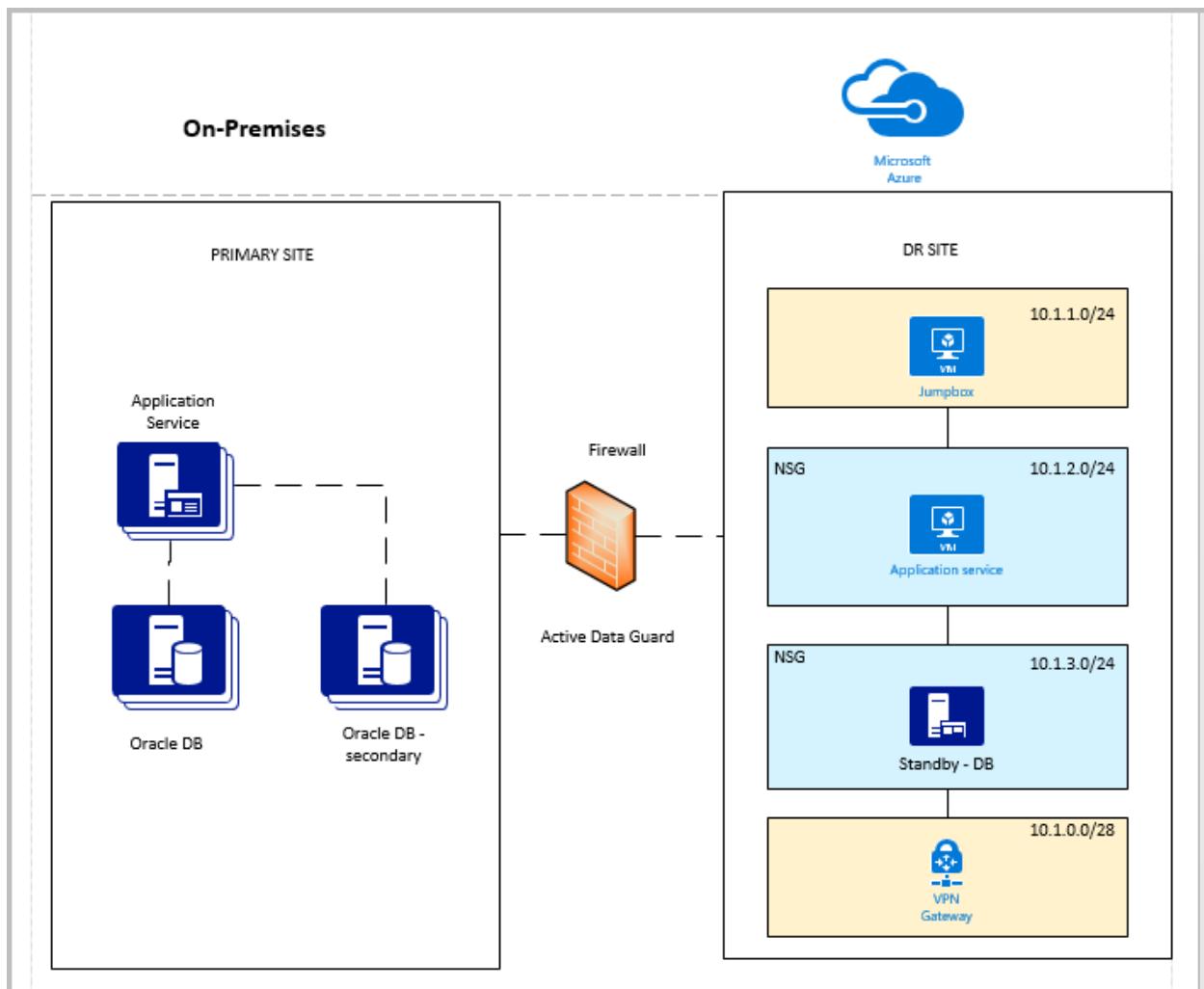
### Approach 1: Direct connections between on-premises and Azure, requiring open TCP ports on the firewall

We don't recommend direct connections because they expose the TCP ports to the outside world.

### Topology

Following is a summary of the Azure setup:

- One DR site
- One virtual network
- One Oracle database with Data Guard (active)
- One application service on the DR site
- One jumpbox, which restricts access to the private network and only allows sign-in by an administrator
- A jumpbox, application service, database, and VPN gateway on separate subnets
- NSG enforced on application and database subnets
- An NSG policy/rule to allow inbound TCP port 1521 (or a user-defined port)
- An NSG policy/rule to restrict only the IP address/addresses on-premises (DB or application) to access the virtual network



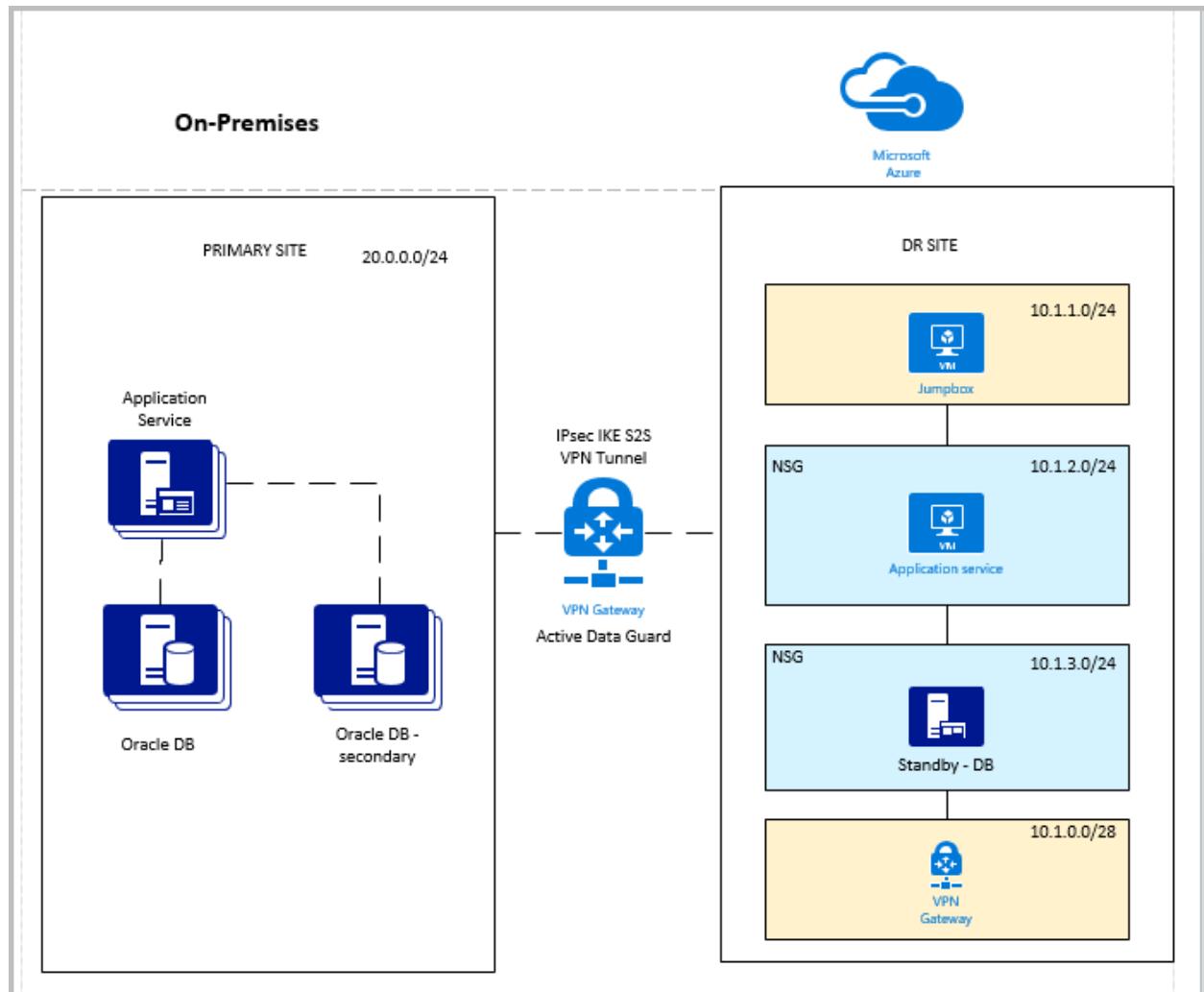
## Approach 2: Site-to-site VPN

Site-to-site VPN is a better approach. For more information about setting up a VPN, see [Create a virtual network with a Site-to-Site VPN connection using CLI](#).

## Topology

Following is a summary of the Azure setup:

- One DR site
- One virtual network
- One Oracle database with Data Guard (active)
- One application service on the DR site
- One jumpbox, which restricts access to the private network and only allows sign-in by an administrator
- A jumpbox, application service, database, and VPN gateway are on separate subnets
- NSG enforced on application and database subnets
- Site-to-site VPN connection between on-premises and Azure



## Additional reading

- Design and implement an Oracle database on Azure
- Configure Oracle Data Guard
- Configure Oracle Golden Gate
- Oracle backup and recovery

# Next steps

- [Tutorial: Create highly available VMs](#)
- [Explore VM deployment Azure CLI samples](#) ↗

# What are solutions for running Oracle WebLogic Server on Azure Virtual Machines?

Article • 10/25/2023

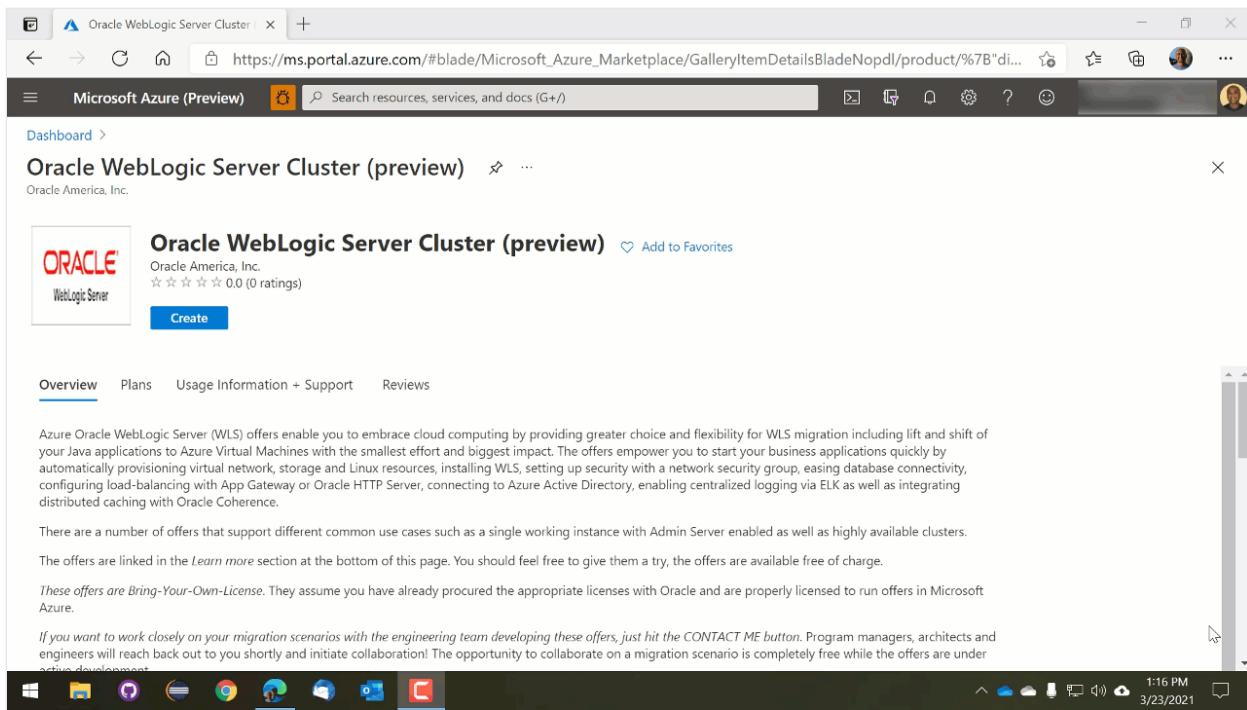
Applies to:  Linux VMs

This page describes the solutions for running Oracle WebLogic Server (WLS) on Azure virtual machines. These solutions are jointly developed and supported by Oracle and Microsoft.

You can also run WLS on the Azure Kubernetes Service. The solutions to do so are described in [this Microsoft article](#).

WLS is a leading Java application server running some of the most mission-critical enterprise Java applications across the globe. WLS forms the middleware foundation for the Oracle software suite. Oracle and Microsoft are committed to empowering WLS customers with choice and flexibility to run workloads on Azure as a leading cloud platform.

The Azure WLS solutions are aimed at making it as easy as possible to migrate your Java applications to Azure virtual machines. The solutions do so by generating deployed resources for most common cloud provisioning scenarios. The solutions automatically provision virtual network, storage, Java, WLS, and Linux resources. With minimal effort, WebLogic Server is provisioned. The solutions can set up security with a network security group, load balancing with Azure App Gateway or Oracle HTTP Server, and distributed caching with Oracle Coherence. You can also automatically connect to your existing database including Azure PostgreSQL, Azure MySQL, Azure SQL, and the Oracle Database on the Oracle Cloud or Azure.



There are solution templates available to meet different scenarios such as [single instance with an admin server](#), and [cluster](#). The solutions are available free of charge. These solutions are described and linked below. You can find detailed documentation on the solutions [here](#).

*These offers are Bring-Your-Own-License.* They assume you have already got the appropriate licenses with Oracle and are properly licensed to run offers in Azure.

The solution templates support a range of operating system, Java, and WLS versions through base images (such as WebLogic Server 14 and Java 11 on Red Hat Enterprise Linux 8). These base images are also available on Azure Marketplace on their own. The base images are suitable for customers that require complex, customized Azure deployments.

If you prefer step-by-step guidance for going from zero to a WLS cluster without any solution templates or base images, see [Install Oracle WebLogic Server on Azure Virtual Machines manually](#).

*If you're interested in working closely on your migration scenarios with the engineering team developing these offers, select the [CONTACT ME](#) button on the [marketplace offer overview page](#).* Program managers, architects, and engineers will reach back out to you shortly and start close collaboration.

## Oracle WebLogic Server with Admin Server

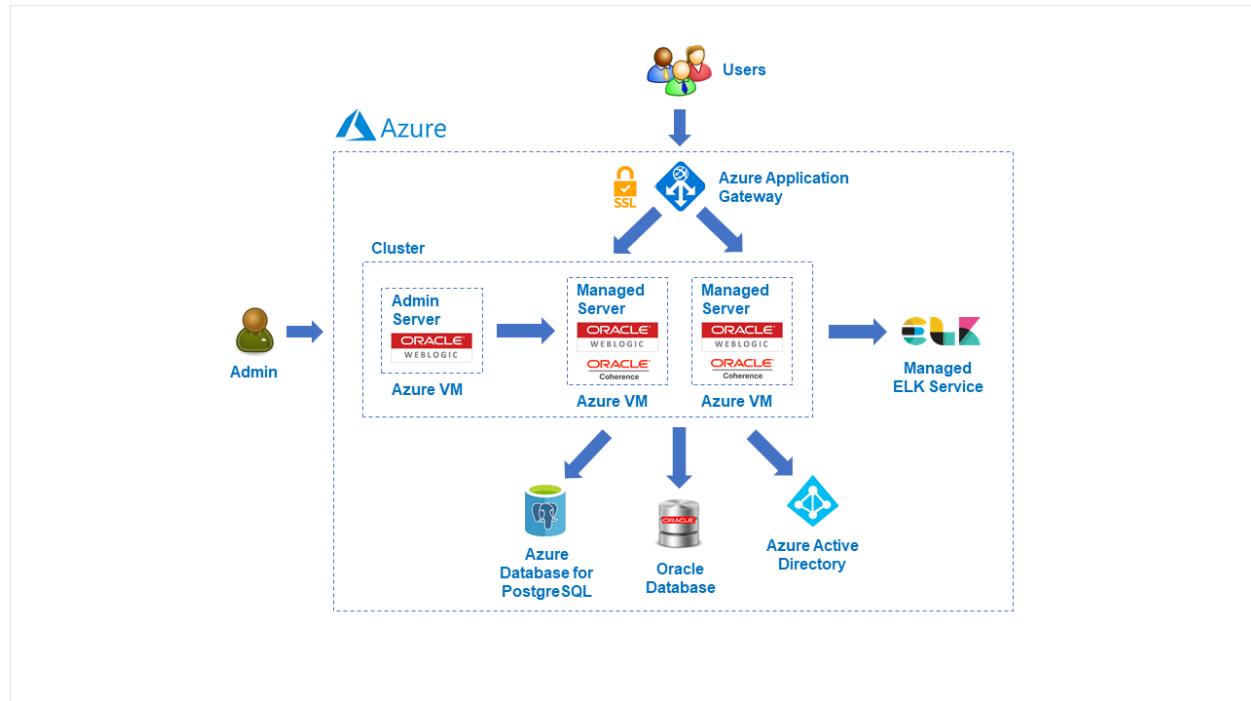
This [solution template](#) provisions a single virtual machine and installs WLS on it. It creates a domain and starts up the administration server. You can manage the domain

and get started with application deployments right away.

## Oracle WebLogic Server Cluster

This solution template [creates a highly available cluster of WLS virtual machines](#). The administration server and all managed servers are started by default. You can manage the cluster and get started with highly available applications right away.

The solutions enable a wide range of production-ready deployment architectures with relative ease. You can meet most migration cases in the most productive way possible by allowing a focus on business application development.



After resources are automatically provisioned by the solutions, you have complete flexibility to customize your deployments further. It's likely on top of deploying applications you'll integrate further Azure resources with your deployments. You're encouraged to [connect with the development team](#) and provide feedback on further improving the solutions.

## Next steps

Explore the offers on Azure.

[Oracle WebLogic Server with Admin Server](#)

[Oracle WebLogic Server Cluster](#)

# What are solutions for running Oracle WebLogic Server on the Azure Kubernetes Service?

Article • 10/25/2023

Applies to:  Linux VMs

This page describes the solutions for running Oracle WebLogic Server (WLS) on the Azure Kubernetes Service (AKS). These solutions are jointly developed and supported by Oracle and Microsoft.

It's also possible to run WebLogic Server on Azure Virtual Machines. The solutions to do so are described in [this Microsoft article](#).

WebLogic Server is a leading Java application server running some of the most mission-critical enterprise Java applications across the globe. WebLogic Server forms the middleware foundation for the Oracle software suite. Oracle and Microsoft are committed to empowering WebLogic Server customers with choice and flexibility to run workloads on Azure as a leading cloud platform.

## WLS on AKS certified and supported

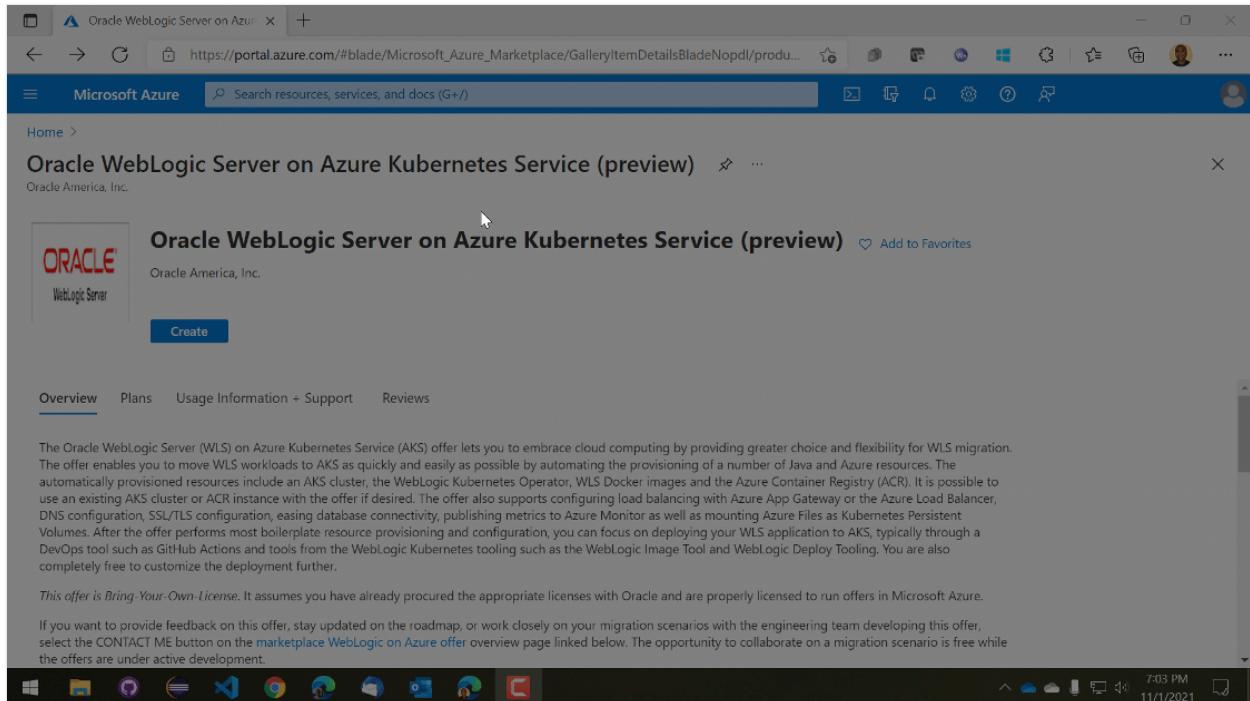
WebLogic Server is certified by Oracle and Microsoft to run well on AKS. The WLS on AKS solutions are aimed at making it as easy as possible to run your containerized and orchestrated Java applications on Kubernetes. The solutions are focused on reliability, scalability, manageability, and enterprise support.

WLS clusters are fully enabled to run on Kubernetes via the WebLogic Kubernetes Operator (referred to simply as the 'Operator' from here onward). The Operator follows the standard Kubernetes Operator pattern. It simplifies the management and operation of WebLogic domains on Kubernetes by automating otherwise manual tasks and adding extra operational reliability features. The Operator supports Oracle WebLogic Server 12c, Oracle Fusion Middleware Infrastructure 12c and beyond. For details on the Operator, refer to the [official documentation from Oracle](#).

## WLS on AKS marketplace solution template

Beyond certifying WLS on AKS, Oracle and Microsoft jointly provide a [marketplace solution template](#) with the goal of making it as quick and easy as possible to migrate

WLS workloads to AKS. The offer does so by automating the provisioning of a number of Java and Azure resources. The automatically provisioned resources include an AKS cluster, the WebLogic Kubernetes Operator, WLS Docker images, and the Azure Container Registry (ACR). It's possible to use an existing AKS cluster or ACR instance with the offer. The offer also supports configuring load balancing with Azure App Gateway or the Azure Load Balancer, easing database connectivity, publishing metrics to Azure Monitor and mounting Azure Files as Kubernetes Persistent Volumes. The currently supported database integrations include Azure PostgreSQL, Azure MySQL, Azure SQL, and the Oracle Database on the Oracle Cloud or Azure.



After the solution template performs most boilerplate resource provisioning and configuration, you can focus on deploying your WLS application to AKS, typically through a DevOps tool such as GitHub Actions and tools from WebLogic Kubernetes tooling such as the WebLogic Image Tool and WebLogic Deploy Tooling. You're completely free to customize the deployment further.

You can find detailed documentation on the solution template [here](#).

## Guidance, scripts, and samples for WLS on AKS

Oracle and Microsoft also provide basic step-by-step guidance, scripts, and samples for running WebLogic Server on AKS. The guidance is suitable for customers that wish to remain as close as possible to a native Kubernetes manual deployment experience as an alternative to using a solution template. The guidance is incorporated into the Azure Kubernetes Service sample section of the [Operator documentation](#). The guidance allows a very high degree of configuration and customization.

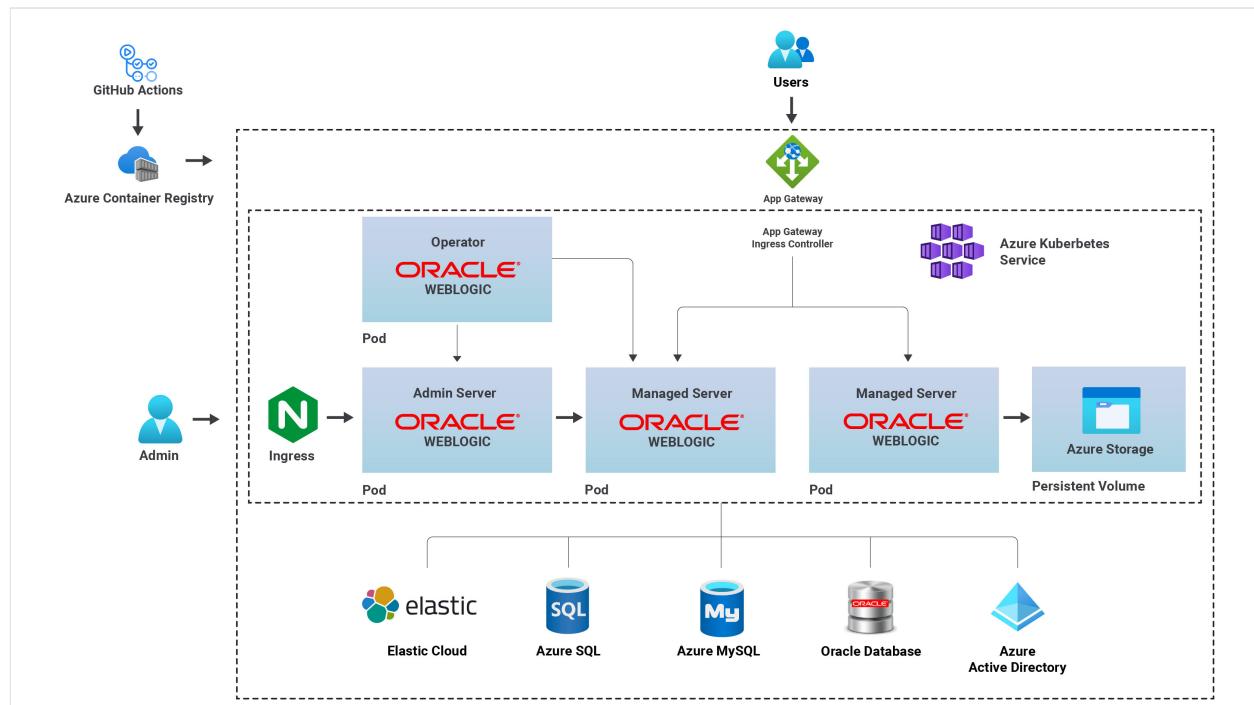
The guidance supports two ways of deploying WLS domains to AKS. Domains can be deployed directly to Kubernetes Persistent Volumes. This deployment option is good if you want to migrate to AKS but still want to administer WLS using the Admin Console or the WebLogic Scripting Tool (WLST). The option also allows you to move to AKS without adopting Docker development. The more Kubernetes native way of deploying WLS domains to AKS is to build custom container images based on official WLS images from the Oracle Container Registry, publish the custom images to ACR and deploy the domain to AKS using the Operator.

*These solutions are all Bring-Your-Own-License.* They assume you've already got the appropriate licenses with Oracle and are properly licensed to run offers in Azure.

*If you're interested in working closely on your migration scenarios with the engineering team developing these solutions, fill out [this short survey](#) and include your contact information.* Program managers, architects, and engineers will reach back out to you shortly and start close collaboration.

## Deployment architectures

The solutions for running Oracle WebLogic Server on the Azure Kubernetes Service enable a wide range of production-ready deployment architectures with relative ease.



Beyond what the solutions provide you have complete flexibility to customize your deployments further. It's likely on top of deploying applications you'll integrate further Azure resources with your deployments or tune the deployments to your specific applications. You're encouraged to provide feedback in the [survey](#) on further improving the solutions.

# Next steps

Explore running Oracle WebLogic Server on the Azure Kubernetes Service.

[WLS on AKS marketplace solution](#)

[WLS on AKS marketplace solution documentation](#)

[Guidance, scripts and samples for running WLS on AKS](#)

[WebLogic Kubernetes Operator](#)

# Migrate WebLogic Server applications to Azure Virtual Machines

Article • 04/03/2023

This guide describes what you should be aware of when you want to migrate an existing WebLogic application to run on Azure Virtual Machines. For an overview of available WebLogic Server solutions in Azure Marketplace, see [What are solutions for running Oracle WebLogic Server on Azure Virtual Machines?](#)

## Pre-migration

To ensure a successful migration, before you start, complete the assessment and inventory steps described in the following sections.

### Define what you mean by "migration complete"

This guide, and the corresponding Azure Marketplace Offers, are a starting point to accelerate the migration of your WebLogic Server workloads to Azure. It's important to define the scope of your migration effort. For example, are you doing a strict "lift and shift" from your existing infrastructure to Azure Virtual Machines? If so, you may be tempted to work in some "lift and improve" as you migrate.

It's better to stick as close to pure "lift and shift" as possible, accounting for the necessary changes as detailed in this guide. Define what you mean by "migration complete" so that you know when you've reached this milestone. When you've reached your "migration complete", you can take a snapshot of your Virtual Machines as described in [Create a snapshot](#). After you've verified that you can successfully restore from your snapshot, you can do the improvements without fear of losing the migration progress you've achieved thus far.

### Ensure that the target is the appropriate target for your migration effort

The first step in a successful migration of a WLS application to Azure is selecting the most appropriate migration target. WLS runs well on Azure virtual machines (VMs) or Azure Kubernetes Service (AKS). The VM target is the easiest choice, because it most closely resembles an on-premises deployment. The administrative and deployment experience for virtual machines is very analogous to what you have on-premises. The

trade-off for this ease is economic cost. Generally speaking, the per-minute cost for a VM-based solution is higher compared with AKS. While an AKS-based solution costs less to run, you must constrain your application to fit within the requirements of AKS. If minimizing change is the most important factor for your migration effort, consider a VM-based migration. In this case, see [Migrate WebLogic applications to Azure Virtual Machines](#). If you can tolerate converting your application to run within Kubernetes to reduce runtime cost, consider an AKS-based migration. In this case, continue with [Migrate WebLogic Server applications to Azure Kubernetes Service](#).

## Determine whether the prebuilt Azure Marketplace offers are a good starting point

Oracle and Microsoft have partnered to bring a set of Azure solution templates to Azure Marketplace to provide a solid starting point for migrating to Azure. Consult the [Oracle Fusion Middleware](#) documentation for the list of offers and choose the one that most closely matches your existing deployment. You can see the list of offers in the overview article [What is Oracle WebLogic Server on Azure?](#)

If none of the existing offers are a good starting point, you have to reproduce the deployment by hand using Azure Virtual Machine resources. You can find the step-by-step guidance in [Install Oracle WebLogic Server on Azure Virtual Machines manually](#). For more information, see [What is IaaS?](#)

## Determine whether the WebLogic version is compatible

Your existing WebLogic version must be compatible with the version in the IaaS offers. To see the offers for WebLogic version 12.2.1.3, [query Azure Marketplace for Oracle WebLogic 12.2.1.3](#). If your existing WebLogic version isn't compatible with that version, you have to reproduce the deployment by hand using Azure IaaS resources. For more information, see [the Azure documentation](#).

## Inventory server capacity

Document the hardware (memory, CPU, disk) of the current production server(s) as well as the average and peak request counts and resource utilization. This information must inform the choice of VM size. For more information, see [Sizes for Cloud Services](#).

## Inventory all secrets

Before the advent of "configuration as a service" technologies such as Azure Key Vault, there wasn't a well-defined concept of "secrets". Instead, you had a disparate set of configuration settings that effectively functioned as what we now call "secrets". With app servers such as WebLogic Server, these secrets are in many different config files and configuration stores. Check all properties and configuration files on the production server(s) for any secrets and passwords. Be sure to check `weblogic.xml` in your WARs. Configuration files containing passwords or credentials may also be found inside your application. For more information, see [Azure Key Vault basic concepts](#).

## Inventory all certificates

Document all the certificates used for public SSL endpoints. You can view all certificates on the production server(s) by running the following command:

Bash

```
keytool -list -v -keystore <path to keystore>
```

## Validate that the supported Java version works correctly

All of the migration paths for WebLogic to Azure require a specific Java version, which varies for each path. You'll need to validate that your application is able to run correctly using that supported version.

### ⓘ Note

This validation is especially important if your current server is running on an unsupported JDK (such as Oracle JDK or IBM OpenJ9).

To obtain your current Java version, sign in to your production server and run the following command:

Bash

```
java -version
```

### ⓘ Note

When migrating to WLS on Azure virtual machines, the requirements for the specific Java versions are determined by the pre-installed Java on the virtual

machines. When migrating to WLS on AKS, the specific Java version is determined by the container image chosen. There are a wide variety of choices, but all of them use the Oracle JDK.

## Inventory JNDI resources

Inventory all JNDI resources. For example, datasources such as databases may have an associated JNDI name that allows JPA to correctly bind instances of `EntityManager` to a particular database. For more information on JNDI resources and databases, see [WebLogic Server Data Sources](#) in the Oracle documentation. Other JNDI-related resources, such as JMS message brokers, may require migration or reconfiguration. For more information on JMS configuration see [Oracle WebLogic Server 12.2.1.4.0](#).

## Inspect your domain configuration

The main configuration unit in WebLogic Server is the domain. As such, the `config.xml` file contains a wealth of configuration that you must carefully consider for migration. The file includes references to additional XML files that are stored in subdirectories. Oracle advises that you should normally use the **Administration Console** to configure WebLogic Server's manageable objects and services and allow WebLogic Server to maintain the `config.xml` file. For more information, see [Domain Configuration Files](#).

## Inside your application

Inspect the `WEB-INF/weblogic.xml` file and/or the `WEB-INF/web.xml` file.

## Determine whether session replication is used

If your application relies on session replication, with or without Oracle Coherence\*Web, you have three options:

- Coherence\*Web can run alongside a WebLogic Server in the Azure virtual machines, but you must manually configure this option after you provision the offer. If you are using standalone Coherence, you can also run it in an Azure virtual machine, but you must manually configure this option after you provision the offer.
- Refactor your application to use a database for session management.
- Refactor your application to externalize the session to Azure Redis Service. For more information, see [Azure Cache for Redis](#).

For all of these options, it's a good idea to master how WebLogic does HTTP Session State Replication. For more information, see [HTTP Session State Replication](#) in the Oracle documentation.

## Document datasources

If your application uses any databases, you need to capture the following information:

- What is the datasource name?
- What is the connection pool configuration?
- Where can I find the JDBC driver JAR file?

For more information on JDBC drivers in WebLogic, see [Using JDBC Drivers with WebLogic Server](#).

## Determine whether WebLogic has been customized

Determine which of the following customizations have been made, and capture what's been done.

- Have the startup scripts been changed? Such scripts include *setDomainEnv*, *commEnv*, *startWebLogic*, and *stopWebLogic*.
- Are there any specific parameters passed to the JVM?
- Are there JARs added to the server classpath?

## Determine whether Management over REST is used

If the lifecycle of your application includes using Management over REST, you need to capture which ports are used to access the REST API and determine how they are authenticated and exposed. After migration, you'll need to ensure that these same ports and authentication mechanisms are exposed so your application lifecycle can operate in a similar fashion as before the migration. For more information, see [Administering Oracle WebLogic Server with RESTful Management Services](#).

## Determine whether a connection to on-premises is needed

If your application needs to access any of your on-premises services, you'll need to provision one of Azure's connectivity services. For more information, see [Choose a solution for connecting an on-premises network to Azure](#). Alternatively, you'll need to

refactor your application to use publicly available APIs that your on-premises resources expose.

## Determine whether Java Message Service (JMS) Queues or Topics are in use

If your application is using JMS Queues or Topics, you'll need to migrate them to an externally-hosted JMS server. Azure Service Bus and the Advanced Message Queuing Protocol can be a great migration strategy for those using JMS. For more information, see [Use JMS with Azure Service Bus and AMQP 1.0](#).

If JMS persistent stores have been configured, you must capture their configuration and apply it after the migration.

If you're using Oracle Message Broker, you can migrate this software to Azure virtual machines and use it as-is.

## Determine whether you are using your own custom created Shared Java EE Libraries

If you're using the Shared Java EE library feature, you have two options:

- Refactor your application code to remove all dependencies on your libraries, and instead incorporate the functionality directly into your application.
- Add the libraries to the server classpath.

## Determine whether OSGi bundles are used

If you used OSGi bundles added to the WebLogic server, you'll need to add the equivalent JAR files directly to your web application.

## Determine whether your application contains OS-specific code

If your application contains any code with dependencies on the host OS, then you'll need to refactor it to remove those dependencies. For example, you may need to replace any use of `/` or `\` in file system paths with [File.Separator](#) or [Paths.get](#).

## Determine whether Oracle Service Bus is in use

If your application is using Oracle Service Bus (OSB), you'll need to capture how OSB is configured. For more information, see [About the Oracle Service Bus Installation](#).

## Determine whether your application is composed of multiple WARs

If your application is composed of multiple WARs, you should treat each of those WARs as separate applications and go through this guide for each of them.

## Determine whether your application is packaged as an EAR

If your application is packaged as an EAR file, be sure to examine the *application.xml* and *weblogic-application.xml* files and capture their configurations.

## Identify all outside processes and daemons running on the production servers

If you have any processes running outside the application server, such as monitoring daemons, you'll need to eliminate them or migrate them elsewhere.

## Determine whether WebLogic Scripting Tool (WLST) is used

If you currently use WLST to perform your deployment, you'll need to assess what it's doing. If WLST is changing any (runtime) parameters of your application as part of the deployment, you'll need to make sure that this behavior continues to work while testing your application after migration.

## Determine whether and how the file system is used

VM filesystems operate the same way as on-premises filesystems with respect to persistence, startup, and shutdown. Even so, it's important to be aware of your filesystem needs and ensure the VMs have adequate storage size and performance.

### Read-only static content

If your application currently serves static content, you'll need an alternate location for it. You may wish to consider moving static content to Azure Blob Storage and adding

Azure CDN for lightning-fast downloads globally. For more information, see [Static website hosting in Azure Storage](#) and [Quickstart: Integrate an Azure storage account with Azure CDN](#). You can also directly deploy the static content to an app in the Azure Spring Apps Enterprise plan. For more information, see [Deploy web static files](#).

## Dynamically published static content

If your application allows for static content that is uploaded/produced by your application but is immutable after its creation, you can use Azure Blob Storage and Azure CDN as described above, with an Azure Function to handle uploads and CDN refresh. We've provided a sample implementation for your use at [Uploading and CDN-preloading static content with Azure Functions](#). You can also directly deploy the static content to an app in the Azure Spring Apps Enterprise plan. For more information, see [Deploy web static files](#).

## Determine the network topology

The current set of Azure Marketplace offers is a starting point for your migration. If the offer does not cover aspects of your architecture that you need to migrate, you'll need to capture the network topology of your existing deployment and reproduce that in Azure, even after standing up the basic offer with one of the solution templates.

This is a very broad topic, but the following references can give some direction to your migration efforts:

- This reference enumerates the high level topics relevant to the migration of network topology to Azure: [Fast Track Deployment Guide](#).
- This reference describes important concerns regarding clustering, which has an impact on network topology: [WebLogic Server Clustering](#).
- Because data sources are separate servers in a WebLogic system, you must consider them as part of the network topology analysis. [WebLogic Server Data Sources](#).
- Messaging sources are also separate servers. [WebLogic Server Messaging](#).
- Load balancing is a fundamental requirement. This reference covers the WebLogic Server side of load balancing: [Load Balancing in a Cluster](#).

## Account for the use of JCA Adapters and Resource Adapters

If your existing application is using JCA Adapters and/or Resource Adapters to connect to other enterprise systems, ensure that the configuration for these artifacts is applied to

the WebLogic Server running in Azure Virtual Machines. For more information, see [Creating and Configuring Resource Adapters](#)

## Account for the use of custom security providers and JAAS

If your application is using JAAS, you need to make sure the configuration of security providers is correctly migrated. For more information, see [About Configuring WebLogic Security Providers](#) in the Oracle documentation.

## Determine whether WebLogic clustering is used

Most likely, you've deployed your application on multiple WebLogic servers to achieve high availability. You can migrate these clusters directly from your on-premises installation to WebLogic running in Azure Virtual Machines. For more information, see [Domain Configuration Files](#) in the Oracle documentation.

## Account for load-balancing requirements

Load balancing is an essential part of migrating your Oracle WebLogic Server cluster to Azure. The easiest solution is to use the built-in support for [Azure Application Gateway](#) provided in the Azure Marketplace offer for Oracle WebLogic Server cluster. For a tutorial on this topic, see [Tutorial: Migrate a WebLogic Server cluster to Azure with Azure Application Gateway as a load balancer](#).

For a summary of the capabilities of Azure Application Gateway compared to other Azure load-balancing solutions, see [Overview of load-balancing options in Azure](#).

## Determine whether the Java EE Application Client feature is used

If your application uses the Java EE Application Client feature, it should continue to work unchanged after migrating to Azure Virtual Machines. For more information, see [Using Java EE Client Application Modules](#).

# Migration

## Select a WebLogic on Azure Virtual Machines offer

The following offers are available for WebLogic on Azure Virtual Machines.

During the deployment of an offer, you're asked to choose the Virtual Machine size for your WebLogic server nodes. It's important to consider all aspects of sizing (memory, processor, disk) in your choice of VM size. For more information, see the [Azure Documentation for virtual machine sizing](#)

## WebLogic Server Single Node with no Admin Server

This offer creates a single VM and installs WebLogic on it, but doesn't configure any domains, which is useful for scenarios where you have a highly customized domain configuration.

## WebLogic Server Single Node with Admin Server

This offer provisions a single VM and installs WebLogic Server on it. It creates a domain and starts up the admin server.

## WebLogic Server N-Node Cluster

This offer creates a highly available cluster of WebLogic Server VMs.

## WebLogic Server N-Node Dynamic Cluster

This offer creates a highly available and scalable dynamic cluster of WebLogic Server VMs

## Provision the offer

After you've selected which offer to start with, follow the instructions in [documentation for the offers](#) to provision that offer. Make sure to choose the domain name that matches your existing domain name. You can even match the domain password with your existing domain password.

## Migrate the domains

After you've provisioned the offer, you can examine the domain configuration and follow [this guidance](#) for details on how to migrate the domains.

## Connect the databases

After you've migrated the domains, you can connect the databases by following the instructions [in the offer documentation](#). These instructions help you account for any database secrets and access strings involved.

## Account for KeyStores

You must account for the migration of any SSL KeyStores used by your application. For more information, see [Configuring Keystores](#).

## Connect the JMS sources

After you've connected the databases, you can configure JMS. For more information, see [Fusion Middleware Administering JMS Resources for Oracle WebLogic Server](#) in the WebLogic documentation.

## Account for authentication and authorization

Most applications have some kind of authentication and authorization. If you use LDAP for authentication, you can set up Microsoft Entra Domain Services with secure LDAP and configure LDAP connections in WebLogic Server. For more information, see [Create and configure a Microsoft Entra Domain Services managed domain](#) and [Configure secure LDAP for a Microsoft Entra Domain Services managed domain](#).

## Account for logging

Use the integration with Elastic on Azure provided by the Oracle WebLogic Server marketplace solution templates. This approach is the easiest way to account for logging. You can see the list of offers in the overview article [What are solutions for running Oracle WebLogic Server on Azure Virtual Machines?](#) Complete tutorials to configure Elastic are provided in:

- [Land Oracle WebLogic Server logs to Elasticsearch and Kibana in admin offer](#)
- [Land Oracle WebLogic Server logs to Elasticsearch and Kibana in cluster offer](#)
- [Land Oracle WebLogic Server logs to Elasticsearch and Kibana in dynamic cluster offer](#)

If the Elastic integration isn't appropriate, you should carry over the existing logging configuration when you migrate the domain. For more information, see [Configure java.util.logging logger levels](#) and [Configuring Log Files and Filtering Log Messages for Oracle WebLogic Server](#) in the Oracle documentation.

# Migrating your applications

The techniques used to deploy applications from the development team into test, staging, and production servers vary greatly from case to case. In some cases, there's a highly evolved CI/CD platform that results in the applications being deployed to the WebLogic Server. In other cases, the process can be more manual. One benefit of using Azure Virtual Machines to migrate WebLogic applications to the cloud is that your existing processes continue to work.

You have to configure the Network Security Group that the offer provisions to allow access from your CI/CD pipeline or manual deployment system. For more information, see [Network security groups](#).

## Testing

Any in-container tests against applications must be configured to access the new servers running within Azure. As with the CI/CD concerns, you must ensure the necessary network security rules allow your tests to access the applications deployed to Azure. For more information, see [Network security groups](#).

## Post-migration

After you've reached the migration goals you defined in the [pre-migration](#) step, perform some end-to-end acceptance testing to verify that everything works as expected. For guidance on some potential post-migration enhancements, see the following recommendations:

- Using Azure Storage to serve static content mounted to the virtual machines. For more information, see [Attach or detach a data disk to a virtual machine](#).
- Deploy your applications to your migrated WebLogic cluster with Azure DevOps. For more information, see [Azure DevOps getting started documentation](#).
- If you deployed WebLogic Server with Azure Application Gateway by following the steps in [Tutorial: Migrate a WebLogic Server cluster to Azure with Azure Application Gateway as a load balancer](#), you may want to do more configuration on the Application Gateway. For more information, see [Application Gateway configuration overview](#).
- Enhance your network topology with advanced load balancing services. For more information, see [Using load-balancing services in Azure](#).

- Use Azure Managed Identities to manage secrets and assign role based access to Azure resources. For more information, see [What are managed identities for Azure resources?](#)
- Integrate WebLogic Java EE authentication and authorization with Microsoft Entra ID. For more information, see [Integrating Microsoft Entra getting started guide](#).
- Use Azure Key Vault to store any information that functions as a "secret". For more information, see [Azure Key Vault basic concepts](#).

# Tutorial: Configure secure LDAP for a Microsoft Entra Domain Services managed domain

Article • 10/06/2023

To communicate with your Microsoft Entra Domain Services managed domain, the Lightweight Directory Access Protocol (LDAP) is used. By default, the LDAP traffic isn't encrypted, which is a security concern for many environments.

With Microsoft Entra Domain Services, you can configure the managed domain to use secure Lightweight Directory Access Protocol (LDAPS). When you use secure LDAP, the traffic is encrypted. Secure LDAP is also known as LDAP over Secure Sockets Layer (SSL) / Transport Layer Security (TLS).

This tutorial shows you how to configure LDAPS for a Domain Services managed domain.

In this tutorial, you learn how to:

- ✓ Create a digital certificate for use with Microsoft Entra Domain Services
- ✓ Enable secure LDAP for Microsoft Entra Domain Services
- ✓ Configure secure LDAP for use over the public internet
- ✓ Bind and test secure LDAP for a managed domain

If you don't have an Azure subscription, [create an account](#) before you begin.

## Prerequisites

To complete this tutorial, you need the following resources and privileges:

- An active Azure subscription.
  - If you don't have an Azure subscription, [create an account](#).
- A Microsoft Entra tenant associated with your subscription, either synchronized with an on-premises directory or a cloud-only directory.
  - If needed, [create a Microsoft Entra tenant](#) or [associate an Azure subscription with your account](#).
- A Microsoft Entra Domain Services managed domain enabled and configured in your Microsoft Entra tenant.
  - If needed, [create and configure a Microsoft Entra Domain Services managed domain](#).

- The *LDPExe* tool installed on your computer.
  - If needed, [install the Remote Server Administration Tools \(RSAT\)](#) for Active Directory Domain Services and LDAP.
- You need [Application Administrator](#) and [Groups Administrator](#) Microsoft Entra roles in your tenant to enable secure LDAP.

## Sign in to the Microsoft Entra admin center

In this tutorial, you configure secure LDAP for the managed domain using the Microsoft Entra admin center. To get started, first sign in to the [Microsoft Entra admin center](#).

## Create a certificate for secure LDAP

To use secure LDAP, a digital certificate is used to encrypt the communication. This digital certificate is applied to your managed domain, and lets tools like *LDPExe* use secure encrypted communication when querying data. There are two ways to create a certificate for secure LDAP access to the managed domain:

- A certificate from a public certificate authority (CA) or an enterprise CA.
  - If your organization gets certificates from a public CA, get the secure LDAP certificate from that public CA. If you use an enterprise CA in your organization, get the secure LDAP certificate from the enterprise CA.
  - A public CA only works when you use a custom DNS name with your managed domain. If the DNS domain name of your managed domain ends in *.onmicrosoft.com*, you can't create a digital certificate to secure the connection with this default domain. Microsoft owns the *.onmicrosoft.com* domain, so a public CA won't issue a certificate. In this scenario, create a self-signed certificate and use that to configure secure LDAP.
- A self-signed certificate that you create yourself.
  - This approach is good for testing purposes, and is what this tutorial shows.

The certificate you request or create must meet the following requirements. Your managed domain encounters problems if you enable secure LDAP with an invalid certificate:

- **Trusted issuer** - The certificate must be issued by an authority trusted by computers connecting to the managed domain using secure LDAP. This authority may be a public CA or an Enterprise CA trusted by these computers.
- **Lifetime** - The certificate must be valid for at least the next 3-6 months. Secure LDAP access to your managed domain is disrupted when the certificate expires.

- **Subject name** - The subject name on the certificate must be your managed domain. For example, if your domain is named *aaddscontoso.com*, the certificate's subject name must be *\*.aaddscontoso.com*.
  - The DNS name or subject alternate name of the certificate must be a wildcard certificate to ensure the secure LDAP works properly with Domain Services. Domain Controllers use random names and can be removed or added to ensure the service remains available.
- **Key usage** - The certificate must be configured for *digital signatures* and *key encipherment*.
- **Certificate purpose** - The certificate must be valid for TLS server authentication.

There are several tools available to create self-signed certificate such as OpenSSL, Keytool, MakeCert, [New-SelfSignedCertificate](#) cmdlet, etc.

In this tutorial, let's create a self-signed certificate for secure LDAP using the [New-SelfSignedCertificate](#) cmdlet.

Open a PowerShell window as **Administrator** and run the following commands. Replace the `$dnsName` variable with the DNS name used by your own managed domain, such as *aaddscontoso.com*:

```
PowerShell

# Define your own DNS name used by your managed domain
$dnsName="aaddscontoso.com"

# Get the current date to set a one-year expiration
$lifetime=Get-Date

# Create a self-signed certificate for use with Azure AD DS
New-SelfSignedCertificate -Subject *.{$dnsName} `
    -NotAfter $lifetime.AddDays(365) -KeyUsage DigitalSignature,
KeyEncipherment `
    -Type SSLServerAuthentication -DnsName *.$dnsName, $dnsName
```

The following example output shows that the certificate was successfully generated and is stored in the local certificate store (*LocalMachine\MY*):

```
Output

PS C:\WINDOWS\system32> New-SelfSignedCertificate -Subject *.{$dnsName} `
>> -NotAfter $lifetime.AddDays(365) -KeyUsage DigitalSignature,
KeyEncipherment `
>> -Type SSLServerAuthentication -DnsName *.$dnsName, $dnsName.com

PSParentPath: Microsoft.PowerShell.Security\Certificate::LocalMachine\MY
```

Thumbprint	Subject
-----	-----
959BD1531A1E674EB09E13BD8534B2C76A45B3E6	CN=aaddscontoso.com

## Understand and export required certificates

To use secure LDAP, the network traffic is encrypted using public key infrastructure (PKI).

- A **private** key is applied to the managed domain.
  - This private key is used to *decrypt* the secure LDAP traffic. The private key should only be applied to the managed domain and not widely distributed to client computers.
  - A certificate that includes the private key uses the *.PFX* file format.
  - When exporting the certificate, you must specify the *TripleDES-SHA1* encryption algorithm. This is applicable to the *.pfx* file only and does not impact the algorithm used by the certificate itself. Note that the *TripleDES-SHA1* option is available only beginning with Windows Server 2016.
- A **public** key is applied to the client computers.
  - This public key is used to *encrypt* the secure LDAP traffic. The public key can be distributed to client computers.
  - Certificates without the private key use the *.CER* file format.

These two keys, the *private* and *public* keys, make sure that only the appropriate computers can successfully communicate with each other. If you use a public CA or enterprise CA, you are issued with a certificate that includes the private key and can be applied to a managed domain. The public key should already be known and trusted by client computers.

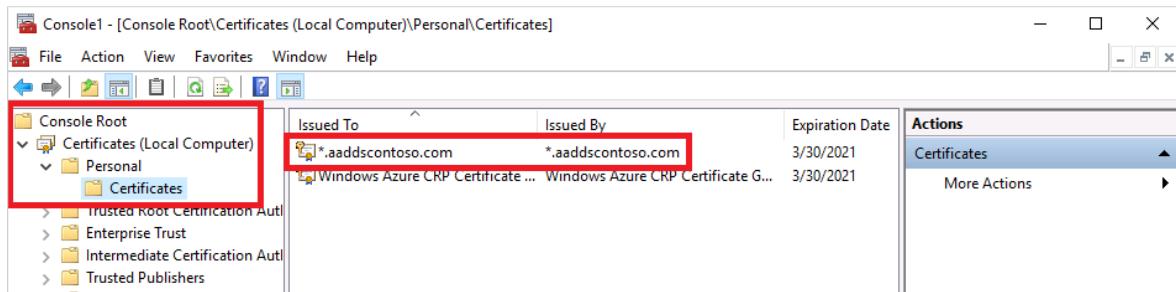
In this tutorial, you created a self-signed certificate with the private key, so you need to export the appropriate private and public components.

## Export a certificate for Microsoft Entra Domain Services

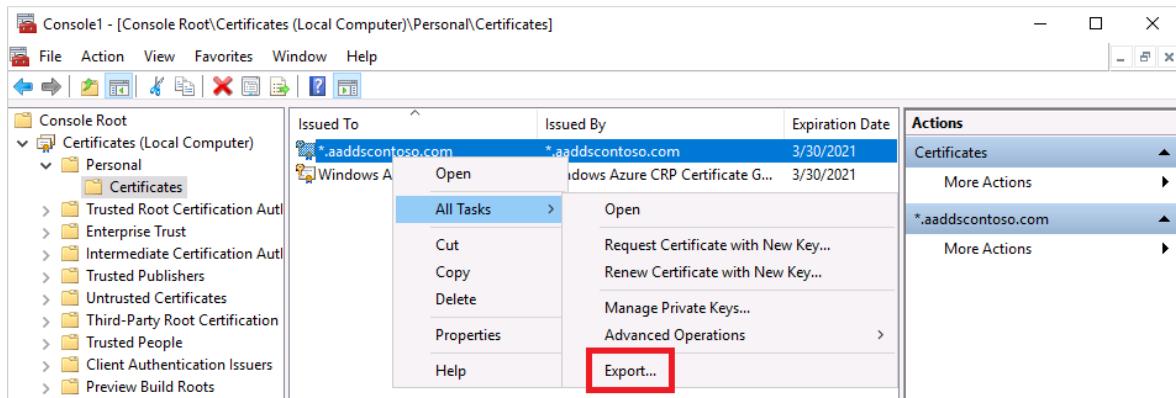
Before you can use the digital certificate created in the previous step with your managed domain, export the certificate to a *.PFX* certificate file that includes the private key.

1. To open the *Run* dialog, select the **Windows + R** keys.
2. Open the Microsoft Management Console (MMC) by entering **mmc** in the *Run* dialog, then select **OK**.

3. On the User Account Control prompt, then select Yes to launch MMC as administrator.
4. From the File menu, select Add/Remove Snap-in...
5. In the Certificates snap-in wizard, choose Computer account, then select Next.
6. On the Select Computer page, choose Local computer: (the computer this console is running on), then select Finish.
7. In the Add or Remove Snap-ins dialog, select OK to add the certificates snap-in to MMC.
8. In the MMC window, expand Console Root. Select Certificates (Local Computer), then expand the Personal node, followed by the Certificates node.



9. The self-signed certificate created in the previous step is shown, such as *aaddscontoso.com*. Right-select this certificate, then choose All Tasks > Export...

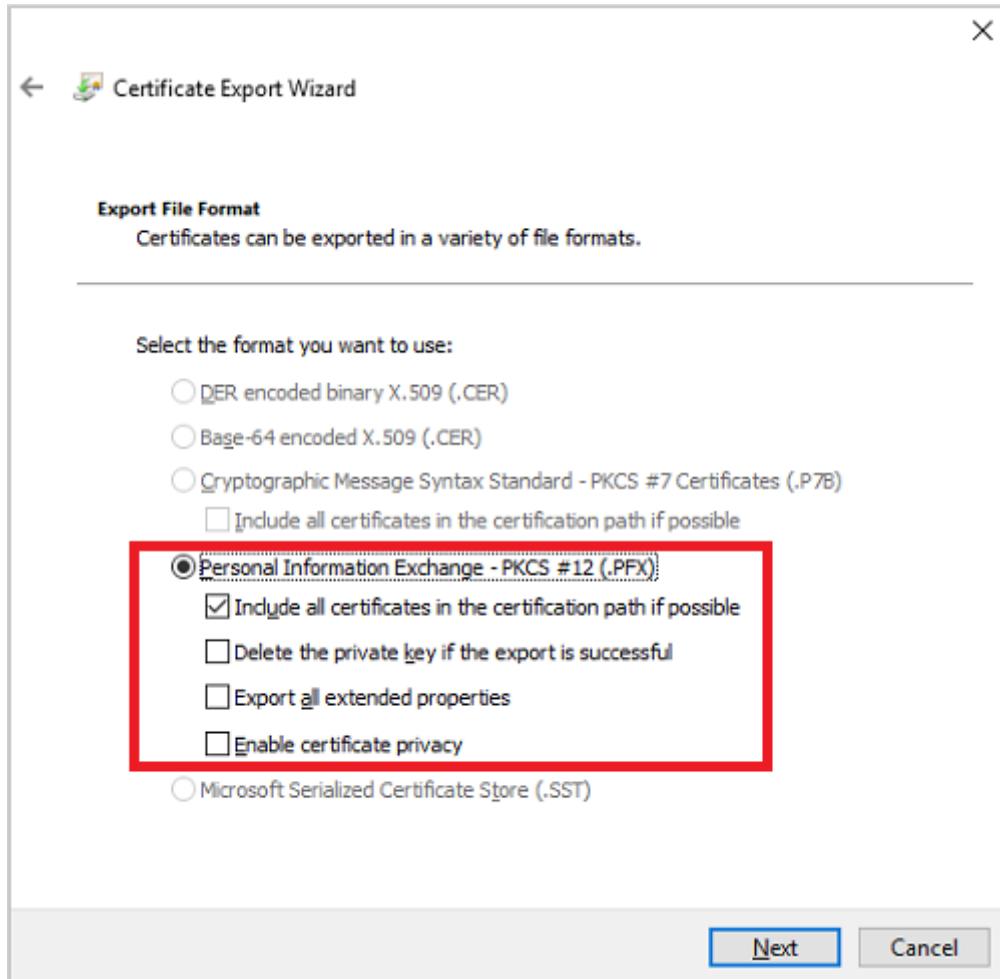


10. In the Certificate Export Wizard, select Next.
11. The private key for the certificate must be exported. If the private key is not included in the exported certificate, the action to enable secure LDAP for your managed domain fails.

On the Export Private Key page, choose Yes, export the private key, then select Next.

12. Managed domains only support the **.PFX** certificate file format that includes the private key. Don't export the certificate as **.CER** certificate file format without the private key.

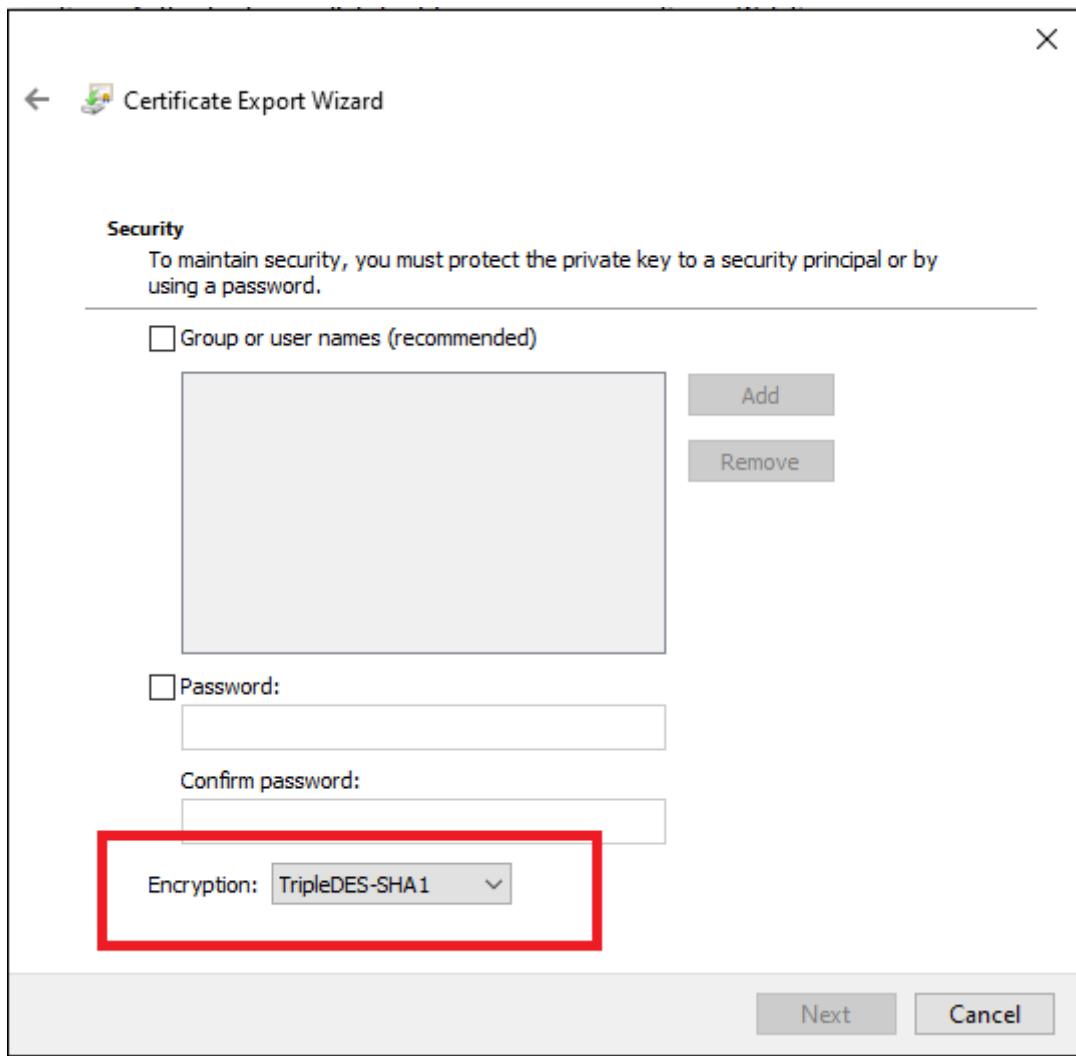
On the **Export File Format** page, select **Personal Information Exchange - PKCS #12 (.PFX)** as the file format for the exported certificate. Check the box for *Include all certificates in the certification path if possible*:



13. As this certificate is used to decrypt data, you should carefully control access. A password can be used to protect the use of the certificate. Without the correct password, the certificate can't be applied to a service.

On the **Security** page, choose the option for **Password** to protect the **.PFX** certificate file. The encryption algorithm must be *TripleDES-SHA1*. Enter and confirm a password, then select **Next**. This password is used in the next section to enable secure LDAP for your managed domain.

If you export using the **PowerShell export-pfxcertificate cmdlet**, you need to pass the **-CryptoAlgorithmOption** flag using **TripleDES\_SHA1**.



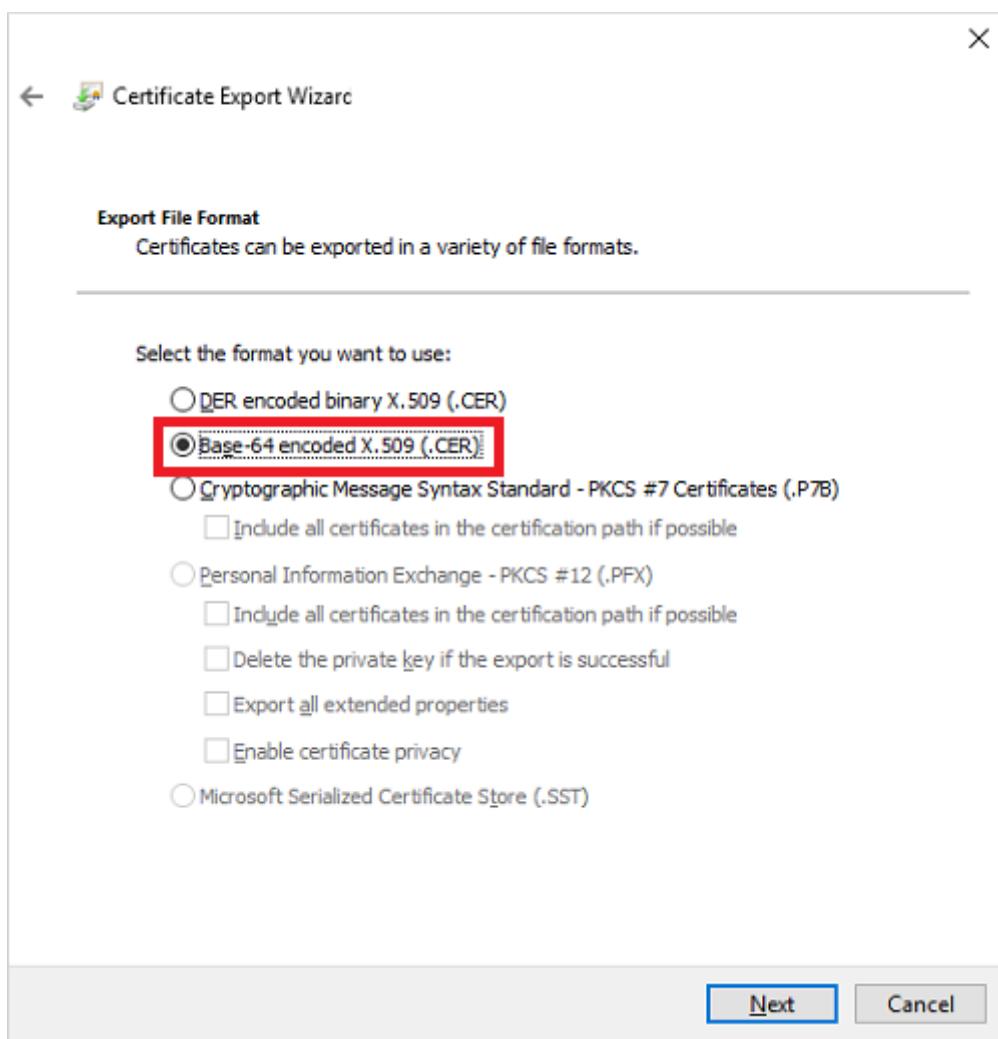
14. On the **File to Export** page, specify the file name and location where you'd like to export the certificate, such as `C:\Users\<account-name>\azure-ad-ds.pfx`. Keep a note of the password and location of the `.PFX` file as this information would be required in next steps.
15. On the review page, select **Finish** to export the certificate to a `.PFX` certificate file. A confirmation dialog is displayed when the certificate has been successfully exported.
16. Leave the MMC open for use in the following section.

## Export a certificate for client computers

Client computers must trust the issuer of the secure LDAP certificate to be able to connect successfully to the managed domain using LDAPS. The client computers need a certificate to successfully encrypt data that is decrypted by Domain Services. If you use a public CA, the computer should automatically trust these certificate issuers and have a corresponding certificate.

In this tutorial you use a self-signed certificate, and generated a certificate that includes the private key in the previous step. Now let's export and then install the self-signed certificate into the trusted certificate store on the client computer:

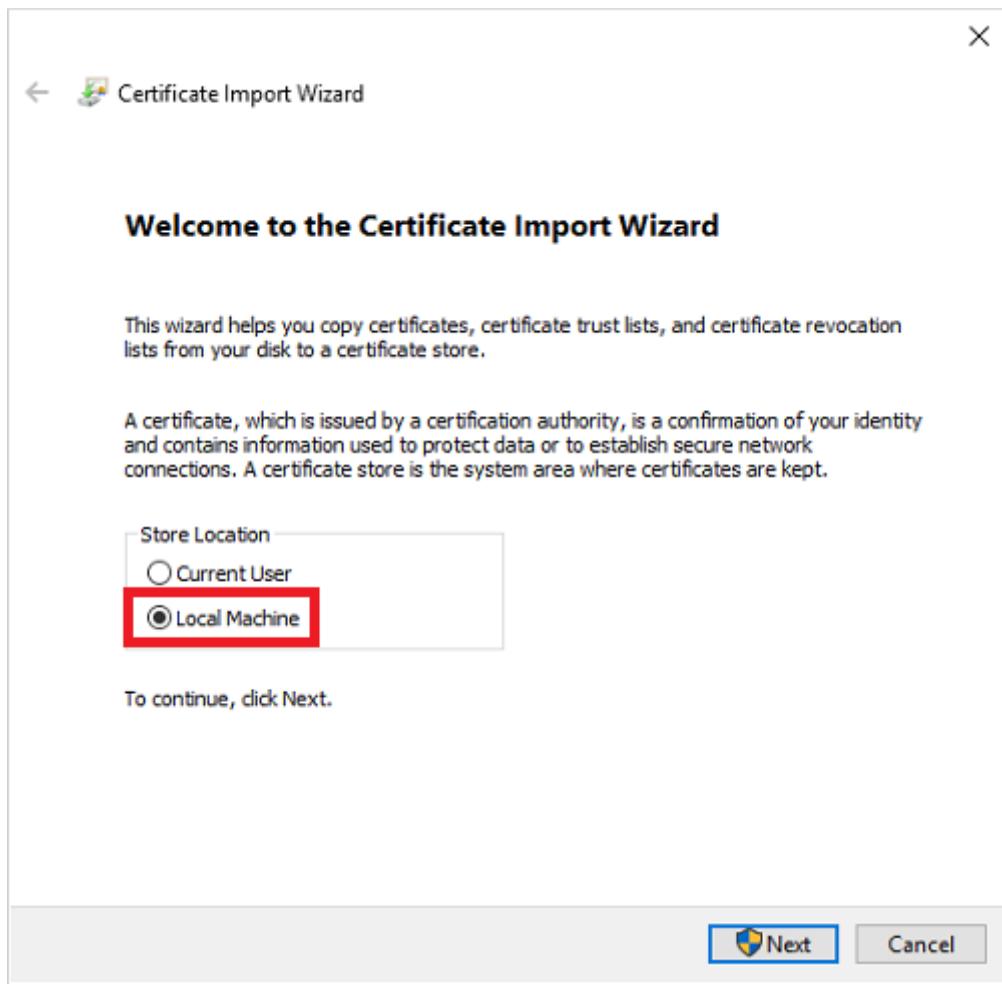
1. Go back to the MMC for *Certificates (Local Computer) > Personal > Certificates* store. The self-signed certificate created in a previous step is shown, such as *aaddscontoso.com*. Right-select this certificate, then choose **All Tasks > Export...**
2. In the **Certificate Export Wizard**, select **Next**.
3. As you don't need the private key for clients, on the **Export Private Key** page choose **No, do not export the private key**, then select **Next**.
4. On the **Export File Format** page, select **Base-64 encoded X.509 (.CER)** as the file format for the exported certificate:



5. On the **File to Export** page, specify the file name and location where you'd like to export the certificate, such as `C:\Users\<account-name>\azure-ad-ds-client.cer`.
6. On the review page, select **Finish** to export the certificate to a *.CER* certificate file. A confirmation dialog is displayed when the certificate has been successfully exported.

The *.CER* certificate file can now be distributed to client computers that need to trust the secure LDAP connection to the managed domain. Let's install the certificate on the local computer.

1. Open File Explorer and browse to the location where you saved the *.CER* certificate file, such as `C:\Users\<account-name>\azure-ad-ds-client.cer`.
2. Right-select the *.CER* certificate file, then choose **Install Certificate**.
3. In the **Certificate Import Wizard**, choose to store the certificate in the *Local machine*, then select **Next**:



4. When prompted, choose **Yes** to allow the computer to make changes.
5. Choose to **Automatically select the certificate store based on the type of certificate**, then select **Next**.
6. On the review page, select **Finish** to import the *.CER* certificate. A confirmation dialog is displayed when the certificate has been successfully imported.

# Enable secure LDAP for Microsoft Entra Domain Services

With a digital certificate created and exported that includes the private key, and the client computer set to trust the connection, now enable secure LDAP on your managed domain. To enable secure LDAP on a managed domain, perform the following configuration steps:

1. In the [Microsoft Entra admin center](#), enter *domain services* in the **Search resources** box. Select **Microsoft Entra Domain Services** from the search result.
2. Choose your managed domain, such as *aaddscontoso.com*.
3. On the left-hand side of the Microsoft Entra Domain Services window, choose **Secure LDAP**.
4. By default, secure LDAP access to your managed domain is disabled. Toggle **Secure LDAP to Enable**.
5. Secure LDAP access to your managed domain over the internet is disabled by default. When you enable public secure LDAP access, your domain is susceptible to password brute force attacks over the internet. In the next step, a network security group is configured to lock down access to only the required source IP address ranges.  
Toggle **Allow secure LDAP access over the internet** to **Enable**.
6. Select the folder icon next to **.PFX file with secure LDAP certificate**. Browse to the path of the *.PFX* file, then select the certificate created in a previous step that includes the private key.

## Important

As noted in the previous section on certificate requirements, you can't use a certificate from a public CA with the default *.onmicrosoft.com* domain.

Microsoft owns the *.onmicrosoft.com* domain, so a public CA won't issue a certificate.

Make sure your certificate is in the appropriate format. If it's not, the Azure platform generates certificate validation errors when you enable secure LDAP.

7. Enter the **Password to decrypt .PFX file** set in a previous step when the certificate was exported to a *.PFX* file.

## 8. Select Save to enable secure LDAP.

The screenshot shows the Azure AD Domain Services blade for a domain named 'aaddscontoso.com'. On the left, there's a navigation menu with options like Overview, Activity log, Access control (IAM), Settings (Properties, Secure LDAP, Synchronization, Health, Notification settings, SKU), Monitoring (Diagnostic settings, Logs, Workbooks), and Support + troubleshooting. The 'Secure LDAP' option is selected and highlighted with a red box. On the right, under the 'Secure LDAP' section, there are two main settings: 'Allow secure LDAP access over the internet' (status: Enabled) and 'Upload a .PFX file containing the certificate to be used for secure LDAP access to this managed domain'. A .PFX file named 'azure-ad-ds.pfx' is uploaded, and a password is entered in the 'Password to decrypt .PFX file' field. A warning message at the bottom states: 'Your subnet is protected by network security group aadds-nsg. To give user access to secure LDAP endpoint, please ensure "Allow" rule on port 636 is configured with proper IP ranges on the network security group.'

A notification is displayed that secure LDAP is being configured for the managed domain. You can't modify other settings for the managed domain until this operation is complete.

It takes a few minutes to enable secure LDAP for your managed domain. If the secure LDAP certificate you provide doesn't match the required criteria, the action to enable secure LDAP for the managed domain fails.

Some common reasons for failure are if the domain name is incorrect, the encryption algorithm for the certificate isn't *TripleDES-SHA1*, or the certificate expires soon or has already expired. You can re-create the certificate with valid parameters, then enable secure LDAP using this updated certificate.

## Change an expiring certificate

1. Create a replacement secure LDAP certificate by following the steps to [create a certificate for secure LDAP](#).
2. To apply the replacement certificate to Domain Services, in the left menu for **Microsoft Entra Domain Services** in the Microsoft Entra admin center, select **Secure LDAP**, and then select **Change Certificate**.
3. Distribute the certificate to any clients that connect by using secure LDAP.

## Lock down secure LDAP access over the internet

When you enable secure LDAP access over the internet to your managed domain, it creates a security threat. The managed domain is reachable from the internet on TCP port 636. It's recommended to restrict access to the managed domain to specific known IP addresses for your environment. An Azure network security group rule can be used to limit access to secure LDAP.

Let's create a rule to allow inbound secure LDAP access over TCP port 636 from a specified set of IP addresses. A default *DenyAll* rule with a lower priority applies to all other inbound traffic from the internet, so only the specified addresses can reach your managed domain using secure LDAP.

1. In the [Microsoft Entra admin center](#), search for and select *Resource groups*.
2. Choose your resource group, such as *myResourceGroup*, then select your network security group, such as *aaads-nsg*.
3. The list of existing inbound and outbound security rules are displayed. On the left-hand side of the network security group window, choose **Settings > Inbound security rules**.
4. Select **Add**, then create a rule to allow *TCP port 636*. For improved security, choose the source as *IP Addresses* and then specify your own valid IP address or range for your organization.

Setting	Value
Source	IP Addresses
Source IP addresses / CIDR ranges	A valid IP address or range for your environment
Source port ranges	*
Destination	Any
Destination port ranges	636
Protocol	TCP
Action	Allow
Priority	401
Name	AllowLDAPS

5. When ready, select **Add** to save and apply the rule.

The screenshot shows the Azure portal interface for managing Network Security Groups (NSGs). On the left, a sidebar lists various settings like Overview, Activity log, Access control (IAM), Tags, Diagnose and solve problems, Settings, Inbound security rules (highlighted with a red box), Outbound security rules, Network interfaces, Subnets, Properties (highlighted with a red box), Locks, Export template, Monitoring, Diagnostic settings, Logs, NSG flow logs, Support + troubleshooting, Effective security rules, and New support request. The main pane displays the 'Inbound security rules' for the 'aadds-nsg' NSG. It shows a table of existing rules with columns for Priority, Name, and Action. A red box highlights the '+ Add' button. The right side of the screen is a detailed configuration form for adding a new rule:

Priority	Name
101	AllowSyncWithAzureAD
201	AllowRD
301	AllowPSRemoting
65000	AllowVnetInBound
65001	AllowAzureLoadBalancer
65500	DenyAllInBound

**Add inbound security rule**

**Source \***

IP Addresses

Source IP addresses/CIDR ranges \*

Source port ranges \*

Destination \*

Destination port ranges \*

Protocol \*

Any  TCP  UDP  ICMP

Action \*

Allow  Deny

Priority \*

401

Name \*

AllowLDAP

Description

Add

## Configure DNS zone for external access

With secure LDAP access enabled over the internet, update the DNS zone so that client computers can find this managed domain. The *Secure LDAP external IP address* is listed on the **Properties** tab for your managed domain:

The screenshot shows the Azure portal interface for managing a managed domain. On the left, a sidebar lists Overview, Activity log, Access control (IAM), Settings, Properties (highlighted with a red box), Secure LDAP, Synchronization, and Health. The main pane displays the properties for the domain 'aaddscontoso.com':

**aaddscontoso.com | Properties**

Secure LDAP

Enabled

Secure LDAP certificate thumbprint

21CB8B92D71331F38B47281F96F9E9B8330...

Secure LDAP certificate expires

Tue, 30 Mar 2021 23:51:43 GMT

Secure LDAP external IP address

168.62.205.103

Configure your external DNS provider to create a host record, such as *ldaps*, to resolve to this external IP address. To test locally on your machine first, you can create an entry in the Windows hosts file. To successfully edit the hosts file on your local machine, open *Notepad* as an administrator, then open the file `C:\Windows\System32\drivers\etc\hosts`.

The following example DNS entry, either with your external DNS provider or in the local hosts file, resolves traffic for `ldaps.aaddscontoso.com` to the external IP address of `168.62.205.103`:



## Test queries to the managed domain

To connect and bind to your managed domain and search over LDAP, you use the *LDPe* tool. This tool is included in the Remote Server Administration Tools (RSAT) package. For more information, see [install Remote Server Administration Tools](#).

1. Open *LDPe* and connect to the managed domain. Select **Connection**, then choose **Connect....**
2. Enter the secure LDAP DNS domain name of your managed domain created in the previous step, such as *ldaps.aaddscontoso.com*. To use secure LDAP, set **Port** to 636, then check the box for **SSL**.
3. Select **OK** to connect to the managed domain.

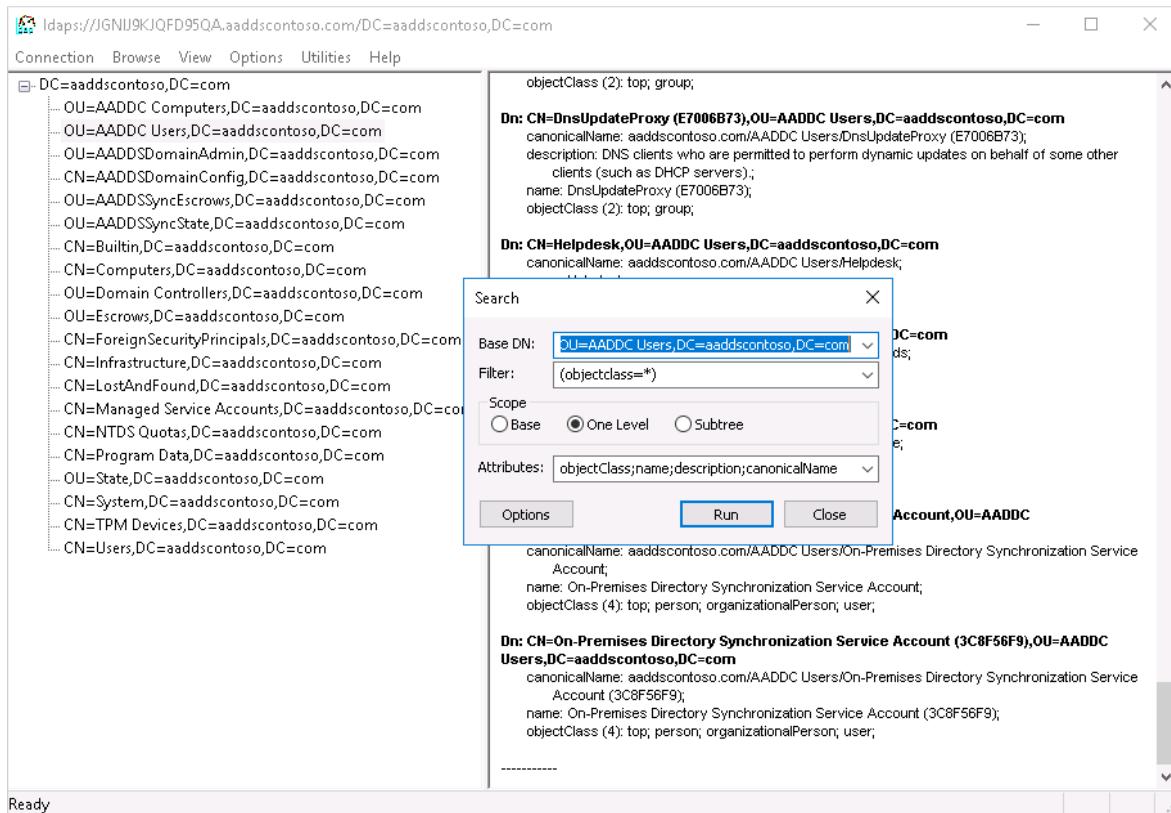
Next, bind to your managed domain. Users (and service accounts) can't perform LDAP simple binds if you have disabled NTLM password hash synchronization on your managed domain. For more information on disabling NTLM password hash synchronization, see [Secure your managed domain](#).

1. Select the **Connection** menu option, then choose **Bind....**
2. Provide the credentials of a user account that belongs to the managed domain. Enter the user account's password, then enter your domain, such as *aaddscontoso.com*.
3. For **Bind type**, choose the option for *Bind with credentials*.
4. Select **OK** to bind to your managed domain.

To see of the objects stored in your managed domain:

1. Select the **View** menu option, and then choose **Tree**.
2. Leave the *BaseDN* field blank, then select **OK**.

3. Choose a container, such as *AADDC Users*, then right-select the container and choose **Search**.
4. Leave the pre-populated fields set, then select **Run**. The results of the query are displayed in the right-hand window, as shown in the following example output:



To directly query a specific container, from the **View > Tree** menu, you can specify a **BaseDN** such as *OU=AADDC Users,DC=AADDSCONTOSO,DC=COM* or *OU=AADDC Computers,DC=AADDSCONTOSO,DC=COM*. For more information on how to format and create queries, see [LDAP query basics](#).

### ⓘ Note

If a Self signed certificate is used, make sure Self signed certificate added on the Trusted Root Certification Authorities for LDAPS to work with LDP.exe

## Clean up resources

If you added a DNS entry to the local hosts file of your computer to test connectivity for this tutorial, remove this entry and add a formal record in your DNS zone. To remove the entry from the local hosts file, complete the following steps:

1. On your local machine, open *Notepad* as an administrator
2. Browse to and open the file `C:\Windows\System32\drivers\etc\hosts`.

3. Delete the line for the record you added, such as `168.62.205.103`

`ldaps.aaddscontoso.com`

## Troubleshooting

If you see an error stating that LDAP.exe cannot connect, try working through the different aspects of getting the connection:

1. Configuring the domain controller
2. Configuring the client
3. Networking
4. Establishing the TLS session

For the certificate subject name match, the DC will use the Domain Services domain name (not the Microsoft Entra domain name) to search its certificate store for the certificate. Spelling mistakes, for example, prevent the DC from selecting the right certificate.

The client attempts to establish the TLS connection using the name you provided. The traffic needs to get all the way through. The DC sends the public key of the server auth cert. The cert needs to have the right usage in the certificate, the name signed in the subject name must be compatible for the client to trust that the server is the DNS name which you're connecting to (that is, a wildcard will work, with no spelling mistakes), and the client must trust the issuer. You can check for any problems in that chain in the System log in Event Viewer, and filter the events where source equals Schannel. Once those pieces are in place, they form a session key.

For more information, see [TLS Handshake](#).

## Next steps

In this tutorial, you learned how to:

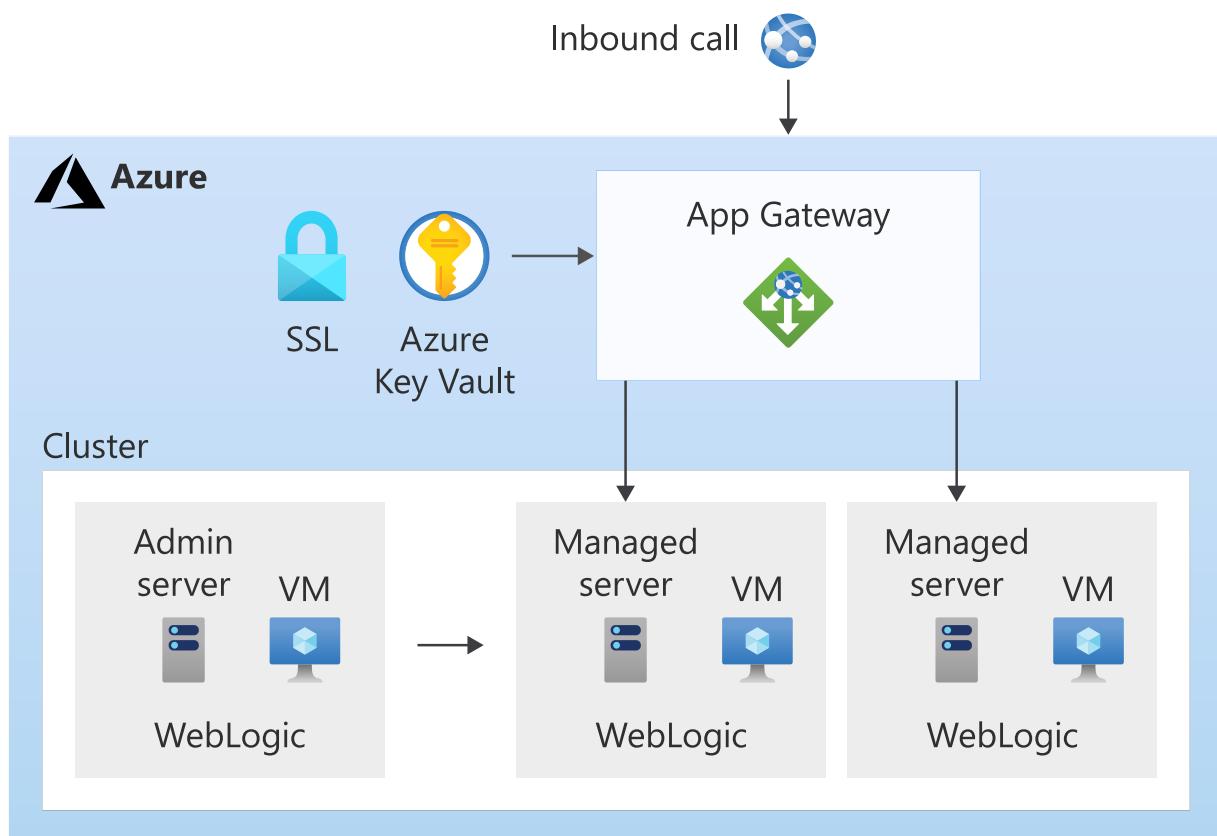
- ✓ Create a digital certificate for use with Microsoft Entra Domain Services
- ✓ Enable secure LDAP for Microsoft Entra Domain Services
- ✓ Configure secure LDAP for use over the public internet
- ✓ Bind and test secure LDAP for a managed domain

[Configure password hash synchronization for a hybrid Microsoft Entra environment](#)

# Tutorial: Migrate a WebLogic Server cluster to Azure with Azure Application Gateway as a load balancer

Article • 05/30/2023

This tutorial walks you through the process of deploying WebLogic Server (WLS) with Azure Application Gateway. It covers the specific steps for creating a Key Vault, storing a TLS/SSL certificate in the Key Vault, and using that certificate for TLS/SSL termination. While all of these elements are well documented in their own right, this tutorial shows the specific way all of these elements come together to create a simple, yet powerful load-balancing solution for WLS on Azure.



Load balancing is an essential part of migrating your Oracle WebLogic Server cluster to Azure. The easiest solution is to use the built-in support for [Azure Application Gateway](#). App Gateway is included as part of the WebLogic Cluster support on Azure. For an overview of WebLogic Cluster support on Azure, see [What is Oracle WebLogic Server on Azure?](#).

In this tutorial, you learn how to:

- ✓ Choose how to provide the TLS/SSL certificate to the App Gateway

- ✓ Deploy WebLogic Server with Azure Application Gateway to Azure
- ✓ Validate successful deployment of WLS and App Gateway

## Prerequisites

- [OpenSSL](#) on a computer running a UNIX-like command-line environment.

While there could be other tools available for certificate management, this tutorial uses OpenSSL. You can find OpenSSL bundled with many GNU/Linux distributions, such as Ubuntu.
- An active Azure subscription.
  - If you don't have an Azure subscription, [create a free account](#).
- The ability to deploy one of the WLS Azure Applications listed at [Oracle WebLogic Server Azure Applications](#).

## Migration context

Here are some things to consider about migrating on-premise WLS installations and Azure Application Gateway. While the steps of this tutorial are the easiest way to stand up a load-balancer in front of your WebLogic Server Cluster on Azure, there are many other ways to do it. This list shows some other things to consider.

- If you have an existing load-balancing solution, ensure that its capabilities are met or exceeded by Azure Application Gateway. For a summary of the capabilities of Azure Application Gateway compared to other Azure load-balancing solutions, see [Overview of load-balancing options in Azure](#).
- If your existing load-balancing solution provides security protection from common exploits and vulnerabilities, the Application Gateway has you covered. Application Gateway's built-in Web Application Firewall (WAF) implements the [OWASP \(Open Web Application Security Project\) core rule sets](#). For more information on WAF support in Application Gateway, see the [Web Application Firewall](#) section of [Azure Application Gateway features](#).
- If your existing load-balancing solution requires end-to-end TLS/SSL encryption, you'll need to do additional configuration after following the steps in this guide. See the [End-to-end TLS encryption](#) section of [Overview of TLS termination and end to end TLS with Application Gateway](#) and the Oracle documentation on [Configuring SSL in Oracle Fusion Middleware](#).
- If you're optimizing for the cloud, this guide shows you how to start from scratch with Azure App Gateway and WLS.

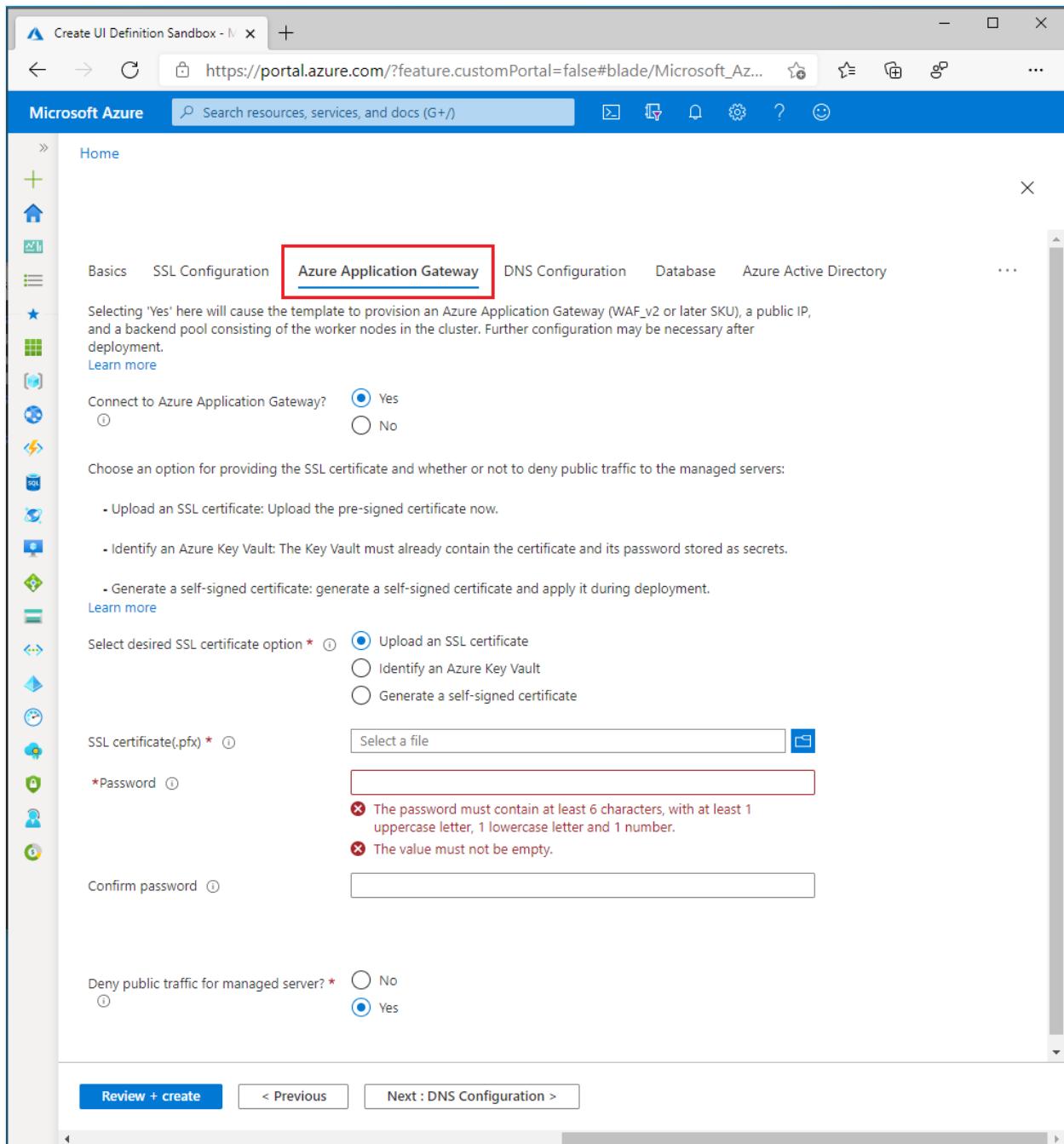
- For a comprehensive survey of migrating WebLogic Server to Azure Virtual Machines, see [Migrate WebLogic Server applications to Azure Virtual Machines](#).

## Deploy WebLogic Server with Application Gateway to Azure

This section will show you how to provision a WLS cluster with Azure Application Gateway automatically created as the load balancer for the cluster nodes. The Application Gateway will use the provided TLS/SSL certificate for TLS/SSL termination. For advanced details on TLS/SSL termination with Application Gateway, see [Overview of TLS termination and end to end TLS with Application Gateway](#).

To create the WLS cluster and Application Gateway, use the following steps.

First, begin the process of deploying a WebLogic Server configured or dynamic cluster as described [in the Oracle documentation](#), but come back to this page when you reach **Azure Application Gateway**, as shown here.



## Choose how to provide the TLS/SSL certificate to the App Gateway

You have several options to provide the TLS/SSL certificate to the application gateway, but can only choose one. This section explains each option so you can choose the best one for your deployment.

### Option one: Upload an TLS/SSL certificate

This option is suitable for production workloads where the App Gateway faces the public Internet, or for intranet workloads that require TLS/SSL. By choosing this option, an

Azure Key Vault is automatically provisioned to contain the TLS/SSL certificate used by the App Gateway.

To upload an existing, signed, TLS/SSL certificate, use the following steps:

1. Follow the steps from your certificate issuer to create a password-protected TLS/SSL certificate and specify the DNS name for the certificate. How to choose wildcard vs. single-name certificate is beyond the scope of this document. Either one will work here.
2. Export the certificate from your issuer using the PFX file format and download it to your local machine. If your issuer doesn't support exporting as PFX, tools exist to convert many certificate formats to PFX format.
3. Select the **Azure Application Gateway** section.
4. Next to **Connect to Azure Application Gateway**, select **Yes**.
5. Select **Upload an SSL certificate**.
6. Select the file browser icon for the field **SSL certificate**. Navigate to the downloaded PFX format certificate and select **Open**.
7. Enter the password for the certificate in the **Password** and **Confirm password** boxes.
8. Choose whether or not to deny public traffic directly to the nodes of the managed servers. Selecting **Yes** will make it so the managed servers are only accessible through the App Gateway.

## Select DNS Configuration

TLS/SSL certificates are associated with a DNS domain name at the time they're issued by the certificate issuer. Follow the steps in this section to configure the deployment with the DNS name for the certificate. You can use a DNS Zone you already have created or allow the deployment to create one for you. Select the **DNS Configuration** section to continue.

## Use an existing Azure DNS Zone

To use an existing Azure DNS Zone with the App Gateway, use the following steps:

1. Next to **Configure Custom DNS Alias**, select **Yes**.
2. Next to **Use an existing Azure DNS Zone** select **Yes**.
3. Enter the name of the Azure DNS Zone next to **DNS Zone Name**.
4. Enter the resource group that contains the Azure DNS Zone from the preceding step.

## Allow the deployment to create a new Azure DNS Zone

To create an Azure DNS Zone to use with the App Gateway, use the following steps:

1. Next to **Configure Custom DNS Alias**, select **Yes**.
2. Next to **Use an existing Azure DNS Zone** select **No**.
3. Enter the name of the Azure DNS Zone next to **DNS Zone Name**. A new DNS Zone will be created in the same resource group as WLS.

Finally, specify the names for the child DNS zones. The deployment will create two child DNS zones for use with WLS: one for the admin console, and one for the App Gateway. For example, if your DNS Zone Name was 'contoso.net', you could enter *admin* and *app* as the values. The admin console would be available at 'admin.contoso.net' and the app gateway would be available at 'app.contoso.net'. Don't forget set up DNS delegation as described in [Delegation of DNS zones with Azure DNS](#).

DNS Zone Name *	contoso.net
Label for Oracle WebLogic Administration Console *	admin
Label for Application Gateway *	app

The other options for providing an TLS/SSL certificate to the App Gateway are detailed in the following sections. If you're satisfied with your chosen option, you can skip to the section [Continue with deployment](#).

## Option two: Identify an Azure Key Vault

This option is suitable for production or non-production workloads, depending on the TLS/SSL certificate provided. If you don't want the deployment to create an Azure Key Vault, you can identify an existing one or create one yourself. This option requires you to store the certificate and its password in the Azure Key Vault before continuing. If you have an existing Key Vault you want to use, skip to the section [Create a TLS/SSL certificate](#). Otherwise, continue to the next section.

### Create an Azure Key Vault

This section shows how to use the Azure portal to create an Azure Key Vault.

1. From the Azure portal menu, or from the **Home** page, select **Create a resource**.
2. In the Search box, enter **Key Vault**.
3. From the results list, choose **Key Vault**.

4. On the Key Vault section, choose **Create**.
5. On the **Create key vault** section provide the following information:

- **Subscription:** Choose a subscription.
- Under **Resource group**, choose **Create new** and enter a resource group name. Take note of the key vault name. *You'll need it later when deploying WLS.*
- **Key Vault Name:** A unique name is required. Take note of the key vault name. *You'll need it later when deploying WLS.*

 **Note**

You may use the same name for both **Resource group** and **Key vault name**.

- In the **Location** pull-down menu, choose a location.
  - Leave the other options to their defaults.
6. Select **Next: Access Policy**.
  7. Under **Enable Access to**, select **Azure Resource Manager for template deployment**.
  8. Select **Review + Create**.
  9. Select **Create**.

Key vault creation is fairly lightweight, typically completing in less than two minutes. When deployment completes, select **Go to resource** and continue to the next section.

## Create a TLS/SSL certificate

This section shows how to create a self-signed TLS/SSL certificate in a format suitable for use by Application Gateway deployed with WebLogic Server on Azure. The certificate must have a non-empty password. If you already have a valid, non-empty password TLS/SSL certificate in **.pfx** format, you can skip this section and move on to the next. If your existing, valid, non-empty password TLS/SSL certificate is not in the **.pfx** format, first convert it to a **.pfx** file before skipping to the next section. Otherwise, open a command shell and enter the following commands.

 **Note**

This section shows how to base 64 encode the certificate before storing it as a secret in the Key Vault. This is required by the underlying Azure deployment that creates the WebLogic Server and Application Gateway.

Follow these steps to create and base 64 encode the certificate:

1. Create an RSA PRIVATE KEY

```
Bash
```

```
openssl genrsa 2048 > private.pem
```

2. Create a corresponding public key.

```
Bash
```

```
openssl req -x509 -new -key private.pem -out public.pem
```

You'll have to answer several questions when prompted by the OpenSSL tool. These values will be included in the certificate. This tutorial uses a self-signed certificate, therefore the values are irrelevant. The following literal values are fine.

- a. For **Country Name**, enter a two letter code.
- b. For **State or Province Name**, enter WA.
- c. For **Organization Name**, enter Contoso. For **Organizational Unit Name** enter billing.
- d. For **Common Name**, enter Contoso.
- e. For **Email Address**, enter billing@contoso.com.

3. Export the certificate as a .pfx file

```
Bash
```

```
openssl pkcs12 -export -in public.pem -inkey private.pem -out mycert.pfx
```

Enter the password twice. Take note of the password. *You'll need it later when deploying WLS.*

4. Base 64 encode the *mycert.pfx* file

```
Bash
```

```
base64 mycert.pfx > mycert.txt
```

Now that you have a Key Vault and a valid TLS/SSL certificate with a non-empty password, you can store the certificate in the Key Vault.

## Store the TLS/SSL certificate in the Key Vault

This section shows how to store the certificate and its password in the Key Vault created in the preceding sections.

To store the certificate, follow these steps:

1. From the Azure portal, put the cursor in the search bar at the top of the page and type the name of the Key Vault you created earlier in the tutorial.
2. Your Key Vault should appear under the **Resources** heading. Select it.
3. In the **Settings** section, select **Secrets**.
4. Select **Generate/Import**.
5. Under **Upload options**, leave the default value.
6. Under **Name**, enter `myCertSecretData`, or whatever name you like.
7. Under **Value**, enter the content of the *mycert.txt* file. The length of the value, and the presence of newlines, aren't a problem for the text field.
8. Leave the remaining values at their defaults and select **Create**.

To store the password for the certificate, follow these steps:

1. You'll be returned to the **Secrets** page. Select **Generate/Import**.
2. Under **Upload options**, leave the default value.
3. Under **Name**, enter `myCertSecretPassword`, or whatever name you like.
4. Under **Value**, enter the password for the certificate.
5. Leave the remaining values at their defaults and select **Create**.
6. You'll be returned to the **Secrets** page.

## Identify the Key Vault

Now that you have a Key Vault with a signed TLS/SSL certificate and its password stored as secrets, return to the **Azure Application Gateway** section to identify the Key Vault for the deployment.

Select desired SSL certificate option \* ⓘ

Upload an SSL certificate  
 Identify an Azure Key Vault  
 Generate a self-signed certificate

Resource group name in current subscription containing the KeyVault \* ⓘ contoso-rg ✓

Name of the Azure KeyVault containing secrets for the certificate for SSL Termination \* ⓘ contoso-kv ✓

The name of the secret in the specified KeyVault whose value is the SSL certificate data \* ⓘ myCertSecretData ✓

The name of the secret in the specified KeyVault whose value is the password for the SSL certificate \* ⓘ myCertSecretPassword

1. Under **Resource group name in current subscription containing the KeyVault**, enter the name of the resource group containing the Key Vault you created earlier.
2. Under **Name of the Azure KeyVault containing secrets for the Certificate for SSL Termination**, enter the name of the Key Vault.
3. Under **The name of the secret in the specified KeyVault whose value is the SSL Certificate Data**, enter `myCertSecretData`, or whatever name you entered previously.
4. Under **The name of the secret in the specified KeyVault whose value is the password for the SSL Certificate**, enter `myCertSecretData`, or whatever name you entered previously.
5. Select **Review + Create**.
6. Select **Create**. This will do a validation the certificate can be obtained from the Key Vault, and that its password matches the value you stored in for the password in the Key Vault. If this validation step fails, review the properties of the Key Vault, ensure the certificate was entered correctly, and ensure the password was entered correctly.
7. Once you see **Validation passed**, select **Create**.

This will start the process of creating the WLS cluster and its front-end Application Gateway, which may take about 15 minutes. When the deployment completes, select **Go to resource group**. From the list of resources in the resource group, select **myAppGateway**.

The final option for providing a TLS/SSL certificate to the App Gateway is detailed in the next section. If you're satisfied with your chosen option, you can skip to the section [Continue with deployment](#).

## Option three: Generate a self-signed certificate

This option is suitable for test and development deployments only. With this option, both an Azure Key Vault and a self-signed certificate are automatically created, and the certificate is provided to App Gateway.

To request the deployment to perform these actions, use the following steps:

1. In the **Azure Application Gateway** section, select **Generate a self-signed certificate**.
2. Select a user-assigned managed identity. This is necessary to allow the deployment to create the Azure Key Vault and certificate.
3. If you don't already have a user-assigned managed identity, select **Add** to begin the process of creating one.
4. To create a user-assigned managed identity, follow the steps in the [Create a user-assigned managed identity](#) section of [Create, list, delete, or assign a role to a user-assigned managed identity using the Azure portal](#). Once you've selected the user-assigned managed identity, make sure the checkbox next to the user-assigned managed identity is checked.

The screenshot shows the 'SSL certificate' configuration page for an Azure Application Gateway. Under 'Select your current usage of SSL', the 'Generate a self-signed certificate' option is selected and highlighted with a red box. Below this, there's a section for 'User assigned managed identity' with an 'Add' button (also highlighted with a red box) and a table listing a single identity named 'myID'. The 'myID' row is also highlighted with a red box.

Name	resource group	subscription
myID	contoso-identity	12345-67890-098787-4321

## Continue with deployment

You can now continue with the other aspects of the WLS deployment as described in the [Oracle documentation ↗](#).

# Validate successful deployment of WLS and App Gateway

This section shows a technique to quickly validate the successful deployment of the WLS cluster and Application Gateway.

If you had selected **Go to resource group** and then **myAppGateway** at the end of the preceding section, you'll be looking at overview page for the Application Gateway. If not, you can find this page by typing **myAppGateway** in the text box at the top of the Azure portal, and then selecting the correct one that appears. Be sure to select the one within the resource group you created for the WLS cluster. Then, complete the following steps.

1. In the left pane of the overview page for **myAppGateway**, scroll down to the **Monitoring** section and select **Backend health**.
2. After the **loading** message disappears, you should see a table in the middle of the screen showing the nodes of your cluster configured as nodes in the backend pool.
3. Verify that the status shows **Healthy** for each node.

## Clean up resources

If you're not going to continue to use the WLS cluster, delete the Key Vault and the WLS Cluster with the following steps:

1. Visit the overview page for **myAppGateway** as shown in the preceding section.
2. At the top of the page, under the text **Resource group**, select the resource group.
3. Select **Delete resource group**.
4. The input focus will be set to the field labeled **TYPE THE RESOURCE GROUP NAME**. Type the resource group name as requested.
5. This will cause the **Delete** button to become enabled. Select the **Delete** button.  
This operation will take some time, but you can continue to the next step while the deletion is processing.
6. Locate the Key Vault by following the first step of the section [Store the TLS/SSL certificate in the Key Vault](#).
7. Select **Delete**.
8. Select **Delete** in the pane that appears.

## Next steps

Continue to explore options to run WLS on Azure.

[Learn more about Oracle WebLogic Server on Azure](#)

# What are solutions for running Oracle WebLogic Server on Azure Virtual Machines?

Article • 10/25/2023

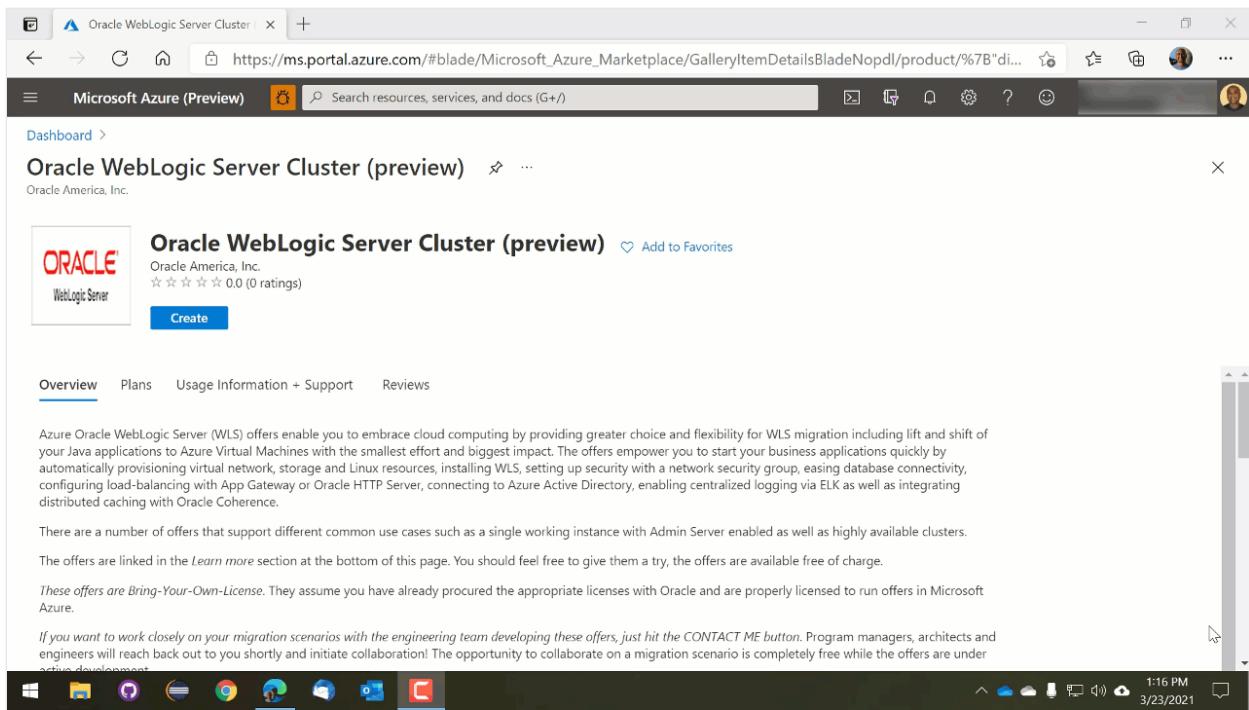
Applies to:  Linux VMs

This page describes the solutions for running Oracle WebLogic Server (WLS) on Azure virtual machines. These solutions are jointly developed and supported by Oracle and Microsoft.

You can also run WLS on the Azure Kubernetes Service. The solutions to do so are described in [this Microsoft article](#).

WLS is a leading Java application server running some of the most mission-critical enterprise Java applications across the globe. WLS forms the middleware foundation for the Oracle software suite. Oracle and Microsoft are committed to empowering WLS customers with choice and flexibility to run workloads on Azure as a leading cloud platform.

The Azure WLS solutions are aimed at making it as easy as possible to migrate your Java applications to Azure virtual machines. The solutions do so by generating deployed resources for most common cloud provisioning scenarios. The solutions automatically provision virtual network, storage, Java, WLS, and Linux resources. With minimal effort, WebLogic Server is provisioned. The solutions can set up security with a network security group, load balancing with Azure App Gateway or Oracle HTTP Server, and distributed caching with Oracle Coherence. You can also automatically connect to your existing database including Azure PostgreSQL, Azure MySQL, Azure SQL, and the Oracle Database on the Oracle Cloud or Azure.



There are solution templates available to meet different scenarios such as [single instance with an admin server](#), and [cluster](#). The solutions are available free of charge. These solutions are described and linked below. You can find detailed documentation on the solutions [here](#).

*These offers are Bring-Your-Own-License.* They assume you have already got the appropriate licenses with Oracle and are properly licensed to run offers in Azure.

The solution templates support a range of operating system, Java, and WLS versions through base images (such as WebLogic Server 14 and Java 11 on Red Hat Enterprise Linux 8). These base images are also available on Azure Marketplace on their own. The base images are suitable for customers that require complex, customized Azure deployments.

If you prefer step-by-step guidance for going from zero to a WLS cluster without any solution templates or base images, see [Install Oracle WebLogic Server on Azure Virtual Machines manually](#).

*If you're interested in working closely on your migration scenarios with the engineering team developing these offers, select the [CONTACT ME](#) button on the [marketplace offer overview page](#).* Program managers, architects, and engineers will reach back out to you shortly and start close collaboration.

## Oracle WebLogic Server with Admin Server

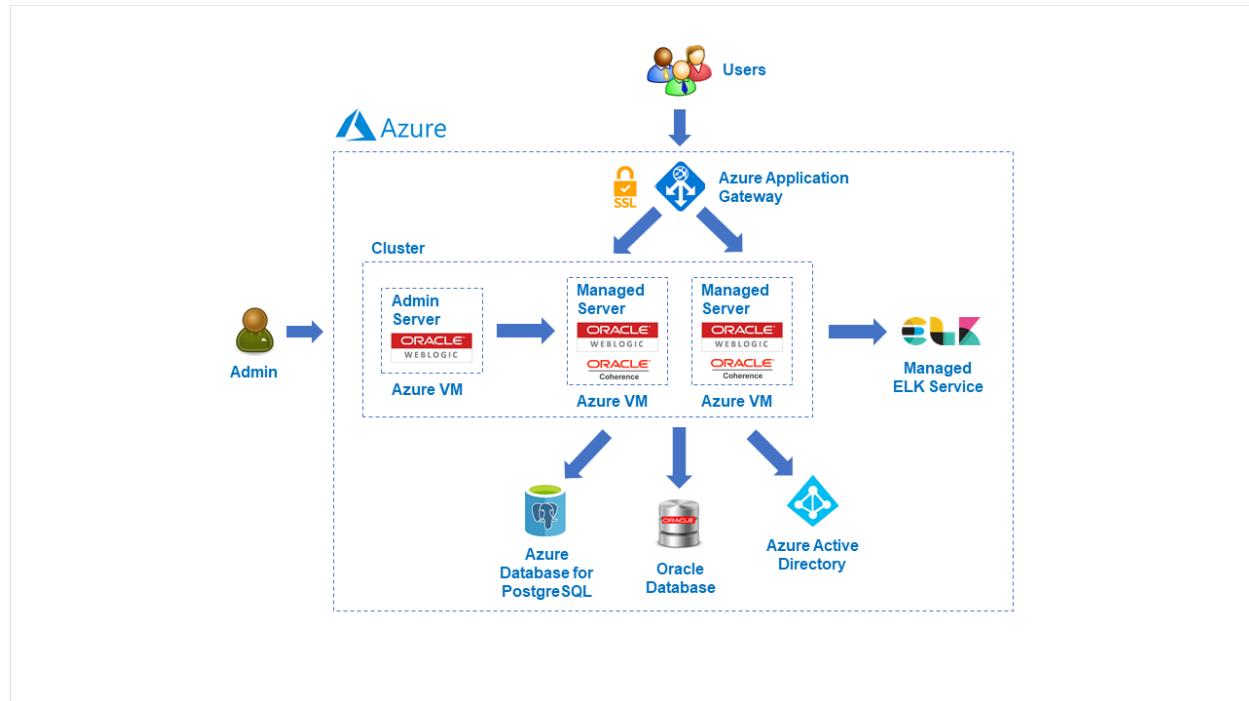
This [solution template](#) provisions a single virtual machine and installs WLS on it. It creates a domain and starts up the administration server. You can manage the domain

and get started with application deployments right away.

## Oracle WebLogic Server Cluster

This [solution template](#) creates a highly available cluster of WLS virtual machines. The administration server and all managed servers are started by default. You can manage the cluster and get started with highly available applications right away.

The solutions enable a wide range of production-ready deployment architectures with relative ease. You can meet most migration cases in the most productive way possible by allowing a focus on business application development.



After resources are automatically provisioned by the solutions, you have complete flexibility to customize your deployments further. It's likely on top of deploying applications you'll integrate further Azure resources with your deployments. You're encouraged to [connect with the development team](#) and provide feedback on further improving the solutions.

## Next steps

Explore the offers on Azure.

[Oracle WebLogic Server with Admin Server](#)

[Oracle WebLogic Server Cluster](#)

# Migrate JBoss EAP applications to JBoss EAP on Azure App Service

Article • 01/16/2023

This guide describes what you should be aware of when you want to migrate an existing JBoss EAP application to run on JBoss EAP in an Azure App Service instance.

## Pre-migration

To ensure a successful migration, before you start, complete the assessment and inventory steps described in the following sections.

### Inventory server capacity

Document the hardware (memory, CPU, disk) of the current production server(s) and the average and peak request counts and resource utilization. You'll need this information regardless of the migration path you choose. It's useful, for example, to help guide selection of the size of the VMs in your node pool, the amount of memory to be used by the container, and how many CPU shares the container needs.

It's possible to resize node pools in AKS. To learn how, see [Resize node pools in Azure Kubernetes Service \(AKS\)](#).

### Inventory all secrets

Check all properties and configuration files on the production server(s) for any secrets and passwords. Be sure to check *jboss-web.xml* in your WARs. Configuration files that contain passwords or credentials may also be found inside your application.

Consider storing those secrets in Azure KeyVault. For more information, see [Azure Key Vault basic concepts](#).

You can use Key Vault secrets in your App Service instance with Key Vault references. Key Vault references allow you to use the secrets in your application while keeping them secured and encrypted at rest. For more information, see [Use Key Vault references for App Service and Azure Functions](#).

### Inventory all certificates

Document all the certificates used for public SSL endpoints. You can view all certificates on the production server(s) by running the following command:

```
Bash
```

```
keytool -list -v -keystore <path to keystore>
```

## Validate that the supported Java version works correctly

JBoss EAP on Azure VMs requires a supported version of Java. For guidance on which version of the JDK to use, see [Supported Configurations](#) in the Red Hat documentation.

### Note

This validation is especially important if your current server is running on an unsupported JDK (such as Oracle JDK or IBM OpenJ9).

To obtain your current Java version, sign in to your production server and run the following command:

```
Bash
```

```
java -version
```

## Inventory external resources

External resources, such as data sources, JMS message brokers, and others are injected via Java Naming and Directory Interface (JNDI). Some such resources may require migration or reconfiguration.

## Inside your application

Inspect the *WEB-INF/jboss-web.xml* and/or *WEB-INF/web.xml* files. Look for `<Resource>` elements inside the `<Context>` element.

## Datasources

Datasources are JNDI resources with the `type` attribute set to `javax.sql.DataSource`. For each datasource, document the following information:

- What is the datasource name?
- What is the connection pool configuration?
- Where can I find the JDBC driver JAR file?

For more information, see [About JBoss EAP Datasources](#) in the JBoss EAP documentation.

## All other external resources

It isn't feasible to document every possible external dependency in this guide. It's your team's responsibility to verify that you can satisfy every external dependency of your application after the migration.

## Determine whether session replication is used

If your application relies on session replication, you'll have to change your application to remove this dependency. App Service does not allow instances to communicate directly with one another.

## Determine whether and how the file system is used

Any usage of the file system on the application server will require reconfiguration or, in rare cases, architectural changes. The file system may be used by JBoss EAP modules or by your application code. You may identify some or all of the scenarios described in the following sections.

### Read-only static content

If your application currently serves static content, you'll need an alternate location for it. You may wish to consider moving static content to Azure Blob Storage and adding Azure CDN for lightning-fast downloads globally. For more information, see [Static website hosting in Azure Storage](#) and [Quickstart: Integrate an Azure storage account with Azure CDN](#). You can also directly deploy the static content to an app in the Azure Spring Apps Enterprise plan. For more information, see [Deploy web static files](#).

### Dynamically published static content

If your application allows for static content that is uploaded/produced by your application but is immutable after its creation, you can use Azure Blob Storage and Azure CDN as described above, with an Azure Function to handle uploads and CDN

refresh. We've provided a sample implementation for your use at [Uploading and CDN-preloading static content with Azure Functions](#). You can also directly deploy the static content to an app in the Azure Spring Apps Enterprise plan. For more information, see [Deploy web static files](#).

## Dynamic or internal content

For files that are frequently written and read by your application (such as temporary data files), or static files that are visible only to your application, you can use local file storage associated with your app service plan. For more information, see [Operating system functionality on Azure App Service](#) and [Understanding the Azure App Service file system](#).

## Determine whether your application relies on scheduled jobs

Scheduled jobs, such as Quartz Scheduler tasks or Unix cron jobs, should NOT be used with Azure App Service. Azure App Service will not prevent you from deploying an application containing scheduled tasks internally. However, if your application is scaled out, the same scheduled job may run more than once per scheduled period. This situation can lead to unintended consequences.

Inventory any scheduled tasks running on the production server(s), inside or outside your application code.

## Determine whether a connection to on-premises is needed

If your application needs to access any of your on-premises services, you'll need to provision one of Azure's connectivity services. For more information, see [Choose a solution for connecting an on-premises network to Azure](#). Alternatively, you'll need to refactor your application to use publicly available APIs that your on-premises resources expose.

## Determine whether Java Message Service (JMS) Queues or Topics are in use

If your application is using JMS Queues or Topics, you'll need to migrate them to an externally hosted JMS server. Azure Service Bus and the Advanced Message Queuing

Protocol (AMQP) can be a great migration strategy for those using JMS. For more information, see [Use JMS with Azure Service Bus and AMQP 1.0](#).

If JMS persistent stores have been configured, you must capture their configuration and apply it after the migration.

## Determine whether JCA connectors are in use

If your application uses JCA connectors, validate that you can use the JCA connector on JBoss EAP. If you can use the JCA connector on JBoss EAP, then for it to be available, you must add the JARs to the server classpath and put the necessary configuration files in the correct location in the JBoss EAP server directories.

## Determine whether JAAS is in use

If your application is using JAAS, you'll need to capture how JAAS is configured. If it's using a database, you can convert it to a JAAS domain on JBoss EAP. If it's a custom implementation, you'll need to validate that it can be used on JBoss EAP.

## Determine whether your application uses a Resource Adapter

If your application needs a Resource Adapter (RA), it needs to be compatible with JBoss EAP. Determine whether the RA works fine on a standalone instance of JBoss EAP by deploying it to the server and properly configuring it. If the RA works properly, you'll need to add the JARs to the server classpath of the App Service and put the necessary configuration files in the correct location in the JBoss EAP server directories for it to be available.

## Determine whether your application is composed of multiple WARs

If your application is composed of multiple WARs, you should treat each of those WARs as separate applications and go through this guide for each of them.

## Determine whether your application is packaged as an EAR

If your application is packaged as an EAR file, be sure to examine the *application.xml* file and capture the configuration.

## Note

If you want to be able to scale each of your web applications independently for better use of your App Service resources you should break up the EAR into separate web applications.

## Identify all outside processes and daemons running on the production servers

If you have any processes running outside the application server, such as monitoring daemons, you'll need to eliminate them or migrate them elsewhere.

## Perform in-place testing

Before creating your container images, migrate your application to the JDK and JBoss EAP versions that you intend to use on App Service. Test the application thoroughly to ensure compatibility and performance.

## JBoss EAP on App Service feature notes

When using JBoss EAP on App Service, be sure to take the following notes into consideration.

- **JBoss EAP management console:** The JBoss web console isn't exposed on App Service. Instead, the Azure portal provides the management APIs for your application, and you should deploy using the Azure CLI, Azure Maven Plugin, or other Azure developer tools.
- **Transactions:** The application instances are run in a stateless manner, so the Transactions API isn't currently supported. For more information, see [Managing transactions on JBoss EAP](#) in the Red Hat documentation.
- **Managed domain mode:** In a multi-server production environment, Managed Domain mode in JBoss EAP offers centralized managed capabilities. However with JBoss EAP on App Service, the App Service platform assumes the responsibility for configuration and management of your server instances. App Service eliminates the need for JBoss EAP's managed domain mode. Domain mode is a good choice for virtual machine-based multi-server deployments. For more information, see [About managed domains](#) in the Red Hat documentation.

- **Server-to-server clustering:** As of March 15th, 2022, clustered deployment of JBoss EAP is supported in Public Preview. This support means that you no longer have to remove the following features from your applications before you can deploy them to App Service:
  - Stateful session beans.
  - Distributed transactions.
  - Similar features that require instance-to-instance communication or high availability.

For more information, see the [release announcement](#) and the [Clustering in JBoss EAP](#) section of [Configure a Java app for Azure App Service](#).

## Migration

### Red Hat Migration Toolkit for Apps

The [Red Hat Migration Toolkit for Applications](#) is a free extension for Visual Studio Code. This extension analyzes your application code and configuration to provide recommendations for migrating to the cloud from on-premises. For more information, see [Migration Toolkit for Applications overview](#).

The contents of this guide will help you address the other components of the migration journey, such as choosing the correct App Service Plan type, externalizing your session state, and using Azure to manage your EAP instances instead of the JBoss Management interface.

### Provision Azure App Service for JBoss EAP runtime

Use the following commands to create a resource group and an Azure App Service Plan. After the App Service Plan is created, a Linux web app plan is created using the JBoss EAP runtime. You can create JBoss EAP sites only on PremiumV3 and IsolatedV2 App Service Plan tiers.

Be sure the specified environment variables have appropriate values.

#### Note

PremiumV3 and IsolatedV2 are both eligible for Reserved Instance pricing, which can reduce your costs. For more information on App Service Plan tiers and Reserved Instance pricing, see [App Service pricing](#).

## Azure CLI

```
az group create --resource-group $resourceGroup --location eastus
az acr create --resource-group $resourceGroup --name $acrName --sku Standard
az appservice plan create \
    --resource-group $resourceGroup \
    --name $jbossAppService \
    --is-linux \
    --sku P1V2
az webapp create \
    --resource-group $resourceGroup \
    --name $jbossWebApp \
    --plan $jbossAppServicePlan \
    --runtime "JBOSSEAP|7-java8"
# Or use "JBOSSEAP|7-java11" if you're using Java 11
```

## Build the application

Build the application using the following Maven command.

### Bash

```
mvn clean install -DskipTests
```

## Deploy the application

If your application is built from a Maven POM file, use the Webapp plugin for Maven to create the Web App and deploy your application. For more information, see [Quickstart: Create a Java app on Azure App Service](#).

To automate the deployment of JBoss EAP applications, you can use [Azure Pipelines task for Web App](#) or [GitHub Action for deploying to Azure WebApp](#).

## Set up data sources

There are three core steps when registering a data source with JBoss EAP: uploading the JDBC driver, adding the JDBC driver as a module, and registering the module. For more information, see [Datasource Management](#) in the JBoss EAP documentation. App Service is a stateless hosting service, so the configuration commands for adding and registering the data source module must be scripted and applied as the container starts.

To set up data sources, use the following steps.

1. Obtain your database's JDBC driver.

2. Create an XML module definition file for the JDBC driver. The example shown below is a module definition for PostgreSQL. Be sure to replace the `resource-root path` value with the path to the JDBC driver you use.

XML

```
<?xml version="1.0" ?>
<module xmlns="urn:jboss:module:1.1" name="org.postgres">
    <resources>
        <!-- ***** IMPORTANT: REPLACE THIS PLACEHOLDER *****-->
        <resource-root path="/home/site/deployments/tools/postgresql-
42.2.12.jar" />
    </resources>
    <dependencies>
        <module name="javax.api"/>
        <module name="javax.transaction.api"/>
    </dependencies>
</module>
```

3. Put your JBoss CLI commands into a file named `jboss-cli-commands.cli`. The JBoss commands must add the module and register it as a data source. The example below shows the JBoss CLI commands for PostgreSQL.

Bash

```
module add --name=org.postgres --
resources=/home/site/deployments/tools/postgresql-42.2.12.jar --module-
xml=/home/site/deployments/tools/postgres-module.xml

/subsystem=datasources/jdbc-driver=postgres:add(driver-
name="postgres",driver-module-name="org.postgres",driver-class-
name=org.postgresql.Driver,driver-xa-datasource-class-
name=org.postgresql.xa.PGXADatasource)

data-source add --name=postgresDS --driver-name=postgres --jndi-
name=java:jboss/datasources/postgresDS --connection-
url=${POSTGRES_CONNECTION_URL},env.POSTGRES_CONNECTION_URL:jdbc:postgres
ql://db:5432/postgres} --user-
name=${POSTGRES_SERVER_ADMIN_FULL_NAME},env.POSTGRES_SERVER_ADMIN_FULL_N
AME:postgres} --
password=${POSTGRES_SERVER_ADMIN_PASSWORD},env.POSTGRES_SERVER_ADMIN_PAS
SWORD:example} --use-ccm=true --max-pool-size=5 --blocking-timeout-
wait-millis=5000 --enabled=true --driver-class=org.postgresql.Driver --
exception-sorter-class-
name=org.jboss.jca.adapters.jdbc.extensions.postgres.PostgreSQLException-
Sorter --jta=true --use-java-context=true --valid-connection-checker-
class-
name=org.jboss.jca.adapters.jdbc.extensions.postgres.PostgreSQLValidCon
nectionChecker
```

4. Create a startup script called *startup\_script.sh* that calls the JBoss CLI commands.

The example below shows how to call your *jboss-cli-commands.cli* file. Later you will configure App Service to run this script when the instance starts.

Bash

```
$JBOSS_HOME/bin/jboss-cli.sh --connect --  
file=/home/site/deployments/tools/jboss-cli-commands.cli
```

5. Using an FTP client of your choice, upload your JDBC driver, *jboss-cli-commands.cli*, *startup\_script.sh*, and the module definition to */site/deployments/tools/*.

6. Configure your site to run *startup\_script.sh* when the container starts. In the Azure portal, navigate to **Configuration > General Settings > Startup Command**. Set the startup command field to */home/site/deployments/tools/startup\_script.sh*, then select **Save**.

7. Restart the web app, which will cause it to run the configuration script.

8. Update the JTA datasource configuration for your application. Open the *src/main/resources/META-INF/persistence.xml* file for your app and find the `<jta-data-source>` element. Replace its contents as shown here:

Bash

```
<jta-data-source>java:jboss/datasources/postgresDS</jta-data-source>
```

## Build the application

Build the application using the following Maven command.

Bash

```
mvn clean install -DskipTests
```

## Deploy the application

If your application is built from a Maven POM file, use the Webapp plugin for Maven to create the Web App and deploy your application. For more information, see [Quickstart: Create a Java app on Azure App Service](#).

To automate the deployment of JBoss EAP applications, you can use [Azure Pipelines task for Web App](#) or [GitHub Action for deploying to Azure WebApp ↗](#).

## Post-migration

Now that you've migrated your application to Azure App Service, you should verify that it works as you expect. After you've done that, we have some recommendations for you that can make your application more cloud-native.

## Recommendations

- If you opted to use the `/home` directory for file storage, consider replacing it with Azure Storage. For more information, see [Access Azure Storage as a network share from a container in App Service](#).
- If you have configuration in the `/home` directory that contains connection strings, SSL keys, and other secret information, consider using Azure Key Vault and/or parameter injection with application settings where possible. For more information, see [Use Key Vault references for App Service and Azure Functions](#) and [Configure an App Service app in the Azure portal](#).
- Consider using deployment slots for reliable deployments with zero downtime. For more information, see [Set up staging environments in Azure App Service](#).
- Design and implement a DevOps strategy. To maintain reliability while increasing your development velocity, consider automating deployments and testing with Azure Pipelines. For more information, see [Build and deploy to a Java web app](#). When you use deployment slots, you can [automate deployment to a slot](#) followed by the slot swap. For more information, see the [Deploy to a slot](#) section of [Deploy an Azure Web App \(Linux\)](#).
- Design and implement a business continuity and disaster recovery strategy. For mission-critical applications, consider a multi-region deployment architecture. For more information, see [Highly available multi-region web application](#).

# Architectures for Oracle applications with Azure Virtual Machines with database on OCI

Article • 08/18/2023

Applies to:  Linux VMs

Microsoft and Oracle have worked together to enable customers to deploy Oracle applications such as Oracle E-Business Suite, JD Edwards EnterpriseOne, and PeopleSoft in the cloud. With the introduction of the preview [private network interconnectivity](#) between Microsoft Azure and Oracle Cloud Infrastructure (OCI), Oracle applications can now be deployed on Azure with their back-end databases in Azure or OCI. Oracle applications can also be integrated with Azure Active Directory, allowing you to set up single sign-on so that users can sign into the Oracle application using their Azure Active Directory (Azure AD) credentials.

OCI offers multiple Oracle database options for Oracle applications, including DBaaS, Exadata Cloud Service, Oracle RAC, and Infrastructure-as-a-Service (IaaS). Currently, Autonomous Database isn't a supported back-end for Oracle applications.

There are [multiple options](#) for deploying Oracle applications in Azure, including in a highly available and secure manner. Azure also offers [Oracle database VM images](#) that you can deploy if you choose to run your Oracle applications entirely on Azure.

The following sections outline architecture recommendations by both Microsoft and Oracle to deploy Oracle E-Business Suite, JD Edwards EnterpriseOne, and PeopleSoft in a cross-cloud configuration or entirely in Azure. Microsoft and Oracle have tested these applications and confirmed that the performance meets standards set by Oracle for these applications.

## Architecture considerations

Oracle applications are made up of multiple services, which can be hosted on the same or multiple virtual machines in Azure and optionally in OCI.

Application instances can be set up with private or public endpoints. Microsoft and Oracle recommend setting up a *bastion host VM* with a public IP address in a separate subnet for management of the application. Then, assign only private IP addresses to the other machines, including the database tier.

When setting up an application in a cross-cloud architecture, planning is required to ensure that the IP address space in the Azure virtual network doesn't overlap the private IP address space in the OCI virtual cloud network.

For added security, set up network security groups at a subnet level to ensure only traffic on specific ports and IP addresses is permitted. For example, machines in the middle tier should only receive traffic from within the virtual network. No external traffic should reach the middle tier machines directly.

For high availability, you can set up redundant instances of the different servers in the same availability set or different availability zones. Availability zones allow you to achieve a 99.99% uptime SLA, while availability sets allow you to achieve a 99.95% uptime SLA in-region. Sample architectures shown in this article are deployed across two availability zones.

When deploying an application using the cross-cloud interconnect, you may continue to use an existing ExpressRoute circuit to connect your Azure environment to your on-premises network. However, you need a separate ExpressRoute circuit for the interconnect to OCI than the one connecting to your on-premises network.

## E-Business Suite

Oracle E-Business Suite (EBS) is a suite of applications including Supply Chain Management (SCM) and Customer Relationship Management (CRM). To take advantage of OCI's managed database portfolio, EBS can be deployed using the cross-cloud interconnect between Microsoft Azure and OCI. In this configuration, the presentation and application tiers run in Azure and the database tier in OCI, as illustrated in the following architecture diagram (Figure 1).

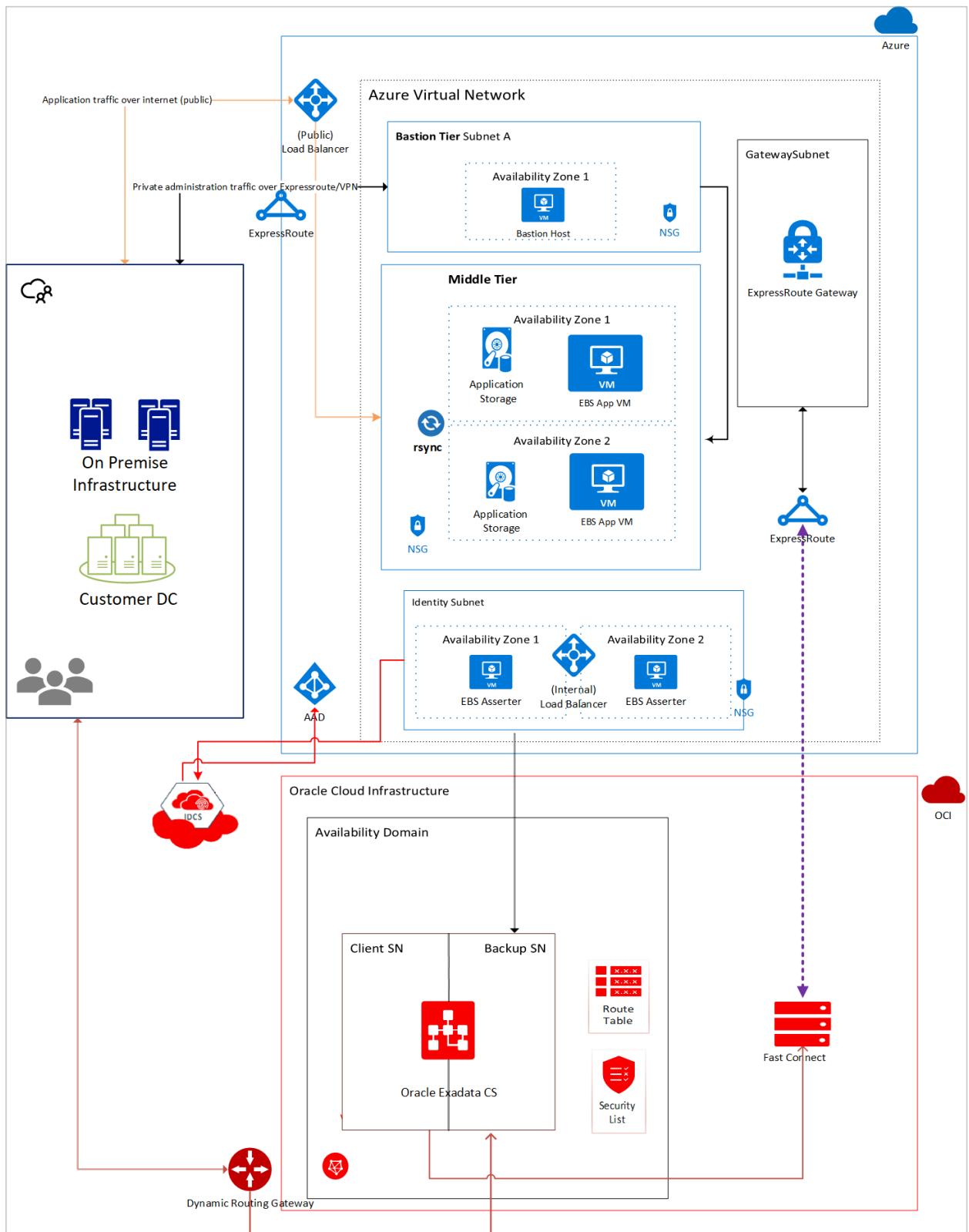


Figure 1: E-Business Suite cross-cloud architecture

In this architecture, the virtual network in Azure is connected to a virtual cloud network in OCI using the cross-cloud interconnect. The application tier is set up in Azure, whereas the database is set up in OCI. It's recommended to deploy each component to its own subnet with network security groups to allow traffic only from specific subnets on specific ports.

The architecture can also be adapted for deployment entirely on Azure with highly available Oracle databases configured using Oracle Data Guard in two availability zones in a region. The following diagram (Figure 2) is an example of this architectural pattern:

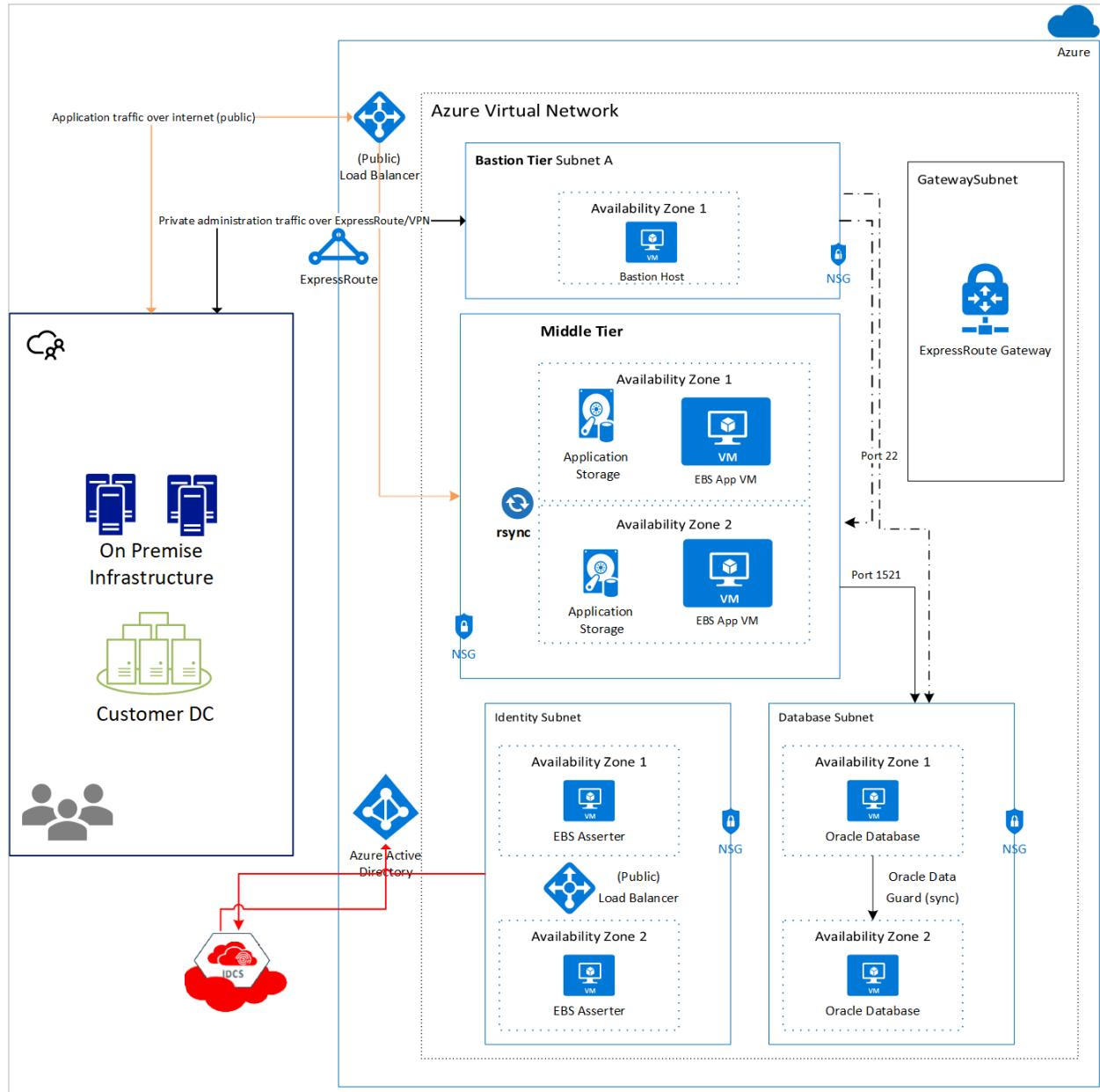


Figure 2: E-Business Suite Azure-only architecture

The following sections describe the different components at a high level.

## Bastion tier

The bastion host is an optional component that you can use as a jump server to access the application and database instances. The bastion host VM can have a public IP address assigned to it, although the recommendation is to set up an ExpressRoute connection or site-to-site VPN with your on-premises network for secure access. Additionally, only SSH (port 22, Linux) or RDP (port 3389, Windows Server) should be

opened for incoming traffic. For high availability, deploy a bastion host in two availability zones or in a single availability set.

You may also enable SSH agent forwarding on your VMs, which allows you to access other VMs in the virtual network by forwarding the credentials from your bastion host. Or, use SSH tunneling to access other instances.

Here's an example of agent forwarding:

```
ssh -A -t user@BASTION_SERVER_IP ssh -A root@TARGET_SERVER_IP`
```

This command connects to the bastion and then immediately runs `ssh` again, so you get a terminal on the target instance. You may need to specify a user other than root on the target instance if your cluster is configured differently. The `-A` argument forwards the agent connection so your private key on your local machine is used automatically. Note that agent forwarding is a chain, so the second `ssh` command also includes `-A` so that any subsequent SSH connections initiated from the target instance also use your local private key.

## Application (middle) tier

The application tier is isolated in its own subnet. There are multiple virtual machines set up for fault tolerance and easy patch management. These VMs can be backed by shared storage, which is offered by Azure NetApp Files and Ultra SSDs. This configuration allows for easier deployment of patches without downtime. The machines in the application tier should be fronted by a public load balancer so that requests to the EBS application tier are processed even if one machine in the tier is offline due to a fault.

## Load balancer

An Azure load balancer allows you to distribute traffic across multiple instances of your workload to ensure high availability. In this case, a public load balancer is set up, because users are allowed to access the EBS application over the web. The load balancer distributes the load to both machines in the middle tier. For added security, allow traffic only from users accessing the system from your corporate network using a site-to-site VPN or ExpressRoute and network security groups.

## Database tier

This tier hosts the Oracle database and is separated into its own subnet. It's recommended to add network security groups that only permit traffic from the application tier to the database tier on the Oracle-specific database port 1521.

Microsoft and Oracle recommend a high availability setup. High availability in Azure can be achieved by setting up two Oracle databases in two availability zones with Oracle Data Guard, or by using Oracle Database Exadata Cloud Service in OCI. When using Oracle Database Exadata Cloud Service, your database is deployed in two subnets. You may also set up Oracle Database in VMs in OCI in two availability domains with Oracle Data Guard.

## Identity tier

The identity tier contains the EBS Asserter VM. EBS Asserter allows you to synchronize identities from Oracle Identity Cloud Service (IDCS) and Azure AD. The EBS Asserter is needed because EBS doesn't support single sign-on protocols like SAML 2.0 or OpenID Connect. The EBS Asserter consumes the OpenID connect token (generated by IDCS), validates it, and then creates a session for the user in EBS.

While this architecture shows IDCS integration, Azure AD unified access and single sign-on also can be enabled with Oracle Access Manager with Oracle Internet Directory or Oracle Unified Directory. For more information, see the whitepapers on [Deploying Oracle EBS with IDCS Integration ↗](#) or [Deploying Oracle EBS with OAM Integration ↗](#).

For high availability, it's recommended that you deploy redundant servers of the EBS Asserter across multiple availability zones with a load balancer in front of it.

Once your infrastructure is set up, E-Business Suite can be installed by following the installation guide provided by Oracle.

## JD Edwards EnterpriseOne

Oracle's JD Edwards EnterpriseOne is an integrated applications suite of comprehensive enterprise resource planning software. It's a multi-tiered application that can be set up with either an Oracle or SQL Server database backend. This section discusses details on deploying JD Edwards EnterpriseOne with an Oracle database back-end either in OCI or in Azure.

In the following recommended architecture (Figure 3), the administration, presentation, and middle tiers are deployed to the virtual network in Azure. The database is deployed in a virtual cloud network in OCI.

As with E-Business Suite, you can set up an optional bastion tier for secure administrative purposes. Use the bastion VM host as a jump server to access the application and database instances.

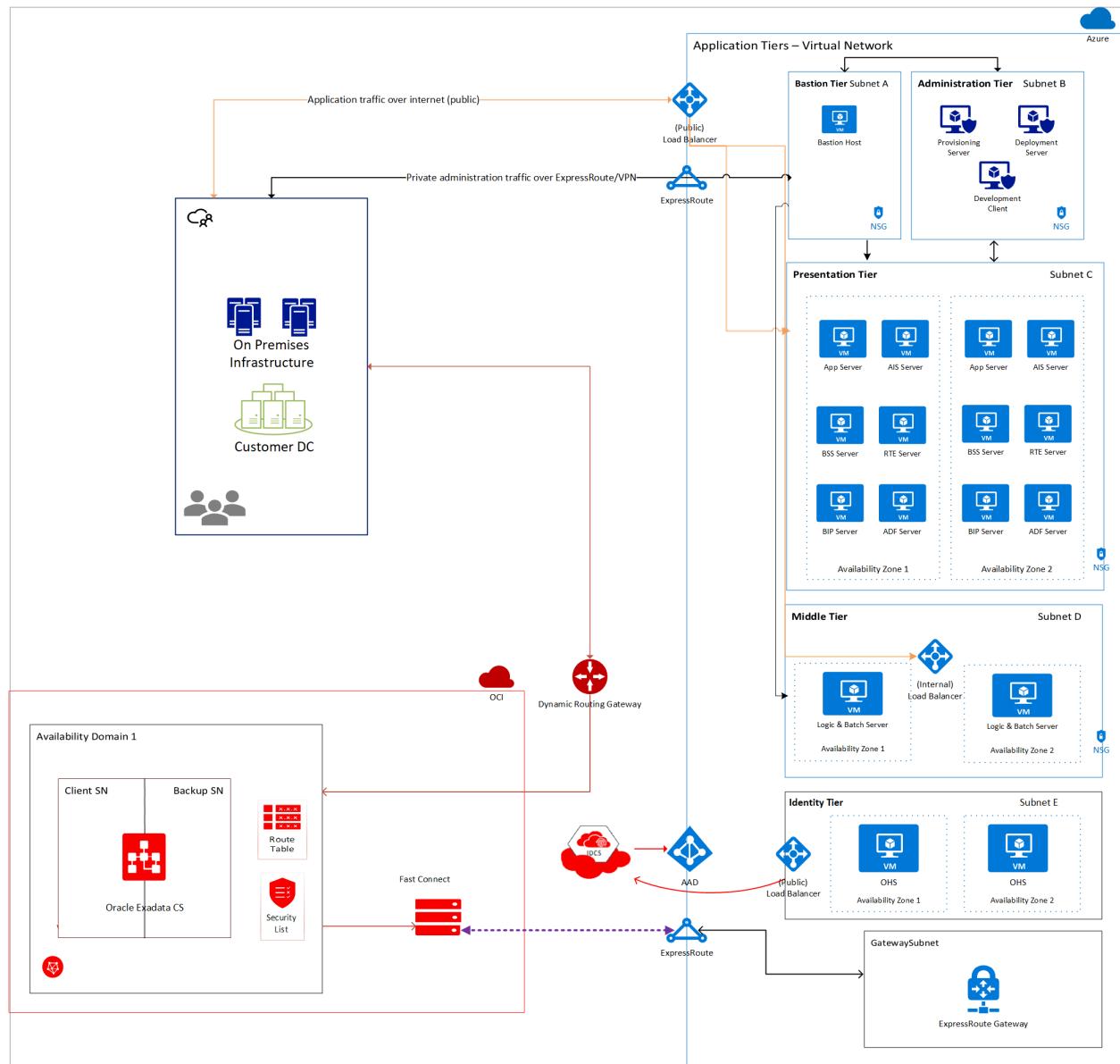
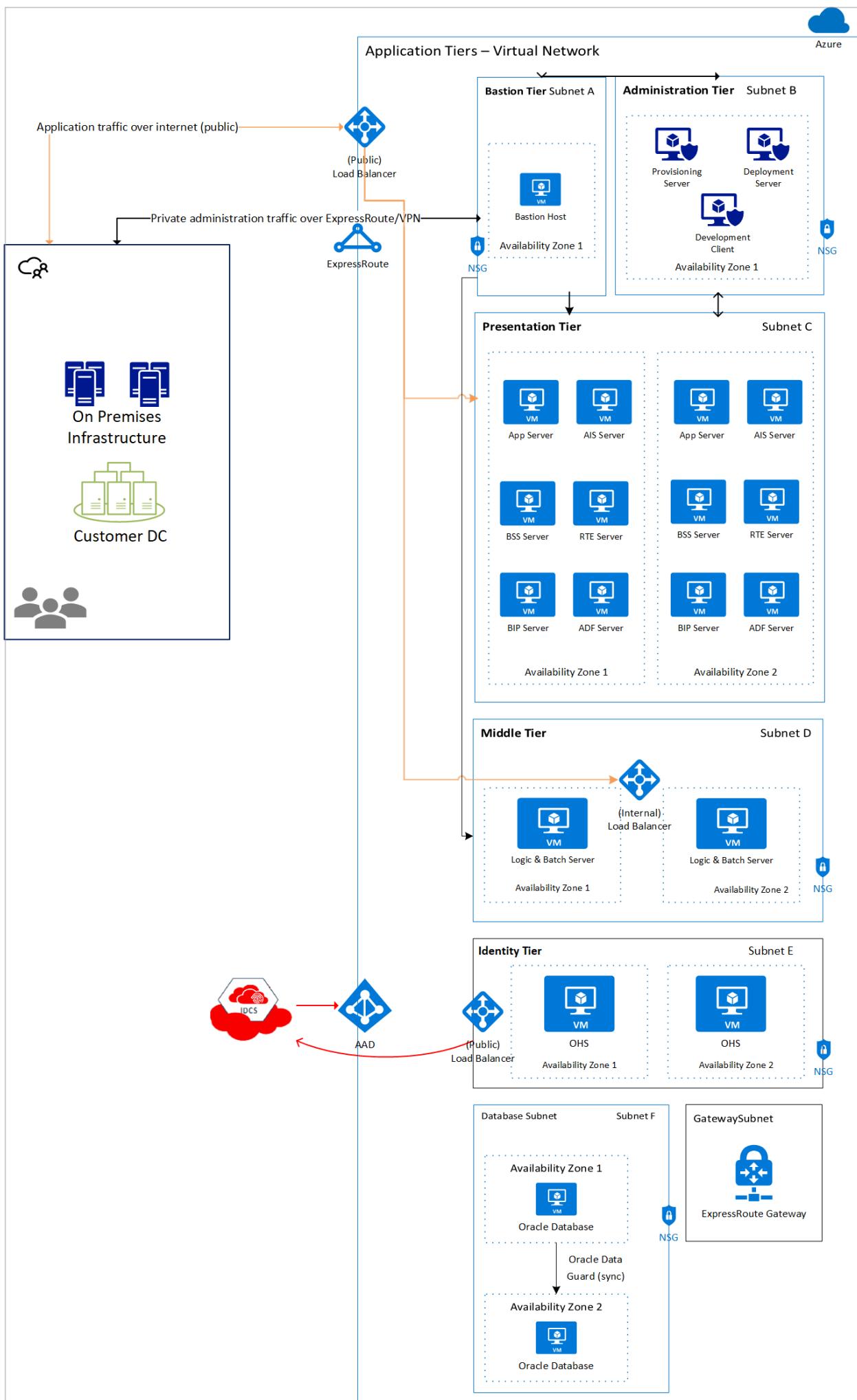


Figure 3: JD Edwards EnterpriseOne cross-cloud architecture

In this architecture, the virtual network in Azure is connected to the virtual cloud network in OCI using the cross-cloud interconnect. The application tier is set up in Azure, whereas the database is set up in OCI. It's recommended to deploy each component to its own subnet with network security groups to allow traffic only from specific subnets on specific ports.

The architecture can also be adapted for deployment entirely on Azure with highly available Oracle databases configured using Oracle Data Guard in two availability zones in a region. The following diagram (Figure 4) is an example of this architectural pattern:



*Figure 4: JD Edwards EnterpriseOne Azure-only architecture*

The following sections describe the different components at a high level.

## Bastion tier

The bastion host is an optional component that you can use as a jump server to access the application and database instances. The bastion host VM can have a public IP address assigned to it, although the recommendation is to set up an ExpressRoute connection or site-to-site VPN with your on-premises network for secure access. Additionally, only SSH (port 22, Linux) or RDP (port 3389, Windows Server) should be opened for incoming traffic. For high availability, deploy a bastion host in two availability zones or in a single availability set.

You may also enable SSH agent forwarding on your VMs, which allows you to access other VMs in the virtual network by forwarding the credentials from your bastion host. Or, use SSH tunneling to access other instances.

Here's an example of agent forwarding:

```
ssh -A -t user@BASTION_SERVER_IP ssh -A root@TARGET_SERVER_IP`
```

This command connects to the bastion and then immediately runs `ssh` again, so you get a terminal on the target instance. You may need to specify a user other than root on the target instance if your cluster is configured differently. The `-A` argument forwards the agent connection so your private key on your local machine is used automatically. Note that agent forwarding is a chain, so the second `ssh` command also includes `-A` so that any subsequent SSH connections initiated from the target instance also use your local private key.

## Administrative tier

As the name suggests, this tier is used for administrative tasks. You can carve out a separate subnet for the administrative tier. The services and servers in this tier are primarily used for installation and administration of the application. Hence, single instances of these servers are sufficient. Redundant instances aren't required for the high availability of your application.

The components of this tier are as follows:

- **Provisioning server** - This server is used for end-to-end deployment of the different components of the application. It communicates with the instances in the other tiers, including the instances in the database tier, over port 22. It hosts the Server Manager Console for JD Edwards EnterpriseOne.
- **Deployment server** - This server is primarily required for the installation of JD Edwards EnterpriseOne. During the installation process, this server acts as the central repository for required files and installation packages. The software is distributed or deployed to other servers and clients from this server.
- **Development client** - This server contains components that run in a web browser and native applications.

## Presentation tier

This tier contains various components such as Application Interface Services (AIS), Application Development Framework (ADF), and Java Application Servers (JAS). The servers in this tier communicate with the servers in the middle tier. They're fronted by a load balancer that routes traffic to the necessary server based on the port number and URL that the traffic is received on. It's recommended that you deploy multiple instances of each server type for high availability.

The following are the components in this tier:

- **Application Interface Services (AIS)** - The AIS server provides the communication interface between JD Edwards EnterpriseOne mobile enterprise applications and JD Edwards EnterpriseOne.
- **Java Application Server (JAS)** - The JAS receives requests from the load balancer and passes it to the middle tier to execute complicated tasks. The JAS has the ability to execute simple business logic.
- **BI Publisher Server (BIP)** - This server presents reports based on the data collected by the JD Edwards EnterpriseOne application. You can design and control how the report presents the data based on different templates.
- **Business Services Server (BSS)** - The BSS enables information exchange and interoperability with other Oracle applications.
- **Real-Time Events Server (RTE)** - The RTE Server allows you to set up notifications to external systems about transactions occurring in the JDE EnterpriseOne system. It uses a subscriber model and allows third-party systems to subscribe to events. To load balance requests to both RTE servers, ensure that the servers are in a cluster.
- **Application Development Framework (ADF) Server** - The ADF server is used to run JD Edwards EnterpriseOne applications developed with Oracle ADF. This is deployed on an Oracle WebLogic server with ADF runtime.

## Middle tier

The middle tier contains the logic server and batch server. In this case, both servers are installed on the same virtual machine. However, for production scenarios, it's recommended that you deploy logic server and batch server on separate servers.

Multiple servers are deployed in the middle tier across two availability zones for higher availability. An Azure load balancer should be created and these servers should be placed in its backend pool to ensure that both servers are active and processing requests.

The servers in the middle tier receive requests from the servers in the presentation tier and the public load balancer only. Network security group rules must be set up to deny traffic from any address other than the presentation tier subnet and the load balancer. An NSG rule can also be set up to allow traffic on port 22 from the bastion host for management purposes. You may be able to use the public load balancer to load balance requests between the VMs in the middle tier.

The following two components are in the middle tier:

- **Logic server** - Contain the business logic or business functions.
- **Batch server** - Used for batch processing

## Database tier

The Database tier contains the database instances for the application. The database can be either an Oracle DB, Oracle RAC, or Oracle Exadata Database system.

If the choice is to use Oracle DB, the database instance may be deployed on Azure via the Oracle DB images available on the Azure Marketplace. Alternatively, you may use the interconnect between Azure and OCI to deploy the Oracle DB in a PaaS model on OCI.

For Oracle RAC, you can use OCI in PaaS model. It is recommended that you use a two-node RAC system. While it is possible to deploy Oracle RAC on Azure CloudSimple in IaaS model, it is not a supported configuration by Oracle. Refer to [Oracle programs eligible for Authorized Cloud Environments ↗](#).

Finally, for Exadata systems, use the OCI interconnect and deploy the Exadata system in OCI. The preceding architecture diagram above shows an Exadata system deployed in OCI across two subnets.

For production scenarios, deploy multiple instances of the database across two availability zones (if deploying in Azure) or two availability domains (in OCI). Use Oracle Active Data Guard to synchronize the primary and standby databases.

The database tier only receives requests from the middle tier. It is recommended that you set up a network security group (security list if deploying the database in OCI) to only allow requests on port 1521 from the middle tier and port 22 from the bastion server for administrative reasons.

For databases deployed in OCI, a separate virtual cloud network must be set up with a dynamic routing gateway (DRG) that is connected to your FastConnect circuit.

## Identity tier

The Microsoft-Oracle partnership enables you to set up a unified identity across Azure, OCI, and your Oracle application. For JD Edwards EnterpriseOne or PeopleSoft application suite, an instance of the Oracle HTTP Server (OHS) is needed to set up single sign-on between Azure AD and Oracle IDCS.

OHS acts as a reverse proxy to the application tier, which means that all the requests to the end applications go through it. Oracle Access Manager WebGate is an OHS web server plugin that intercepts every request going to the end application. If a resource being accessed is protected (requires an authenticated session), the WebGate initiates OIDC authentication flow with Identity Cloud Service through the user's browser. For more information about the flows supported by the OpenID Connect WebGate, see the [Oracle Access Manager documentation](#).

With this setup, a user already logged in to Azure AD can navigate to the JD Edwards EnterpriseOne or PeopleSoft application without logging in again, through Oracle Identity Cloud Service. Customers that deploy this solution gain the benefits of single sign-on, including a single set of credentials, an improved sign-on experience, improved security, and reduced help-desk cost.

To learn more about setting up single sign-on for JD Edwards EnterpriseOne or PeopleSoft with Azure AD, see the associated [Oracle whitepaper](#).

## PeopleSoft

Oracle's PeopleSoft application suite contains software for human resources and financial management. The application suite is multi-tiered and applications include human resource management systems (HRMS), customer relationship management (CRM), financials and supply chain management (FSCM), and enterprise performance management (EPM).

It's recommended that each tier of the software suite be deployed in its own subnet. An Oracle database or Microsoft SQL Server is required as the backend database for the

application. This section discusses details on deploying PeopleSoft with an Oracle database backend.

The following is a canonical architecture for deploying the PeopleSoft application suite in a cross-cloud architecture (Figure 5).

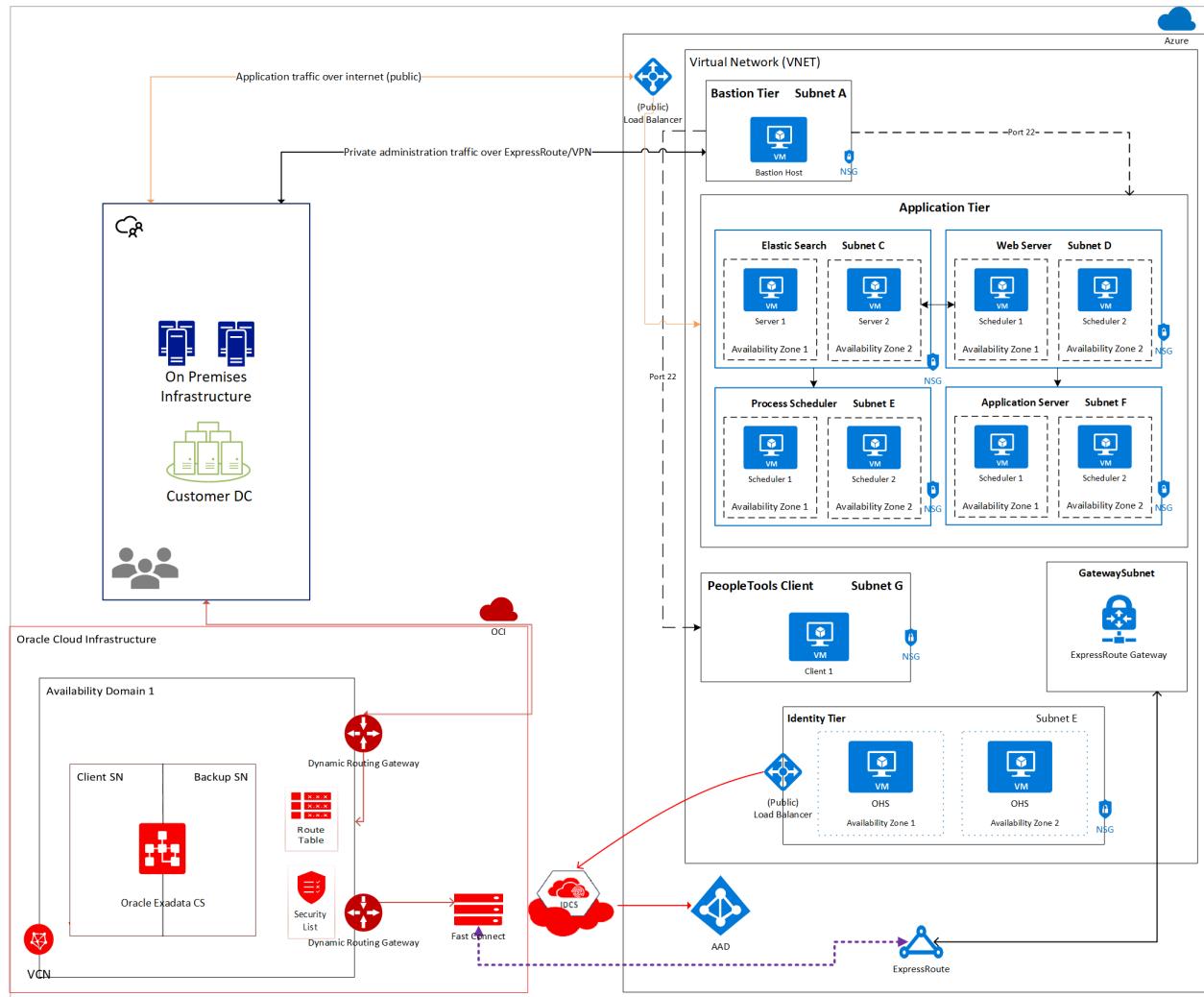


Figure 5: PeopleSoft cross-cloud architecture

In this sample architecture, the virtual network in Azure is connected to the virtual cloud network in OCI using the cross-cloud interconnect. The application tier is set up in Azure, whereas the database is set up in OCI. It's recommended to deploy each component to its own subnet with network security groups to allow traffic only from specific subnets on specific ports.

The architecture can also be adapted for deployment entirely on Azure with highly available Oracle databases configured using Oracle Data Guard in two availability zones in a region. The following diagram (Figure 6) is an example of this architectural pattern:

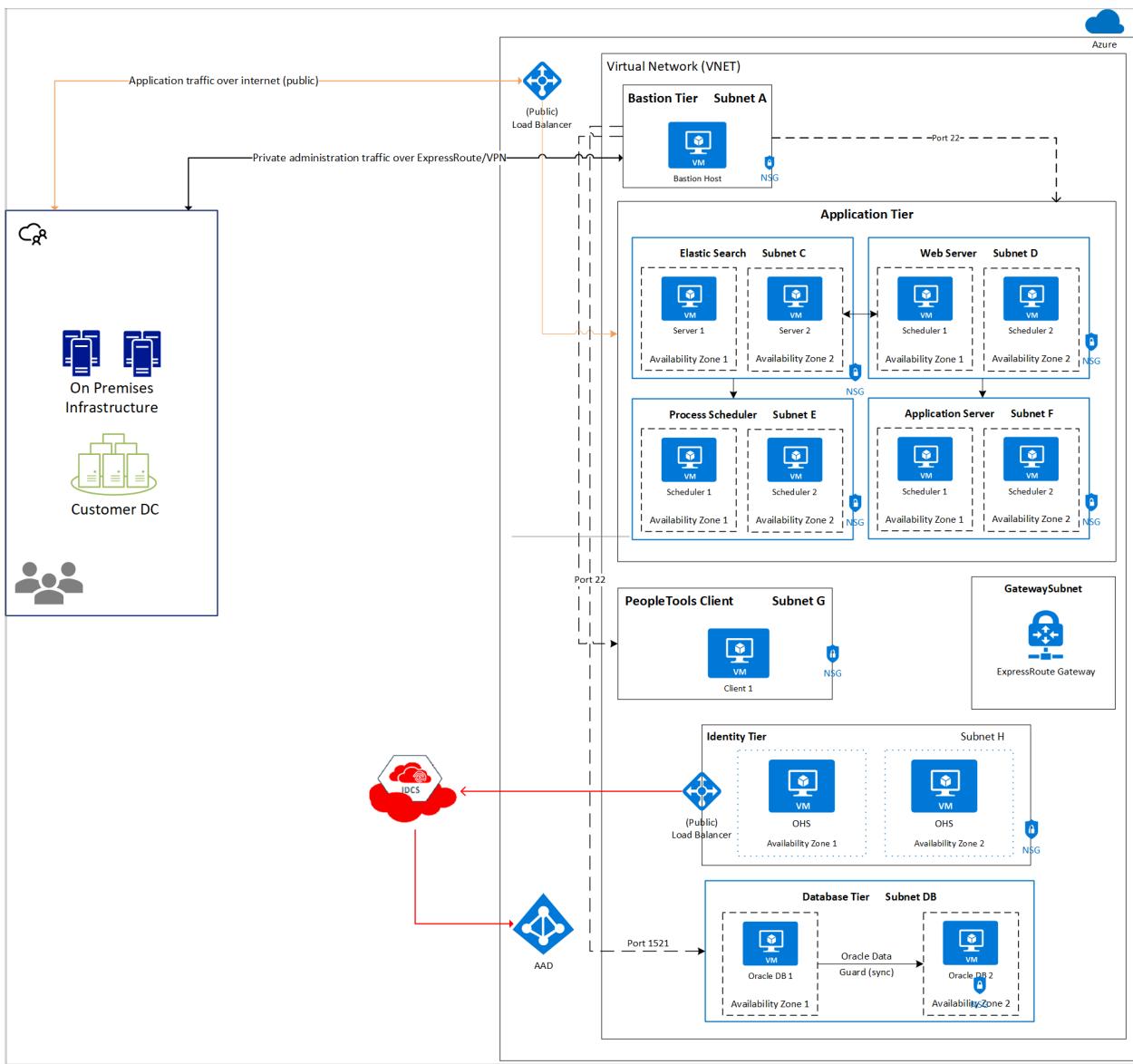


Figure 6: PeopleSoft Azure-only architecture

The following sections describe the different components at a high level.

## Bastion tier

The bastion host is an optional component that you can use as a jump server to access the application and database instances. The bastion host VM can have a public IP address assigned to it, although the recommendation is to set up an ExpressRoute connection or site-to-site VPN with your on-premises network for secure access. Additionally, only SSH (port 22, Linux) or RDP (port 3389, Windows Server) should be opened for incoming traffic. For high availability, deploy a bastion host in two availability zones or in a single availability set.

You may also enable SSH agent forwarding on your VMs, which allows you to access other VMs in the virtual network by forwarding the credentials from your bastion host. Or, use SSH tunneling to access other instances.

Here's an example of agent forwarding:

```
ssh -A -t user@BASTION_SERVER_IP ssh -A root@TARGET_SERVER_IP`
```

This command connects to the bastion and then immediately runs `ssh` again, so you get a terminal on the target instance. You may need to specify a user other than root on the target instance if your cluster is configured differently. The `-A` argument forwards the agent connection so your private key on your local machine is used automatically. Note that agent forwarding is a chain, so the second `ssh` command also includes `-A` so that any subsequent SSH connections initiated from the target instance also use your local private key.

## Application tier

The application tier contains instances of the PeopleSoft application servers, PeopleSoft web servers, elastic search, and PeopleSoft Process Scheduler. An Azure load balancer is set up to accept requests from users, which are routed to the appropriate server in the application tier.

For high availability, consider setting up redundant instances of each server in the application tier across different availability zones. The Azure load balancer can be set up with multiple back-end pools to direct each request to the right server.

## PeopleTools Client

The PeopleTools Client is used to perform administration activities, such as development, migration, and upgrade. Because the PeopleTools Client isn't required for achieving high availability of your application, redundant servers of PeopleTools Client aren't needed.

## Database tier

The Database tier contains the database instances for the application. The database can be either an Oracle DB, Oracle RAC, or Oracle Exadata Database system.

If the choice is to use Oracle DB, the database instance may be deployed on Azure via the Oracle DB images available on the Azure Marketplace. Alternatively, you may use the interconnect between Azure and OCI to deploy the Oracle DB in a PaaS model on OCI.

For Oracle RAC, you can use OCI in PaaS model. It is recommended that you use a two-node RAC system. While it is possible to deploy Oracle RAC on Azure CloudSimple in IaaS model, it is not a supported configuration by Oracle. Refer to [Oracle programs eligible for Authorized Cloud Environments](#).

Finally, for Exadata systems, use the OCI interconnect and deploy the Exadata system in OCI. The preceding architecture diagram above shows an Exadata system deployed in OCI across two subnets.

For production scenarios, deploy multiple instances of the database across two availability zones (if deploying in Azure) or two availability domains (in OCI). Use Oracle Active Data Guard to synchronize the primary and standby databases.

The database tier only receives requests from the middle tier. It is recommended that you set up a network security group (security list if deploying the database in OCI) to only allow requests on port 1521 from the middle tier and port 22 from the bastion server for administrative reasons.

For databases deployed in OCI, a separate virtual cloud network must be set up with a dynamic routing gateway (DRG) that is connected to your FastConnect circuit.

## Identity tier

The Microsoft-Oracle partnership enables you to set up a unified identity across Azure, OCI, and your Oracle application. For JD Edwards EnterpriseOne or PeopleSoft application suite, an instance of the Oracle HTTP Server (OHS) is needed to set up single sign-on between Azure AD and Oracle IDCS.

OHS acts as a reverse proxy to the application tier, which means that all the requests to the end applications go through it. Oracle Access Manager WebGate is an OHS web server plugin that intercepts every request going to the end application. If a resource being accessed is protected (requires an authenticated session), the WebGate initiates OIDC authentication flow with Identity Cloud Service through the user's browser. For more information about the flows supported by the OpenID Connect WebGate, see the [Oracle Access Manager documentation](#).

With this setup, a user already logged in to Azure AD can navigate to the JD Edwards EnterpriseOne or PeopleSoft application without logging in again, through Oracle Identity Cloud Service. Customers that deploy this solution gain the benefits of single sign-on, including a single set of credentials, an improved sign-on experience, improved security, and reduced help-desk cost.

To learn more about setting up single sign-on for JD Edwards EnterpriseOne or PeopleSoft with Azure AD, see the associated [Oracle whitepaper](#).

## Next steps

Use [Terraform scripts](#) to set up Oracle apps in Azure and establish cross-cloud connectivity with OCI.

For more information and whitepapers about OCI, see the [Oracle Cloud](#) documentation.

# Oracle application solutions integrating Microsoft Azure and Oracle Cloud Infrastructure

Article • 10/13/2023

Applies to:  Linux VMs

Microsoft and Oracle have partnered to provide low latency, high throughput cross-cloud connectivity, allowing you to take advantage of the best of both clouds.

Using this cross-cloud connectivity, you can partition a multi-tier application to run your database tier on Oracle Cloud Infrastructure (OCI), and the application and other tiers on Microsoft Azure. The experience is similar to running the entire solution stack in a single cloud.

If you're interested in running your middleware, including WebLogic Server, on Azure infrastructure, but have the Oracle database running within OCI, see [WebLogic Server Azure Applications](#).

If you're interested in deploying Oracle solutions entirely on Azure infrastructure, see [Oracle VM images and their deployment on Microsoft Azure](#).

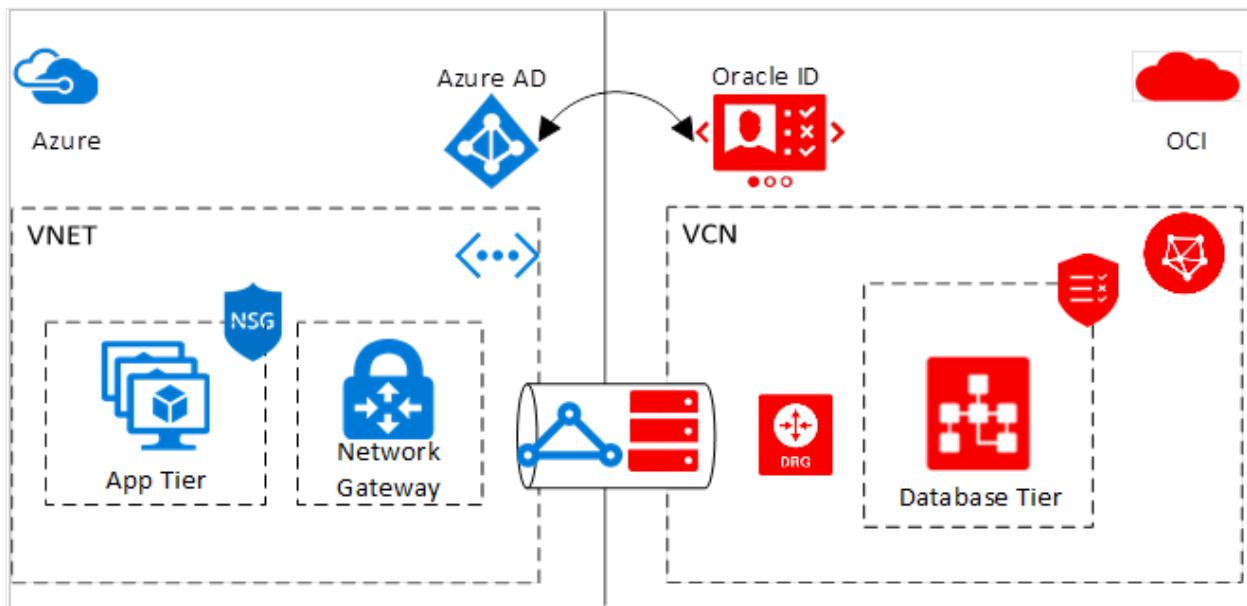
## Scenario overview

*Cross-cloud connectivity* provides a solution for you to run Oracle's industry-leading applications and your own custom applications on Azure virtual machines while enjoying the benefits of hosted database services in OCI.

The following applications are certified in a cross-cloud configuration:

- E-Business Suite
- JD Edwards EnterpriseOne
- PeopleSoft
- Oracle Retail applications
- Oracle Hyperion Financial Management

The following diagram is a high-level overview of the connected solution. For simplicity, the diagram shows only an application tier and a data tier. Depending on the application architecture, your solution could include other tiers such as a WebLogic Server cluster or web tier in Azure.



## Region availability

Cross-cloud connectivity is limited to the following regions:

- Azure Brazil South & OCI Vinhedo (Brazil Southeast)
- Azure Canada Central & OCI Toronto (Canada Southeast)
- Azure East US & OCI Ashburn, VA (US East)
- Azure Germany West Central & OCI Germany Central (Frankfurt)
- Azure Japan East & OCI Tokyo (Japan East)
- Azure Korea Central region & OCI South Korea Central (Seoul)
- Azure South Africa North & South Africa Central (Johannesburg)
- Azure Southeast Asia region & OCI Singapore (Singapore)
- Azure UK South & OCI London (UK South)
- Azure West Europe & OCI Amsterdam (Netherlands Northwest)
- Azure West US & OCI San Jose (US West)
- Azure West US 3 & OCI US West (Phoenix)

## Networking

Enterprise customers often choose to diversify and deploy workloads over multiple clouds for various business and operational reasons. To diversify, you can interconnect cloud networks using the internet, IPSec VPN, or using the cloud provider's direct connectivity solution with your on-premises network. Interconnecting cloud networks can require significant investments in time, money, design, procurement, installation, testing, and operations.

To address these challenges, Oracle and Microsoft have enabled an integrated multicloud experience. Establish *cross-cloud networking* by connecting an [ExpressRoute](#)

circuit in Microsoft Azure with a [FastConnect](#) circuit in OCI. This connectivity is possible where an Azure ExpressRoute peering location is in proximity to or in the same peering location as the OCI FastConnect. This setup allows for secure, fast connectivity between the two clouds without the need for an intermediate service provider.

Using ExpressRoute and FastConnect, you can peer a virtual network in Azure with a virtual cloud network in OCI, if the private IP address space doesn't overlap. Peering the two networks allows a resource in the virtual network to communicate to a resource in the OCI virtual cloud network as if they're both in the same network.

## Network security

Network security is a crucial component of any enterprise application, and is central to this multicloud solution. Any traffic going over ExpressRoute and FastConnect passes over a private network. This configuration allows for secure communication between an Azure virtual network and an Oracle virtual cloud network. You don't need to provide a public IP address to any virtual machines in Azure. Similarly, you don't need an internet gateway in OCI. All communication happens by using the private IP address of the machines.

Additionally, you can set up [security lists](#) on your OCI virtual cloud network and security rules, attached to Azure [network security groups](#). Use these rules to control the traffic flowing between machines in the virtual networks. You can add network security rules at a machine level, at a subnet level, and at the virtual network level.

The [WebLogic Server Azure Applications](#) each create a network security group preconfigured to work with WebLogic Server's port configurations.

## Identity

Identity is one of the core pillars of the partnership between Microsoft and Oracle. Significant work has been done to integrate [Oracle Identity Cloud Service](#) (IDCS) with [Microsoft Entra ID](#) (Microsoft Entra ID). Microsoft Entra ID is Microsoft's cloud-based identity and access management service. Your users can sign in and access various resources with help from Microsoft Entra ID. Microsoft Entra ID also allows you to manage your users and their permissions.

Currently, this integration allows you to manage in one central location. Microsoft Entra ID synchronizes any changes in the directory with the corresponding Oracle directory and is used for single sign-on to cross-cloud Oracle solutions.

## Next steps

- Get started with a [cross-cloud network](#) between Azure and OCI.
- For more information and whitepapers about OCI, see [Oracle Cloud Infrastructure ↗](#).

# Reference architectures for Oracle Database Enterprise Edition on Azure

Article • 06/13/2023

Applies to:  Linux VMs

This article includes information on deploying a highly available Oracle database on Azure. In addition, this guide dives into disaster recovery considerations. These architectures have been created based on customer deployments. This guide only applies to Oracle Database Enterprise Edition.

If you're interested in learning more about maximizing the performance of your Oracle database, see [Design and implement an Oracle database in Azure](#).

## Prerequisites

- An understanding of the different concepts of Azure such as [availability zones](#)
- Oracle Database Enterprise Edition 12c or later
- Awareness of the licensing implications when using the solutions in this article

## High availability for Oracle databases

Achieving high availability in the cloud is an important part of every organization's planning and design. Azure offers [availability zones](#) and [availability sets](#) to be used in regions where availability zones are unavailable. For more information about how to design for the cloud, see [Availability options for Azure Virtual Machines](#).

In addition to cloud-native tools and offerings, Oracle provides solutions for high availability that can be set up on Azure:

- [Oracle Data Guard ↗](#)
- [Data Guard with FSFO ↗](#)
- [Sharding ↗](#)
- [GoldenGate ↗](#)

This guide covers reference architectures for each of these solutions.

When you migrate or create applications for the cloud, we recommend using cloud-native patterns such as [retry pattern](#) and [circuit breaker pattern](#). For other patterns that could help make your application more resilient, see [Cloud Design Patterns guide](#).

## Oracle RAC in the cloud

Oracle Real Application Cluster (RAC) is a solution by Oracle to help customers achieve high throughputs by having many instances accessing one database storage. This pattern is a shared-all architecture. While Oracle RAC can be used for high availability on-premises, Oracle RAC alone can't be used for high availability in the cloud. Oracle RAC only protects against instance level failures and not against rack-level or datacenter-level failures. For this reason, Oracle recommends using Oracle Data Guard with your database, whether single instance or RAC, for high availability.

Customers generally require a high SLA to run mission critical applications. Oracle doesn't currently certify or support Oracle RAC on Azure. However, Azure offers features such as availability zones and planned maintenance windows to help protect against instance-level failures. In addition to these offerings, you can use Oracle Data Guard, Oracle GoldenGate, and Oracle Sharding for high performance and resiliency. These technologies can help protect your databases from rack-level, datacenter-level, and geo-political failures.

When you run Oracle Databases on multiple [availability zones](#) with Oracle Data Guard or GoldenGate, you can get an uptime SLA of 99.99%. In Azure regions where availability zones aren't yet present, you can use [availability sets](#) and achieve an uptime SLA of 99.95%.

### Note

You can have a uptime target that is much higher than the uptime SLA provided by Microsoft.

## Disaster recovery for Oracle databases

When hosting your mission-critical applications in the cloud, it's important to design for high availability and disaster recovery.

For Oracle Database Enterprise Edition, Oracle Data Guard is a useful feature for disaster recovery. You can set up a standby database instance in a [paired Azure region](#) and set up Data Guard failover for disaster recovery. For zero data loss, we recommend that you deploy an Oracle Data Guard Far Sync instance in addition to Active Data Guard.

If your application permits the latency, consider setting up the Data Guard Far Sync instance in a different availability zone than your Oracle primary database. Test the configuration thoroughly. Use a *Maximum Availability* mode to set up synchronous

transport of your redo files to the Far Sync instance. These files are then transferred asynchronously to the standby database.

Your application might not allow for the performance loss when setting up Far Sync instance in another availability zone in *Maximum Availability* mode (synchronous). If not, you might set up a Far Sync instance in the same availability zone as your primary database. For added availability, consider setting up multiple Far Sync instances close to your primary database and at least one instance close to your standby database, if the role transitions. For more information, see [Oracle Active Data Guard Far Sync](#).

When you use Oracle Standard Edition databases, there are ISV solutions that allow you to set up high availability and disaster recovery, such as DBVisit Standby.

## Reference architectures

### Oracle Data Guard

Oracle Data Guard ensures high availability, data protection, and disaster recovery for enterprise data. Data Guard maintains standby databases as transactionally consistent copies of the primary database. Depending on the distance between the primary and secondary databases and the application tolerance for latency, you can set up synchronous or asynchronous replication. If the primary database is unavailable because of a planned or an unplanned outage, Data Guard can switch any standby database to the primary role, minimizing downtime.

When using Oracle Data Guard, you might also open your secondary database for read-only purposes. This configuration is called Active Data Guard. Oracle Database 12c introduced a feature called Data Guard Far Sync Instance. This instance allows you to set up a zero data loss configuration of your Oracle database without having to compromise on performance.

#### Note

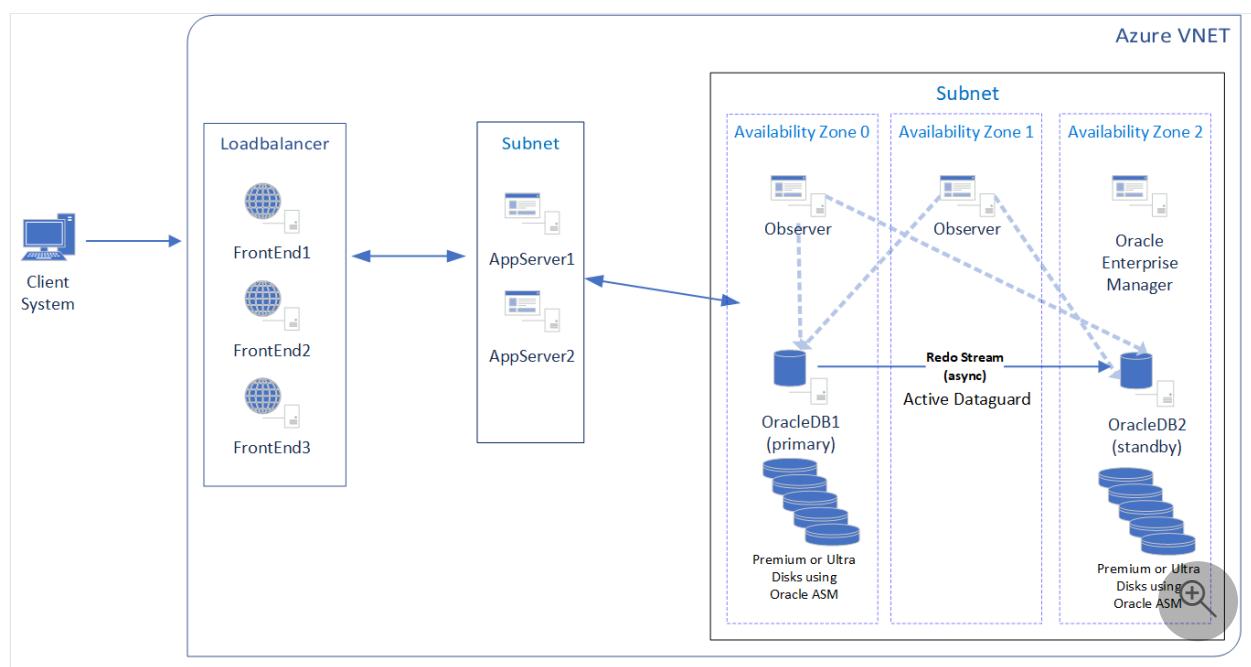
Active Data Guard requires additional licensing. This license is also required to use the Far Sync feature. Contact with your Oracle representative to discuss the licensing implications.

### Oracle Data Guard with Fast-Start Failover

Oracle Data Guard with Fast-Start Failover (FSFO) can provide more resiliency by setting up the broker on a separate machine. The Data Guard broker and the secondary database both run the observer and observe the primary database for downtime. This approach allows for redundancy in your Data Guard observer setup as well.

With Oracle Database version 12.2 and above, it's also possible to configure multiple observers with a single Oracle Data Guard broker configuration. This setup provides extra availability, in case one observer and the secondary database experience downtime. Data Guard Broker is lightweight and can be hosted on a relatively small virtual machine. For more information about Data Guard Broker and its advantages, see [Oracle Data Guard Broker Concepts](#).

The following diagram is a recommended architecture for using Oracle Data Guard on Azure with availability zones. This architecture allows you to get a VM uptime SLA of 99.99%.



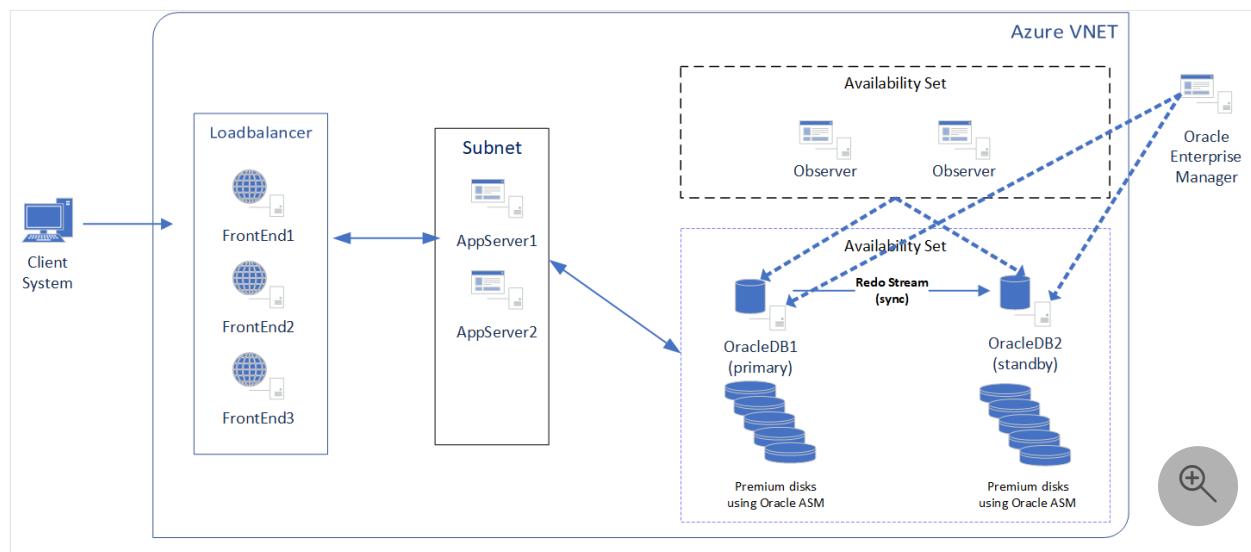
In the preceding diagram, the client system accesses a custom application with Oracle backend by using the web. The web frontend is configured in a load balancer. The web frontend makes a call to the appropriate application server to handle the work. The application server queries the primary Oracle database. The Oracle database has been configured using a hyperthreaded [memory optimized virtual machine](#) with [constrained core vCPUs](#) to save on licensing costs and maximize performance. Multiple premium or ultra disks (Managed Disks) are used for performance and high availability.

The Oracle databases are placed in multiple availability zones for high availability. Each zone is made up of one or more data centers equipped with independent power, cooling, and networking. To ensure resiliency, a minimum of three separate zones are set up in all enabled regions. The physical separation of availability zones within a region

protects the data from data center failures. Additionally, two FSFO observers are set up across two availability zones to initiate and fail over the database to the secondary when an outage occurs.

You might set up other observers or standby databases in a different availability zone, AZ 1, in this case, than the zone shown in the preceding architecture. Finally, Oracle Enterprise Manager (OEM) monitors Oracle databases for uptime and performance. OEM also allows you to generate various performance and usage reports.

In regions where availability zones aren't supported, you might use availability sets to deploy your Oracle Database in a highly available manner. The following diagram is a reference architecture of this use:



### ⓘ Note

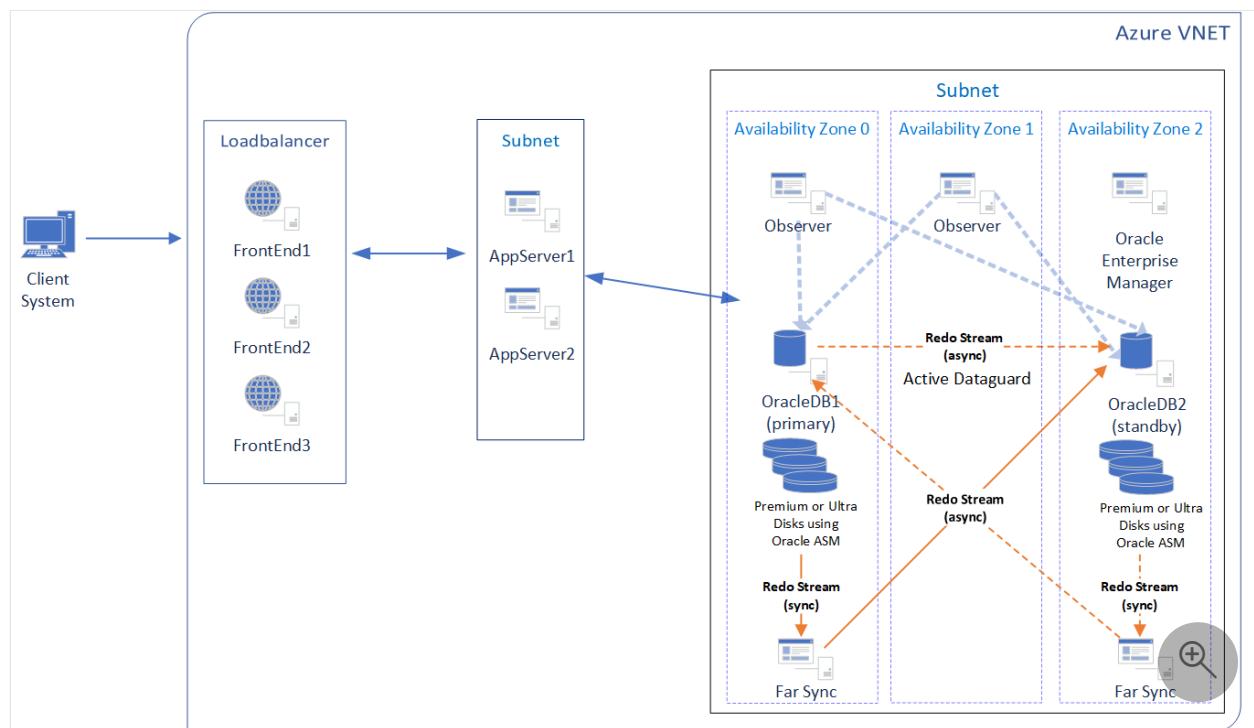
- Because there is only one instance of OEM being deployed, you don't have to place the Oracle Enterprise Manager VM in an availability set.
- Ultra disks aren't currently supported in an availability set configuration.

## Oracle Data Guard Far Sync

Oracle Data Guard Far Sync provides zero data loss protection capability for Oracle Databases. This capability allows you to protect against data loss if your database machine fails. Oracle Data Guard Far Sync needs to be installed on a separate VM. Far Sync is a lightweight Oracle instance that only has a control file, password file, spfile, and standby logs. There are no data files or redo log files.

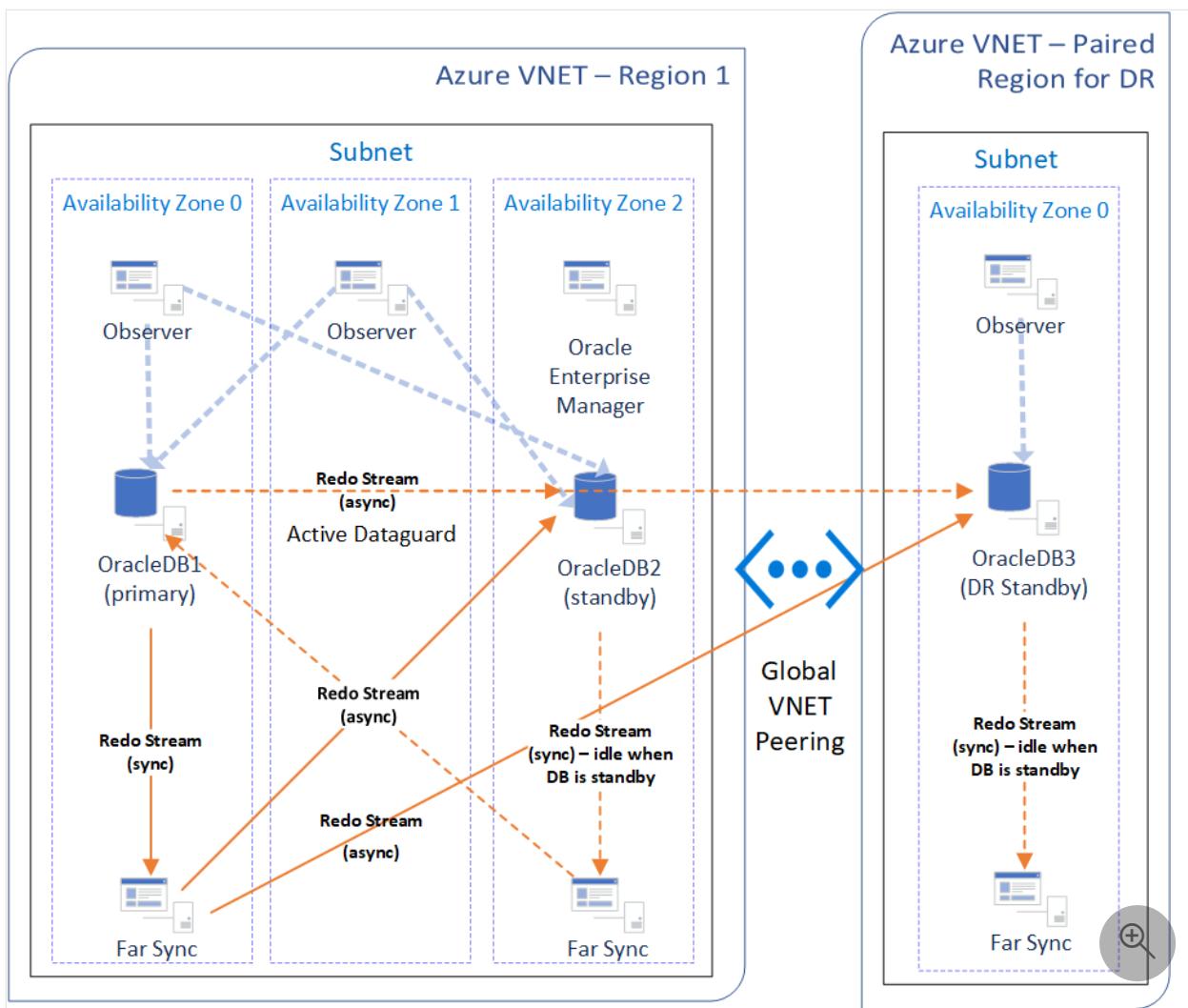
For zero data loss protection, there must be synchronous communication between your primary database and the Far Sync instance. The Far Sync instance receives redo from the primary in a synchronous manner and forwards it immediately to all the standby databases in an asynchronous manner. This setup also reduces the overhead on the primary database, because it only has to send the redo to the Far Sync instance rather than all the standby databases. If a Far Sync instance fails, Data Guard automatically uses asynchronous transport to the secondary database from the primary database to maintain near-zero data loss protection. For added resiliency, customers might deploy multiple Far Sync instances per each database instance, including primary and secondaries.

The following diagram is a high availability architecture using Oracle Data Guard Far Sync:



In the preceding architecture, there's a Far Sync instance deployed in the same availability zone as the database instance to reduce the latency between the two. In cases where the application is latency sensitive, consider deploying your database and Far Sync instance or instances in a [proximity placement group](#).

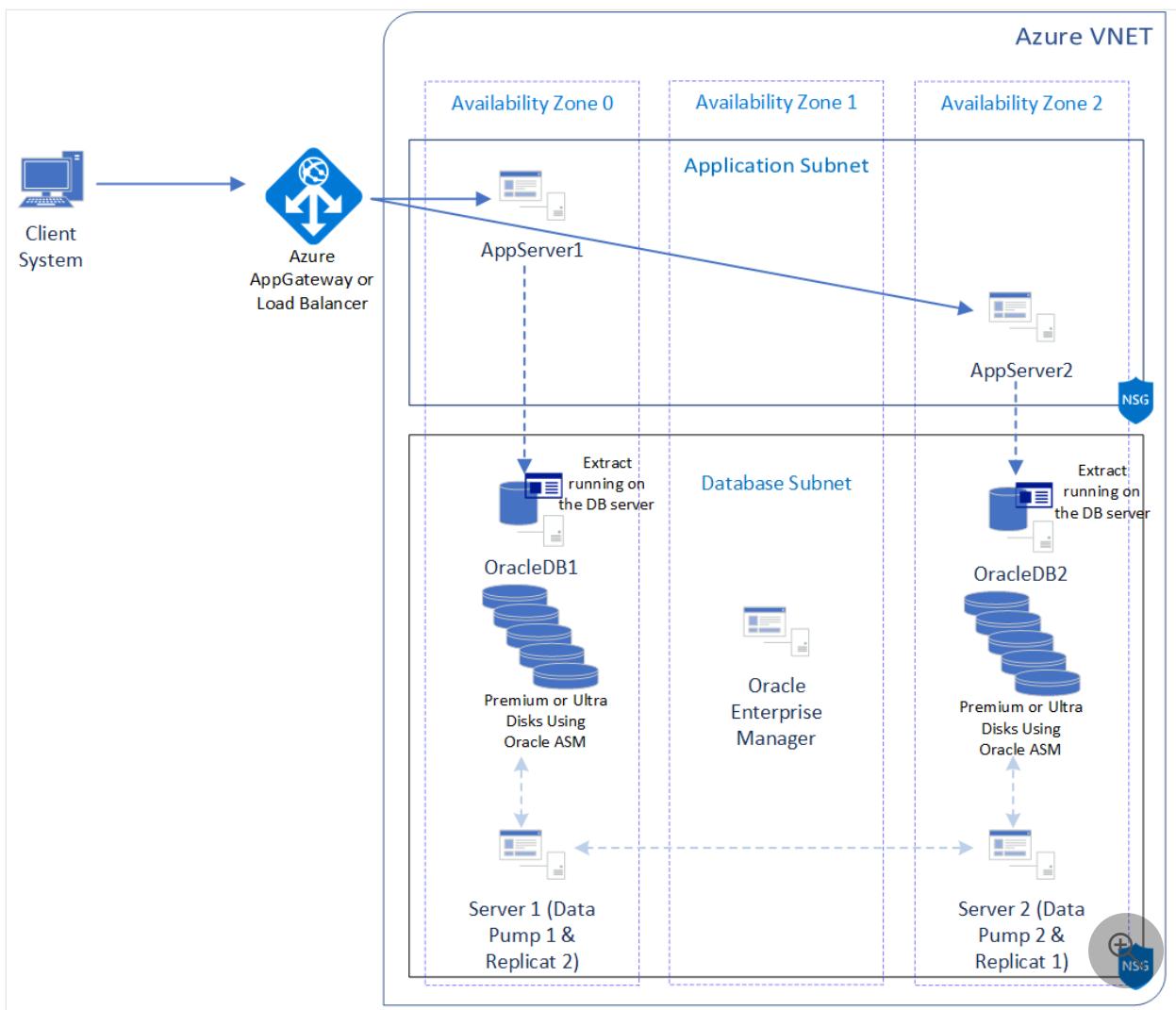
The following diagram is an architecture that uses Oracle Data Guard FSFO and Far Sync to achieve high availability and disaster recovery:



## Oracle GoldenGate

GoldenGate enables the exchange and manipulation of data at the transaction level among multiple, heterogeneous platforms across the enterprise. It moves committed transactions with transaction integrity and minimal overhead on your existing infrastructure. Its modular architecture gives you the flexibility to extract and replicate selected data records, transactional changes, and changes to data definition language (DDL) across various topologies.

Oracle GoldenGate allows you to configure your database for high availability by providing bidirectional replication. This approach allows you to set up a *multi-master* or *active-active configuration*. The following diagram is a recommended architecture for Oracle GoldenGate active-active setup on Azure. In the following architecture, the Oracle database has been configured using a hyperthreaded [memory optimized virtual machine](#) with [constrained core vCPUs](#) to save on licensing costs and maximize performance. The architecture uses multiple premium or ultra disks (managed disks) for performance and availability.



### ① Note

A similar architecture can be set up using availability sets in regions where availability zones aren't currently available.

Oracle GoldenGate has processes such as *Extract*, *Pump*, and *Replicat* that help you asynchronously replicate your data from one Oracle database server to another. These processes allow you to set up a bidirectional replication to ensure high availability of your database if there's availability zone-level downtime.

In the preceding diagram, the *Extract* process runs on the same server as your Oracle database. The *Data Pump* and *Replicat* processes run on a separate server in the same availability zone. The *Replicat* process is used to receive data from the database in the other availability zone and commit the data to the Oracle database in its availability zone. Similarly, the *Data Pump* process sends data that the *Extract* process extracts to the *Replicat* process in the other availability zone.

While the preceding architecture diagram shows the *Data Pump* and *Replicat* processes configured on a separate server, you might set up all the Oracle GoldenGate processes

on the same server, based on the capacity and usage of your server. Always consult your AWR report and the metrics in Azure to understand the usage pattern of your server.

When setting up Oracle GoldenGate bidirectional replication in different availability zones or different regions, it's important to ensure that the latency between the different components is acceptable for your application. The latency between availability zones and regions can vary. Latency depends on multiple factors. We recommend that you set up performance tests between your application tier and your database tier in different availability zones or regions. The tests can confirm that the configuration meets your application performance requirements.

The application tier can be set up in its own subnet and the database tier can be separated into its own subnet. When possible, consider using [Azure Application Gateway](#) to load-balance traffic between your application servers. Application Gateway is a robust web traffic load balancer. It provides cookie-based session affinity that keeps a user session on the same server, minimizing the conflicts on the database. Alternatives to Application Gateway are [Azure Load Balancer](#) and [Azure Traffic Manager](#).

## Oracle Sharding

Sharding is a data tier pattern that was introduced in Oracle 12.2. It allows you to horizontally partition and scale your data across independent databases. It's a share-nothing architecture where each database is hosted on a dedicated virtual machine. This pattern enables high read and write throughput in addition to resiliency and increased availability.

This pattern eliminates single points of failure, provides fault isolation, and enables rolling upgrades without downtime. The downtime of one shard or a data center-level failure doesn't affect the performance or availability of the other shards in other data centers.

Sharding is suitable for high throughput OLTP applications that can't afford any downtime. All rows with the same sharding key are always guaranteed to be on the same shard. This fact increases performance, providing high consistency. Applications that use sharding must have a well-defined data model and data distribution strategy, such as consistent hash, range, list, or composite. The strategy primarily accesses data using a sharding key, for example, *customerId* or *accountNum*. Sharding also allows you to store particular sets of data closer to the end customers, thus helping meet your performance and compliance requirements.

We recommend that you replicate your shards for high availability and disaster recovery. This setup can be done using Oracle technologies such as Oracle Data Guard or Oracle

GoldenGate. A unit of replication can be a shard, a part of a shard, or a group of shards. An outage or slowdown of one or more shards doesn't affect the availability of a sharded database.

For high availability, the standby shards can be placed in the same availability zone where the primary shards are placed. For disaster recovery, the standby shards can be located in another region. You might also deploy shards in multiple regions to serve traffic in those regions. To learn more about configuring high availability and replication of your sharded database, see [Shard-Level High Availability](#).

Oracle Sharding primarily consists of the following components. For more information, see [Oracle Sharding Overview](#):

- **Shard catalog.** Special-purpose Oracle database that is a persistent store for all Shard database configuration data. All configuration changes such as adding or removing shards, mapping of the data, and DDLs in a shard database are initiated on the shard catalog. The shard catalog also contains the master copy of all duplicated tables in an SDB.

The shard catalog uses materialized views to automatically replicate changes to duplicated tables in all shards. The shard catalog database also acts as a query coordinator used to process multi-shard queries and queries that don't specify a sharding key.

We recommend using Oracle Data Guard with availability zones or availability sets for shard catalog high availability as a best practice. The availability of the shard catalog has no effect on the availability of the sharded database. A downtime in the shard catalog only affects maintenance operations and multishard queries during the brief period that the Data Guard failover completes. The SDB continues to route and run online transactions. A catalog outage doesn't affect them.

- **Shard directors.** Lightweight services that need to be deployed in each region/availability zone that your shards reside in. Shard Directors are Global Service Managers deployed in the context of Oracle Sharding. For high availability, we recommend that you deploy at least one shard director in each availability zone that your shards exist in.

When connecting to the database initially, the shard director sets up the routing information and caches the information for subsequent requests, which bypass the shard director. Once the session is established with a shard, all SQL queries and DMLs are supported and executed in the scope of the given shard. This routing is fast and is used for all OLTP workloads that perform intra-shard transactions. We recommend that you to use direct routing for all OLTP workloads that require the

highest performance and availability. The routing cache automatically refreshes when a shard becomes unavailable or changes occur to the sharding topology.

For high-performance, data-dependent routing, Oracle recommends using a connection pool when accessing data in the sharded database. Oracle connection pools, language-specific libraries, and drivers support Oracle Sharding. For more information, see [Oracle Sharding Overview](#).

- **Global service.** Global service is similar to the regular database service. In addition to all the properties of a database service, a global service has properties for sharded databases. These properties include region affinity between clients and shard and replication lag tolerance. Only one global service needs to be created to read/write data to and from a sharded database. When using Active Data Guard and setting up read-only replicas of the shards, you can create another global service for read-only workloads. The client can use these global services to connect to the database.
- **Shard databases.** Shard databases are your Oracle databases. Each database is replicated using Oracle Data Guard in a Broker configuration with FSFO enabled. You don't need to set up Data Guard failover and replication on each shard. This aspect is automatically configured and deployed when the shared database is created. If a particular shard fails, Oracle Sharing fails over database connections from the primary to the standby.

You can deploy and manage Oracle sharded databases with two interfaces: Oracle Enterprise Manager Cloud Control GUI and the `GDSCTL` command-line utility. You can even monitor the different shards for availability and performance using Cloud control. The `GDSCTL DEPLOY` command automatically creates the shards and their respective listeners. In addition, this command automatically deploys the replication configuration used for shard-level high availability specified by the administrator.

There are different ways to shard a database:

- System-managed sharding: Automatically distributes across shards using partitioning
- User-defined sharding: Allows you to specify the mapping of the data to the shards, which works well when there are regulatory or data-localization requirements
- Composite sharding: A combination of system-managed and user-defined sharding for different *shardspaces*
- Table subpartitions: Similar to a regular partitioned table

For more information, see [Sharding Methods](#).

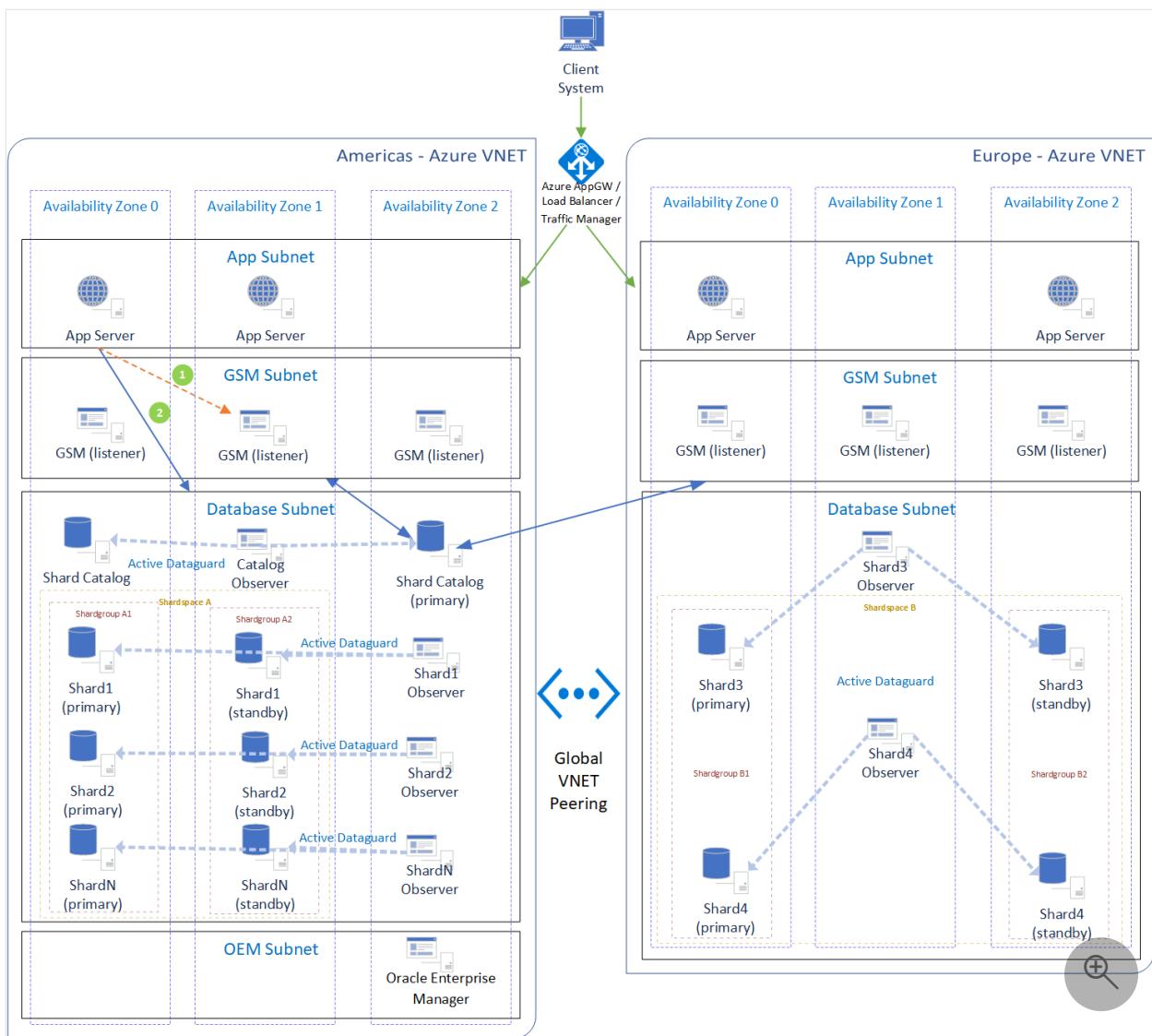
A sharded database looks like a single database to applications and developers. When you migrate to a sharded database, plan carefully to understand which tables are duplicated versus sharded.

Duplicated tables are stored on all shards, whereas sharded tables are distributed across different shards. We recommend that you duplicate small and dimensional tables and distribute/shard the fact tables. Data can be loaded into your sharded database using either the shard catalog as the central coordinator or by running Data Pump on each shard. For more information, see [Migrating Data to a Sharded Database](#).

## Oracle Sharding with Data Guard

Oracle Data Guard can be used for sharding with system-managed, user-defined, and composite sharding methods.

The following diagram is a reference architecture for Oracle Sharding with Oracle Data Guard used for high availability of each shard. The architecture diagram shows a *composite sharding method*. The architecture diagram likely differs for applications with different requirements for data locality, load balancing, high availability, and disaster recovery. Applications might use different method for sharding. Oracle Sharding allows you to meet these requirements and scale horizontally and efficiently by providing these options. A similar architecture can even be deployed using Oracle GoldenGate.



System-managed sharding is the easiest to configure and manage. User-defined sharding or composite sharding is well suited for scenarios where your data and application are geo-distributed or in scenarios where you need to have control over the replication of each shard.

In the preceding architecture, composite sharding is used to geodistribute the data and horizontally scale out your application tiers. Composite sharding is a combination of system-managed and user-defined sharding and thus provides the benefit of both methods. In the preceding scenario, data is first sharded across multiple shardspaces separated by region. Then, the data is further partitioned by using consistent hash across multiple shards in the shardspace.

Each shardspace contains multiple shardgroups. Each shardgroup has multiple shards and is a unit of replication. Each shardgroup contains all the data in the shardspace. Shardgroups A1 and B1 are primary shardgroups, while shardgroups A2 and B2 are standbys. You might choose to have individual shards be the unit of replication, rather than a shardgroup.

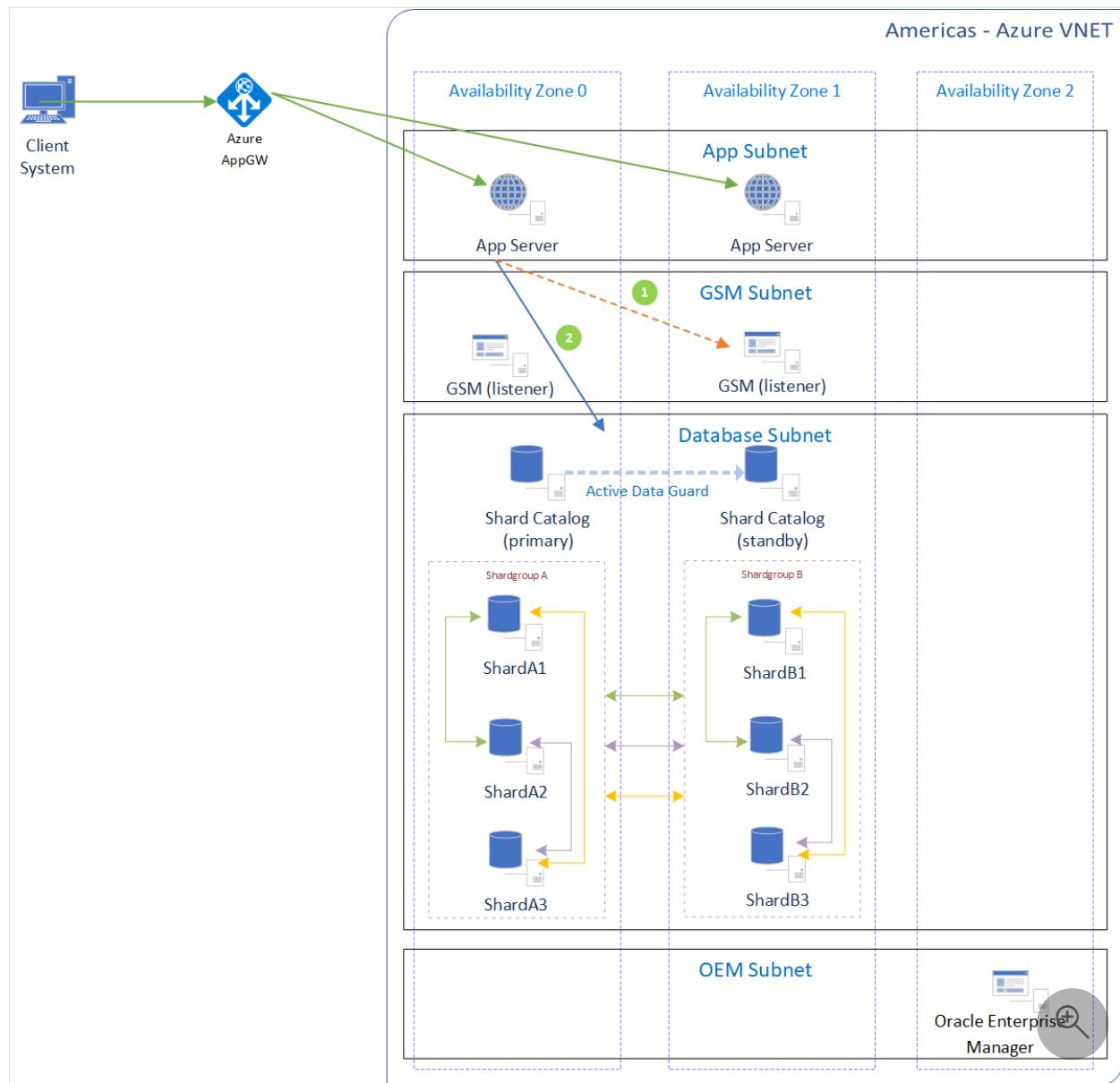
In the preceding architecture, a Global Service Manager (GSM)/shard director is deployed in every availability zone for high availability. We recommend that you deploy at least one GSM/shard director per data center/region. Additionally, an instance of the application server is deployed in every availability zone that contains a shardgroup. This setup allows the application to keep the latency between the application server and the database/shardgroup low. If a database fails, the application server in the same zone as the standby database can handle requests once the database role transition happens. Azure Application Gateway and the shard director keep track of the request and response latency and route requests accordingly.

From an application standpoint, the client system makes a request to Azure Application Gateway or other load-balancing technologies in Azure, which redirects the request to the region closest to the client. Azure Application Gateway also supports sticky sessions, so any requests coming from the same client are routed to the same application server. The application server uses connection pooling in data access drivers. This feature is available in drivers such as JDBC, ODP.NET, and OCI. The drivers can recognize sharding keys specified as part of the request. Oracle Universal Connection Pool (UCP) for JDBC clients can enable non-Oracle application clients such as Apache Tomcat and IIS to work with Oracle Sharding. For more information, see [Overview of UCP Shared Pool for Database Sharding](#).

During the initial request, the application server connects to the shard director in its region to get routing information for the shard that the request needs to be routed to. Based on the sharding key passed, the director routes the application server to the respective shard. The application server caches this information by building a map, and for subsequent requests, bypasses the shard director and routes requests straight to the shard.

## Oracle Sharding with GoldenGate

The following diagram is a reference architecture for Oracle Sharding with Oracle GoldenGate for in-region high availability of each shard. As opposed to the preceding architecture, this architecture only portrays high availability within a single Azure region, with multiple availability zones. You can deploy a multi-region high availability sharded database, similar to the preceding example, by using Oracle GoldenGate.



The preceding reference architecture uses the *system-managed* sharding method to shard the data. Since Oracle GoldenGate replication is done at a chunk level, half the data replicated to one shard can be replicated to another shard. The other half can be replicated to a different shard.

The way the data gets replicated depends on the replication factor. With a replication factor of two, you have two copies of each chunk of data across your three shards in the shardgroup. Similarly, with a replication factor of three and three shards in your shardgroup, all the data in each shard is replicated to every other shard in the shardgroup. Each shard in the shardgroup can have a different replication factor. This setup helps you define your high availability and disaster recovery design efficiently within a shardgroup and across multiple shardgroups.

In the preceding architecture, shardgroup A and shardgroup B both contain the same data but reside in different availability zones. If both shardgroup A and shardgroup B have the same replication factor of three, each row/chunk of your sharded table is replicated six times across the two shardgroups. If shardgroup A has a replication factor

of three and shardgroup B has a replication factor of two, each row/chunk is replicated five times across the two shardgroups.

This setup prevents data loss if an instance-level or availability zone-level failure occurs. The application layer is able to read from and write to each shard. To minimize conflicts, Oracle Sharding designates a *master chunk* for each range of hash values. This feature ensures that write requests for a particular chunk are directed to the corresponding chunk. In addition, Oracle GoldenGate provides automatic conflict detection and resolution to handle any conflicts that might arise. For more information and limitations of implementing GoldenGate with Oracle Sharding, see [Using Oracle GoldenGate with a Sharded Database ↗](#).

In the preceding architecture, a GSM/shard director is deployed in every availability zone for high availability. We recommend that you deploy at least one GSM/shard director per data center or region. An instance of the application server is deployed in every availability zone that contains a shardgroup. This setup allows the application to keep the latency between the application server and the database/shardgroup low. If a database fails, the application server in the same zone as the standby database can handle requests once the database role transitions. Azure Application Gateway and the shard director keep track of the request and response latency and route requests accordingly.

From an application standpoint, the client system makes a request to Azure Application Gateway or other load-balancing technologies in Azure, which redirects the request to the region closest to the client. Azure Application Gateway also supports sticky sessions, so any requests coming from the same client are routed to the same application server. The application server uses connection pooling in data access drivers. This feature is available in drivers such as JDBC, ODP.NET, and OCI. The drivers can recognize sharding keys specified as part of the request. [Oracle Universal Connection Pool \(UCP\) ↗](#) for JDBC clients can enable non-Oracle application clients such as Apache Tomcat and IIS to work with Oracle Sharding.

During the initial request, the application server connects to the shard director in its region to get routing information for the shard that the request needs to be routed to. Based on the sharding key passed, the director routes the application server to the respective shard. The application server caches this information by building a map, and for subsequent requests, bypasses the shard director and routes requests straight to the shard.

## Patching and maintenance

When you deploy your Oracle workloads to Azure, Microsoft takes care of all host operating system level patching. Microsoft communicates any planned operating system level maintenance to customers in advance. Two servers from two different availability zones are never patched simultaneously. For more information on VM maintenance and patching, see [Availability options for Azure Virtual Machines](#).

Patching your virtual machine operating system can be automated using [Azure Automation Update Management](#). Patching and maintaining your Oracle database can be automated and scheduled using [Azure Pipelines](#) or [Azure Automation Update Management](#) to minimize downtime. For more information about continuous delivery and blue/green deployments, see [Progressive exposure techniques](#).

## Architecture and design considerations

- Consider using hyperthreaded [memory optimized virtual machine](#) with [constrained core vCPUs](#) for your Oracle Database VM to save on licensing costs and maximize performance. Use multiple premium or ultra disks (managed disks) for performance and availability.
- When you use managed disks, the disk/device name might change on restart. We recommend that you use the device UUID instead of the name to ensure your mounts persist in spite of restarting. For more information, see [Add the new file system to /etc/fstab](#).
- Use availability zones to achieve high availability in-region.
- Consider using ultra disks when available or premium disks for your Oracle database.
- Consider setting up a standby Oracle database in another Azure region using Oracle Data Guard.
- Consider using [proximity placement groups](#) to reduce the latency between your application and database tier.
- Set up [Oracle Enterprise Manager](#) for management, monitoring, and logging.
- Consider using Oracle Automatic Storage Management for streamlined storage management for your database.
- Use [Azure Pipelines](#) to manage patching and updates to your database without any downtime.
- Tweak your application code to add cloud-native patterns that might help your application be more resilient. Consider patterns such as [retry pattern](#), [circuit breaker pattern](#), and others defined in the [Cloud Design Patterns guide](#).

## Next steps

Review the following Oracle reference articles that apply to your scenario.

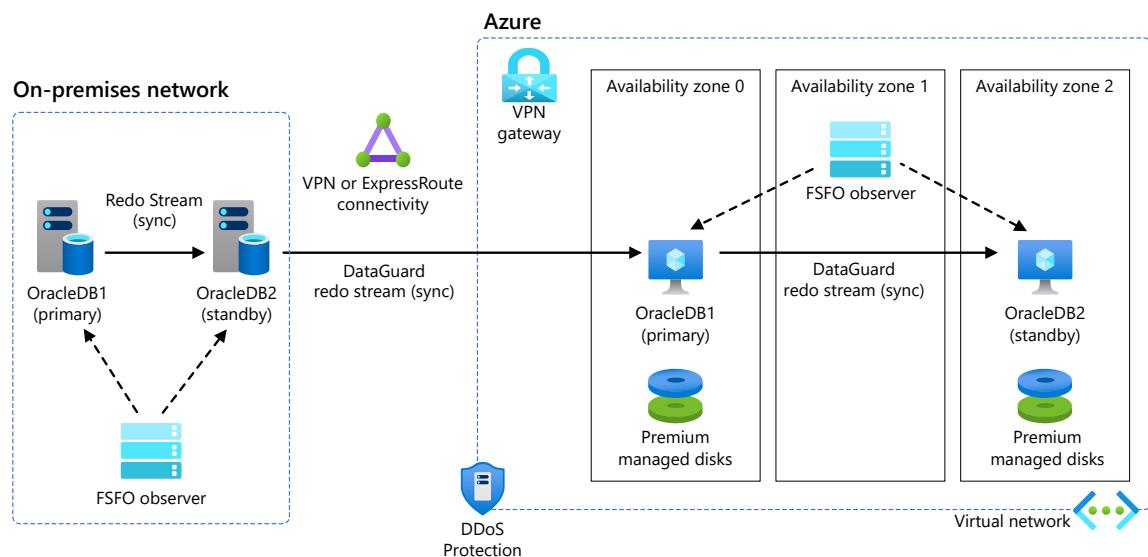
- [Introduction to Oracle Data Guard ↗](#)
- [Oracle Data Guard Broker Concepts ↗](#)
- [Configuring Oracle GoldenGate for Active-Active High Availability ↗](#)
- [Oracle Sharding Overview ↗](#)
- [Oracle Active Data Guard Far Sync Zero Data Loss at Any Distance ↗](#)

# Oracle database migration to Azure

Azure Load Balancer    Azure ExpressRoute    Azure VPN Gateway

This solution migrates an Oracle database and its applications to Azure. We use Oracle Active Data Guard for the database, and we use Azure Load Balancer for the application tier.

## Architecture



Download a [Visio file](#) of this architecture.

## Dataflow

1. Connect your Azure environment with your on-premises network via site-to-site VPN or ExpressRoute.
2. Use DataGuard to mark your OracleDB1 in Azure as your active stand-by.

3. Switch your OracleDB1 in Azure as primary and set up your OracleDB2 in Azure as your standby to finish your migration.

#### ⓘ Note

- This method works only when migrating to and from the same OS version and database version.
- *Assumption:* You're using DataGuard on-premises.

## Components

- [Azure Virtual Network](#)
- [Azure VPN Gateway](#)
- [Azure ExpressRoute](#)
- [Azure Virtual Machines](#)
- [Azure Managed Disks](#)

## Alternatives

If your database is over 2 TB, you can use Oracle Data Guard with Oracle Recovery Manager (RMAN), or use Data Pump to replicate changes after an initial *bulk* data transfer, which provides a minimal downtime migration.

## SQL Server Migration Assistant for Oracle

[Microsoft SQL Server Migration Assistant \(SSMA\) for Oracle](#) is a tool for migrating Oracle databases to Microsoft SQL Server and Azure SQL Database. SSMA for Oracle converts Oracle database objects to SQL Server database objects, creates those objects in SQL Server, and then migrates data from Oracle to SQL Server or Azure SQL Database.

## Scenario details

Oracle DB migrations can be accomplished in multiple ways. This solution covers one of these options, wherein Oracle Active Data Guard is used to migrate the Database. It's assumed that Oracle Data Guard (or Active Data Guard) is used for HA/DR purposes. Depending on the application, either the application can be migrated first or the database. In this case, the application is migrated to Azure using Azure Load Balancer. This enables you to split your traffic between on-premises and Azure, allowing you to gradually migrate your application tier. The database migration is performed in multiple

steps. As a first step, Oracle Data Guard is used to set up a Secondary/Standby Database in Azure. This allows you to migrate your data to Azure. Once the secondary in Azure is in-sync with the primary, you can flip the database in Azure to be your primary database while maintaining your secondary on-premises. As a next step, you may set up a secondary database in a different Availability Zone (or region) for HA/DR purposes. At this point, you can decommission your on-premises environment. All data traffic between on-premises and Azure flows over Azure ExpressRoute or Site-to-Site VPN connectivity.

## Potential use cases

This solution applies when migrating Oracle DB to Azure.

## Considerations

These considerations implement the pillars of the Azure Well-Architected Framework, which is a set of guiding tenets that can be used to improve the quality of a workload. For more information, see [Microsoft Azure Well-Architected Framework](#).

## Migration

You can migrate your entire Oracle database from on-premises to Azure VM with minimal downtime by using Oracle Recovery Manager (RMAN) and Oracle Data Guard. Use RMAN to restore your database to the target standby Azure VM, using either backup/restore or the duplicate database method. You can then configure the target database as a physical standby database with Oracle Data Guard, allowing all the transaction/redo data changes from the primary on-premises database to the standby database. When the primary on-premises Oracle database is in sync with the target standby database on the Azure VM instance, you can switch over to the target database, which will convert it to a read-write database. You can then point your application connections to the new primary database. This option provides a minimum downtime while migrating your database to Azure.

The Oracle Data Pump utility is used to export and import data and metadata from or to Oracle databases. You can run Data Pump export/import on an entire database, selective schemas, tablespaces, or database objects. Data Pump is the recommended tool for migrating data to Azure, for large databases that range from 10 GB to 20 TB in size. It allows a high degree of parallelism, flexible data extraction options, and scalable operations, which enable high-speed movement of data and metadata from a source database to the target database. Oracle Data Pump also supports encryption and

compression, when exporting your data to data dump files. You can use Oracle Data Pump with Oracle Data Guard or Golden Gate to handle the initial data transfer for large databases. Note that Data Pump is available only on Oracle Database 10g Release 1 (10.1) and later.

## Design considerations

### VM sizing

Consider using a hyperthreaded memory-optimized virtual machine with constrained core vCPUs for your Oracle Database VM, to save on licensing costs and to maximize performance. Oracle has guaranteed license mobility from on-premises to Azure. See the Oracle-Azure FAQ.

### Storage

Use multiple premium or ultra disks (managed disks) for performance and availability on your Oracle database. The disk/device name may change on reboots when using managed disks. It's recommended that you use the device UUID instead of the name, to ensure your mounts persist across reboots. Consider using Oracle Automatic Storage Management (ASM) for streamlined storage management for your database.

### Testing and tuning

We recommend the following tests to validate your application against your new Oracle database:

- Run performance tests to ensure that they meet your business expectations.
- Test database failover, recovery, and restoration to make sure that you're meeting RPO and RTO requirements.
- List all critical jobs and reports, and run them on new Oracle instance to evaluate their performance against your service-level agreements (SLAs).
- Finally, when migrating or creating applications for the cloud, it's important to tweak your application code to add cloud-native patterns such as retry pattern and circuit breaker pattern. Other patterns defined in the Cloud Design Patterns guide could help your application be more resilient.

### Oracle licensing

If you're using hyper-threading enabled technology in your Azure VMs, count two vCPUs as equivalent to one Oracle Processor license. See [Licensing Oracle Software in the Cloud Computing Environment](#) for details.

## Backup strategy

One backup strategy is to use Oracle Recovery Manager (RMAN) and Azure Backup for application-consistent backups. You can also use the Azure backup method.

Optionally use Azure Blob Fuse to mount a highly redundant Azure Blob Storage account and write your RMAN backups to it for added resiliency.

## Security

[Azure DDoS Protection](#), combined with application-design best practices, provides enhanced DDoS mitigation features to provide more defense against DDoS attacks. You should enable [Azure DDOS Protection](#) on any perimeter virtual network.

## Business continuity and disaster recovery

For business continuity and disaster recovery, consider deploying the following software:

- Oracle Data Guard Fast-Start Failover (FSFO) for database availability
- Oracle Data Guard Far Sync for zero data loss protection.
- Oracle GoldenGate for multi-primary or active-active mode on Azure availability set or availability zone depends on SLA requirements.

Use Availability Zones to achieve high availability in-region. For more information, see the [Reference architectures for Oracle databases on Azure](#).

An uptime availability of 99.99% for your database tier can be achieved by using a combination of Azure Availability Zones and Oracle Active DataGuard with FSFO.

Consider using proximity placement groups to reduce the latency between your application and database tier.

## Monitoring

Set up Oracle Enterprise Manager for management, monitoring, and logging.

## Next steps

Refer to these articles for supporting information:

- [Implement Oracle Data Guard on an Azure Linux virtual machine](#).
- [Implement Oracle Golden Gate on an Azure Linux VM](#).
- [Reference architectures for Oracle Database Enterprise Edition on Azure](#).

Learn more about the various architectural components:

- [Design and implement an Oracle database on Azure - Azure Virtual Machines](#)
- [Introduction to Oracle Data Guard ↗](#)
- [Oracle Data Guard Broker Concepts ↗](#)
- [Configuring Oracle GoldenGate for Active-Active High Availability ↗](#)
- [Oracle Active Data Guard Far Sync Zero Data Loss at Any Distance ↗](#)
- [Oracle Enterprise Manager ↗](#)
- [Azure Proximity Placement Groups](#)
- [Oracle Recovery Manager \(RMAN\) ↗](#)
- [Licensing Oracle Software in the cloud ↗](#)

## Related resources

- [Firewall and Application Gateway for virtual networks](#)
- [Virtual network integrated serverless microservices](#)
- [Choose between virtual network peering and VPN gateways](#)
- [Spoke-to-spoke networking](#)
- [Add IP address spaces to peered virtual networks](#)
- [Hub-spoke network topology in Azure](#)

# Oracle database migration decision process

Azure Database Migration service    Azure Managed Applications    Azure SQL Database  
Azure Virtual Machines

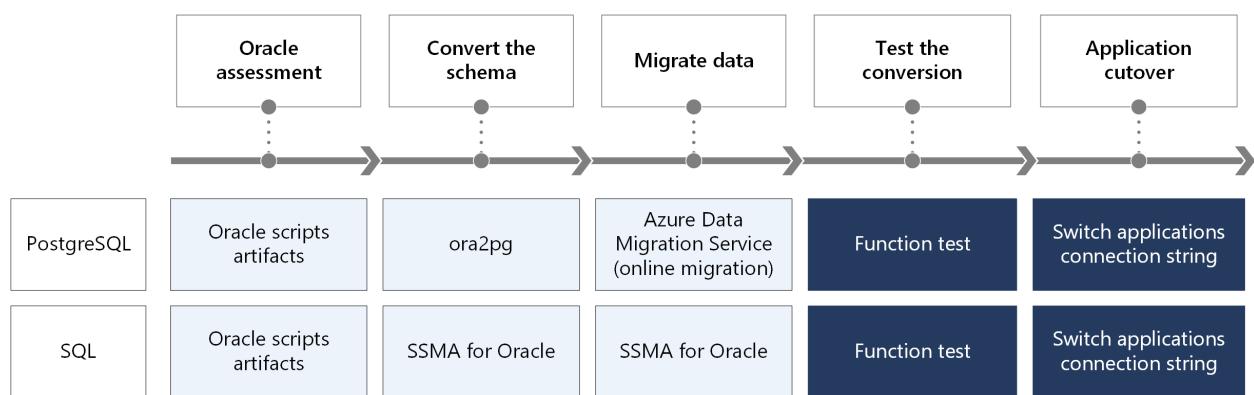
This series of articles provides a way for you to assess your current Oracle database environment, figure out your best migration path to Azure, and links to documents that help you make your migratory move. Your migration path can be to an Azure Virtual Machine (VM). It might also be to an Azure Managed Database that is running on an Azure VM.

To migrate an Oracle database to an Azure environment, you have to:

- Choose Azure resources as target database.
- Decide data migration method to evaluate downtime window.
- Figure out how to achieve business continuity and disaster recovery requirements.

## Architecture

This flow chart shows you the steps to move an Oracle database to either a PostgreSQL or a SQL database in Azure. The steps and the details are similar for both migration paths. Pay attention to the schema conversion and the data migration sections.



Download a [Visio file](#) of this architecture.

## Workflow

1. Use Oracle script artifacts to evaluate Oracle database.
2. Schema conversion is different for both database types:
  - PostgreSQL: Use ora2pg to convert your Oracle schema.
  - SQL: Use SQL Server Migration Assistant (SSMA) to convert your Oracle schema.
3. Data migration is different for both database types:
  - PostgreSQL: Use Azure Data Migration Service to migrate your Oracle data.
  - SQL: Use SSMA to migrate your Oracle data.
4. Test the conversion using functional tests.
5. Switch the application's connection strings to complete the application cutover.

## Oracle database discoveries

You can create and run scripts on your Oracle databases to evaluate how many tables, stored procedures, views, and packages exist in the environment. This table shows an example of the assessment principles:

[\[+\] Expand table](#)

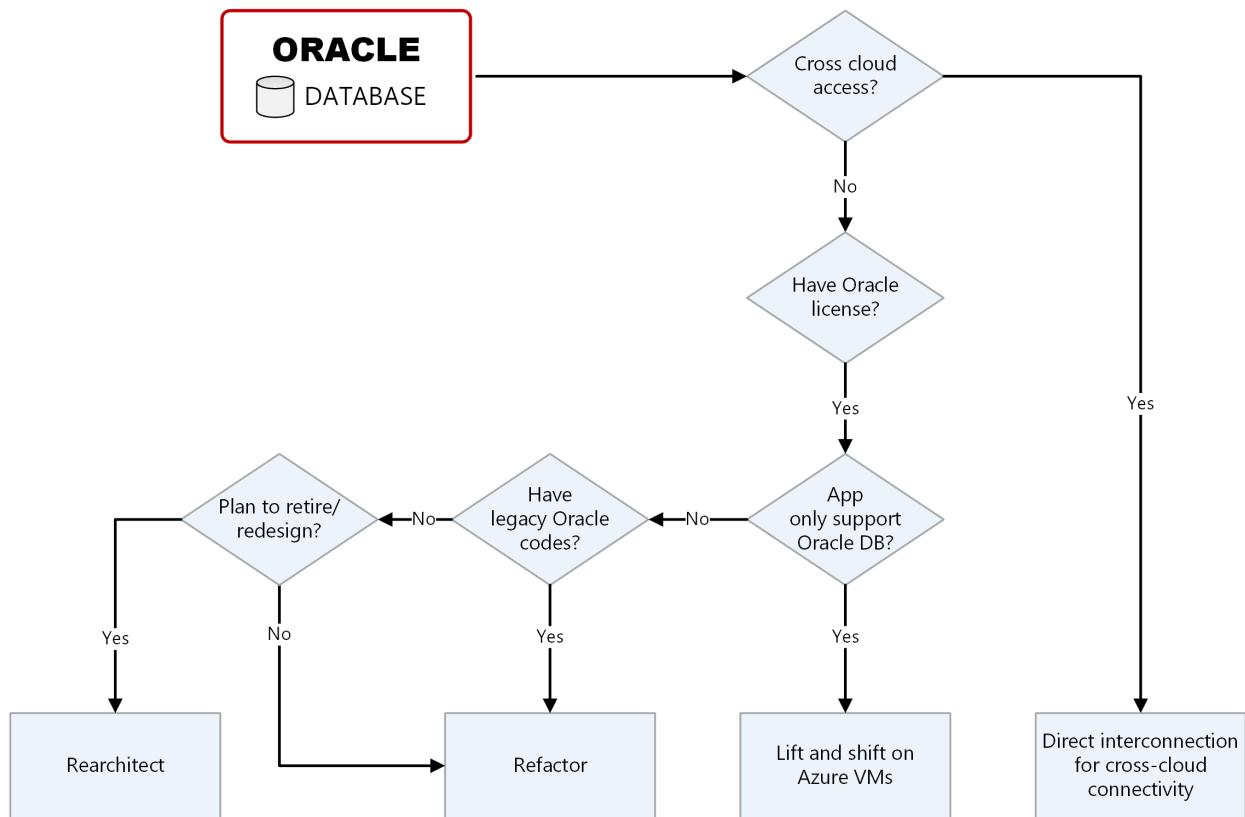
Category	Simple	Medium	Large	Complex	Custom
Number of tables in schema	<500	501-1000	1001-2000	2001-3000	>3000
Total number of SP, Trigger, Functions, Views	<100	101-200	201-400	401-800	>800
Collection Types per schema	<10	11-20	21-40	41-80	>80
Packages per schema	<10	11-25	26-50	51-100	>100

Category	Simple	Medium	Large	Complex	Custom
Schema Data Size	<10 GB	11-75 GB	76-500 GB	501-2000	>2000

Use the [Microsoft Assessment and Planning \(MAP\) Toolkit](#) to evaluate the existing Oracle database and schemas. For more information, refer to the [Oracle to SQL Server: Migration guide](#).

## Migration decision tree

The migration decision tree helps you find the appropriate path of your Oracle database migration.



## Contributors

This article is maintained by Microsoft. It was originally written by the following contributors.

Principal author:

- [Amber Zhao](#) | Principal Customer Engineer

*To see non-public LinkedIn profiles, sign in to LinkedIn.*

## Related resources

What you do next depends on where you wind up on the decision tree:

- **Cross-cloud connectivity:** If you already use Oracle Cloud Infrastructure (OCI), the easy migration path for you is direct interconnection between Azure and OCI. Go to [Oracle database migration: Cross-cloud Connectivity](#).
- **Lift and shift on Azure VMs:** You can deploy your Oracle databases in Azure based on a "bring your own license" model. Go to [Oracle database migration: Lift and Shift](#).
- **Refactor:** You have legacy Oracle code and you prefer using an Azure Managed Service. Go to [Oracle database migration: Refactor](#).
- **Rearchitect:** If you're planning to retire your old code and redesign your architecture, Azure SQL Database Managed Instance is a good option. Go to [Oracle database migration: Rearchitect](#).

# Oracle database migration: Cross-cloud connectivity

Azure ExpressRoute

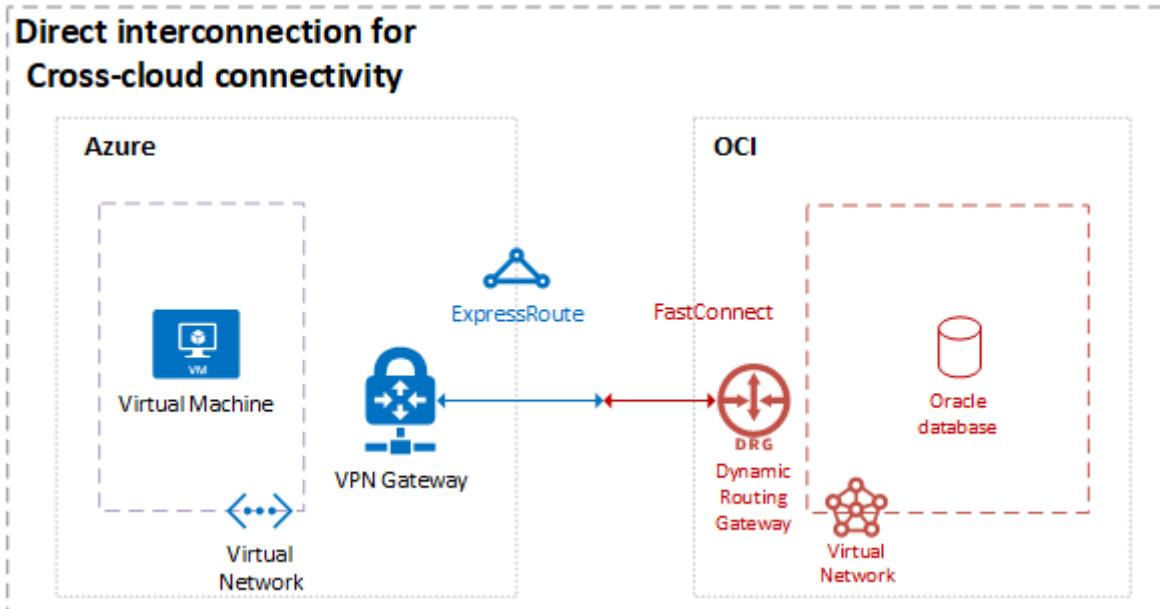
Azure Virtual Machines

Azure Virtual Network

Azure VPN Gateway

To support a [multicloud experience](#), you can create a direct interconnection between Azure and Oracle Cloud Infrastructure (OCI) by using [Azure ExpressRoute](#) and [FastConnect](#). The connection between the services allows applications hosted on Azure to communicate with Oracle database hosted on OCI. You can expect low latency and high throughput by connecting an ExpressRoute circuit in Azure with a FastConnect circuit in OCI.

## Architecture



## Workflow

1. Establish a connection between Azure ExpressRoute and OCI FastConnect.
2. Your Azure application can communicate with your OCI-hosted Oracle database.

## Components

- [Azure Virtual Machines](#) lets you migrate your business and important workloads to Azure to increase operational efficiencies.

- [Azure Virtual Network](#) is your private network in your Azure environment.
- [Azure VPN Gateway](#) connects your infrastructure to the cloud.
- [Azure ExpressRoute](#) creates a faster private connection to Azure.

## Contributors

*This article is maintained by Microsoft. It was originally written by the following contributors.*

Principal author:

- [Amber Zhao](#) | Principal Customer Engineer

*To see non-public LinkedIn profiles, sign in to LinkedIn.*

## Next steps

See [Set up a direct interconnection between Azure and Oracle Cloud Infrastructure](#) for step-by-step configuration instructions. Refer to the **Important** alert at the beginning of the article. It shows a list of Oracle applications that Oracle has certified to run in Azure when using the Azure/Oracle Cloud interconnect solution.

 **Note**

If this migration path doesn't seem like the right one for your business needs, refer back to the [Migration decision tree](#).

# Oracle database migration: Refactor

Azure Database for PostgreSQL

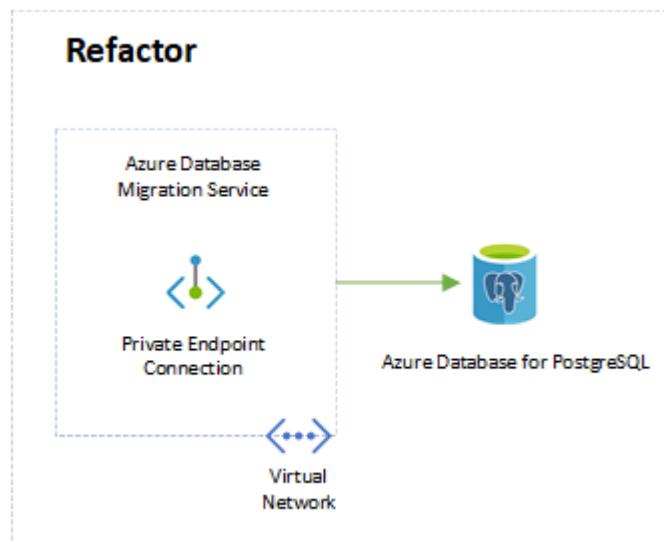
Azure Database Migration service

Azure Virtual Network

Do you have legacy Oracle code and prefer using a managed service on Azure? If so, you can use the Azure Database Migration Service to migrate your Oracle database to Azure Database for PostgreSQL. It's the option for you because it:

- Provides built-in [business continuity](#) and [disaster recovery](#) capacity.
- Allows you to [replicate data](#) from an Azure Database for PostgreSQL server to a read-only server.
- Lets you easily [migrate data to PostgreSQL online](#).

## Architecture



## Workflow

1. Use Azure Database Migration Service to automate your Oracle database migration to Azure.
2. Migrate the database to Azure Database for PostgreSQL.

## Components

- [Azure Database Migration Service](#) is a tool that helps you simplify, guide, and automate your database migration to Azure.
- [Azure Database for PostgreSQL](#) lets you focus on application innovation instead of database management and scale your workload quickly and easily.
- [Azure Virtual Network](#) is your private network in your Azure environment.

## Deploy this scenario

### Prerequisites

- Create an [Azure subscription](#).
- Create an instance of the [Azure Database Migration Service](#) by using the [Azure portal](#).

### Microsoft Assessment and Planning (MAP) Toolkit

Use the [Microsoft Assessment and Planning \(MAP\) Toolkit](#) to evaluate the existing Oracle database and schemas. For more information, refer to the [Oracle to SQL Server: Migration guide](#).

### Assess the migration complexity

Download [Ora2Pg](#). Run the following command to get the migration complexity assessment:

```
Console
```

```
ora2pg -t SHOW_REPORT --estimate_cost
```

Here's an example of the output of a schema assessment:

```
Console
```

```
Migration levels:
```

```
A - Migration that might be run automatically
```

```
B - Migration with code rewrite and a human-days cost up to 5 days
```

```
C - Migration with code rewrite and a human-days cost above 5 days
```

Technical levels:

- 1 = trivial: no stored functions and no triggers
- 2 = easy: no stored functions but with triggers, no manual rewriting
- 3 = simple: stored functions and/or triggers no manual rewriting
- 4 = manual: no stored functions but with triggers or views with code rewriting
- 5 = difficult: stored functions and/or triggers with code rewriting

## Oracle Objects conversion and Data migration

Use ora2pg to convert Oracle tables, stored procedures, packages, and other database objects. After the conversion, they'll be PostgreSQL-compatible. Next, start a migration pipeline in Azure Database Migration Service.

### Convert Oracle objects

Install Ora2Pg on an Azure Virtual Machine (VM). Refer to the [Step-by-Step Guide to install ora2pg on Linux & Windows](#).

Connect to Ora2pg to convert the schemas. Refer to the [Oracle to Azure Database for PostgreSQL Migration Cookbook](#).

### Migrate data online

You can migrate your data online to reduce downtime. Refer to [Create a DMS instance](#) for more info.

### Workaround list

Below you'll find a workaround list. It's useful when you're migrating Oracle database to PostgreSQL. Refer to the [Oracle migrate to PostgreSQL workaround list](#) to get detailed scripts.

[+] Expand table

Oracle	PostgreSGL
Database Link	Foreign Data Wrapper
External Table	Foreign Table
Synonym	View / Set search_path
Global Temporary Table	Unlogged Table / Temp Table
Virtual column	View / Function / Trigger
Connected by	With Recursive
Reverse Index	Functional Index
Index Organized Table (IOT)	Cluster the table according to an Index

## Contributors

*This article is maintained by Microsoft. It was originally written by the following contributors.*

Principal author:

- [Amber Zhao](#) | Principal Customer Engineer

*To see non-public LinkedIn profiles, sign in to LinkedIn.*

## Next steps

To begin migrating your Oracle database to Azure Database for PostgreSQL, see [Tutorial: Migrate Oracle to Azure Database for PostgreSQL online using DMS](#).

### Note

If this migration path doesn't seem like the right one for your business needs, refer back to the [Migration decision tree](#).

# Oracle database migration: Rearchitect

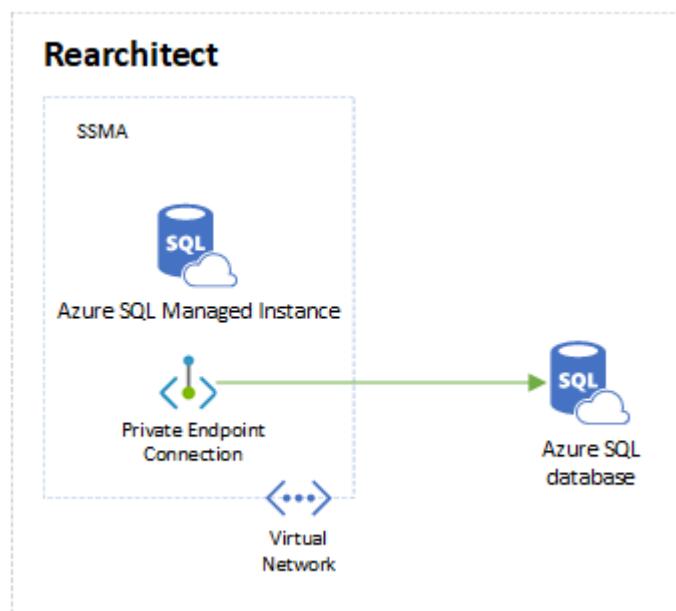
Azure SQL Managed Instance

Azure Virtual Machines

Are you comfortable working with Microsoft SQL Server? If so, you can use Azure SQL Managed Instance to rearchitect your database. It's a good option because it:

- Provides built-in [business continuity and disaster recovery capabilities](#).
- Offers [high level security and stability](#).
- Interfaces with [SQL Server Migration Assistant for Oracle \(SSMA\)](#). This tool allows for easy conversion of Oracle objects and migration of data to SQL Managed Instance.

## Architecture



## Workflow

1. Use SSMA to convert your Oracle schema to SQL schema.
2. Migrate the new schema to Azure SQL Managed Instance.
3. Connect the SQL Managed instance to your Azure SQL database.

# Components

- [Azure SQL Managed Instance](#) is the intelligent, scalable, cloud database service that combines the broadest SQL Server engine compatibility with all the benefits of a fully managed and evergreen platform as a service (PAAS).
- [Azure SQL](#) gives you a unified experience across your entire SQL portfolio and a full range of deployment options.
- [Azure Virtual Network](#) is your private network in your Azure environment.

## Deploy this scenario

### Evaluate your Oracle database

Use the [Microsoft Assessment and Planning \(MAP\) Toolkit](#) to evaluate the existing Oracle database and schemas. For more information, refer to the [Oracle to SQL Server: Migration guide](#).

### Oracle objects conversion results

Download SSMA and use it for Oracle schema and data migration: [Microsoft SQL Server Migration Assistant for Oracle](#).

This table shows the conversion results of Oracle objects to SQL Server objects carried out by SSMA.

[+] Expand table

Oracle objects	Resulting SQL Server objects
Functions	If the function can be directly converted to Transact-SQL, SSMA creates a function. In some cases, the function must be converted to a stored procedure. In this case, SSMA creates a stored procedure and a function that calls the stored procedure.
Procedures	If the procedure can be directly converted to Transact-SQL, SSMA creates a stored procedure. In some cases, a stored procedure must be called in an autonomous transaction. In this case, SSMA creates two stored

Oracle objects	Resulting SQL Server objects
	procedures: one that implements the procedure, and another that is used for calling the implementing stored procedure.
Packages	SSMA creates a set of stored procedures and functions that are unified by similar object names.
Sequences	SSMA creates sequence objects (SQL Server 2012 or SQL Server 2014) or emulates Oracle sequences.
Tables with dependent objects such as indexes and triggers	SSMA creates tables with dependent objects.
View with dependent objects, such as triggers	SSMA creates views with dependent objects.
Materialized Views	<p><b>SSMA creates indexed views on SQL server with some exceptions. Conversion will fail if the materialized view includes one or more of the following constructs:</b></p> <p>User-defined function</p> <p>Non-deterministic field / function / expression in SELECT, WHERE, or GROUP BY clauses</p> <p>Usage of Float column in SELECT*, WHERE, or GROUP BY clauses (special case of previous issue)</p> <p>Custom data type (incl. nested tables)</p> <p>COUNT(distinct &lt;field&gt;)</p> <p>FETCH</p> <p>OUTER joins (LEFT, RIGHT, or FULL)</p> <p>Subquery, other view</p> <p>OVER, RANK, LEAD, LOG</p> <p>MIN, MAX</p> <p>UNION, MINUS, INTERSECT</p> <p>HAVING</p>
Trigger	<b>SSMA creates triggers based on the following rules:</b>

Oracle objects	Resulting SQL Server objects
	<p>BEFORE triggers are converted to INSTEAD OF triggers.</p> <p>AFTER triggers are converted to AFTER triggers.</p> <p>INSTEAD OF triggers are converted to INSTEAD OF triggers. Multiple INSTEAD OF triggers defined on the same operation are combined into one trigger.</p> <p>Row-level triggers are emulated using cursors.</p> <p>Cascading triggers are converted into multiple individual triggers.</p>
Synonyms	<p><b>Synonyms are created for the following object types:</b></p> <ul style="list-style-type: none"> <li>Tables and object tables</li> <li>Views and object views</li> <li>Stored procedures</li> <li>Functions</li> </ul> <p><b>Synonyms for the following objects are resolved and replaced by direct object references:</b></p> <ul style="list-style-type: none"> <li>Sequences</li> <li>Packages</li> <li>Java class schema objects</li> <li>User-defined object types</li> <li>Synonyms for another synonym can't be migrated and will be marked as errors.</li> </ul> <p>Synonyms aren't created for Materialized views.</p>
User-Defined Types	<p><b>SSMA does not provide support for conversion of user-defined types. User-Defined Types, including its usage in PL/SQL programs are marked with special conversion errors guided by the following rules:</b></p> <p>Table column of a user-defined type is converted to VARCHAR(8000).</p> <p>Argument of user-defined type to a stored procedure or function is converted to VARCHAR(8000).</p> <p>Variable of user-defined type in PL/SQL block is converted to</p>

Oracle objects	Resulting SQL Server objects
	VARCHAR(8000).  Object Table is converted to a Standard table. Object view is converted to a Standard view.

For more info, see [Converting Oracle Schemas \(OracleToSQL\)](#).

## Convert Oracle objects conversion and migrate data

Once you've installed SSMA, create a report to convert Oracle schema and migrate the data to Azure SQL Managed Instance. For a step-by-step guide, see [Migrate an Oracle schema to SQL Server 2017 on Linux with the SQL Server Migration Assistant](#).

## Post-migration tasks

After the whole migration, uninstall the client components to remove the `ssma_oracle` schema.

### Note

Don't uninstall the extension pack from SQL Server unless your migrated database no longer uses functions in the `ssma_oracle` schema of the `sysdb` database.

For more information, see [Removing SSMA for Oracle Components](#).

## Contributors

*This article is maintained by Microsoft. It was originally written by the following contributors.*

Principal author:

- [Amber Zhao](#) | Principal Customer Engineer

*To see non-public LinkedIn profiles, sign in to LinkedIn.*

## Next steps

To begin migrating your Oracle database to SQL, see [SQL Server Migration Assistant for Oracle \(OracleToSQL\)](#).

 **Note**

If this migration path doesn't seem like the right one for your business needs, refer back to the [Migration decision tree](#).

# Oracle Database with Azure NetApp Files

Azure NetApp Files    Azure Virtual Machines    Azure Virtual Network

The most demanding Oracle Database workloads require very high I/O capacity. They also need low-latency access to storage. This document describes a high-bandwidth, low-latency solution for Oracle Database workloads.

The solution provides shared file access with the network file system (NFS) protocol. The architecture uses Azure NetApp Files, a shared file-storage service. Azure NetApp Files offers benefits:

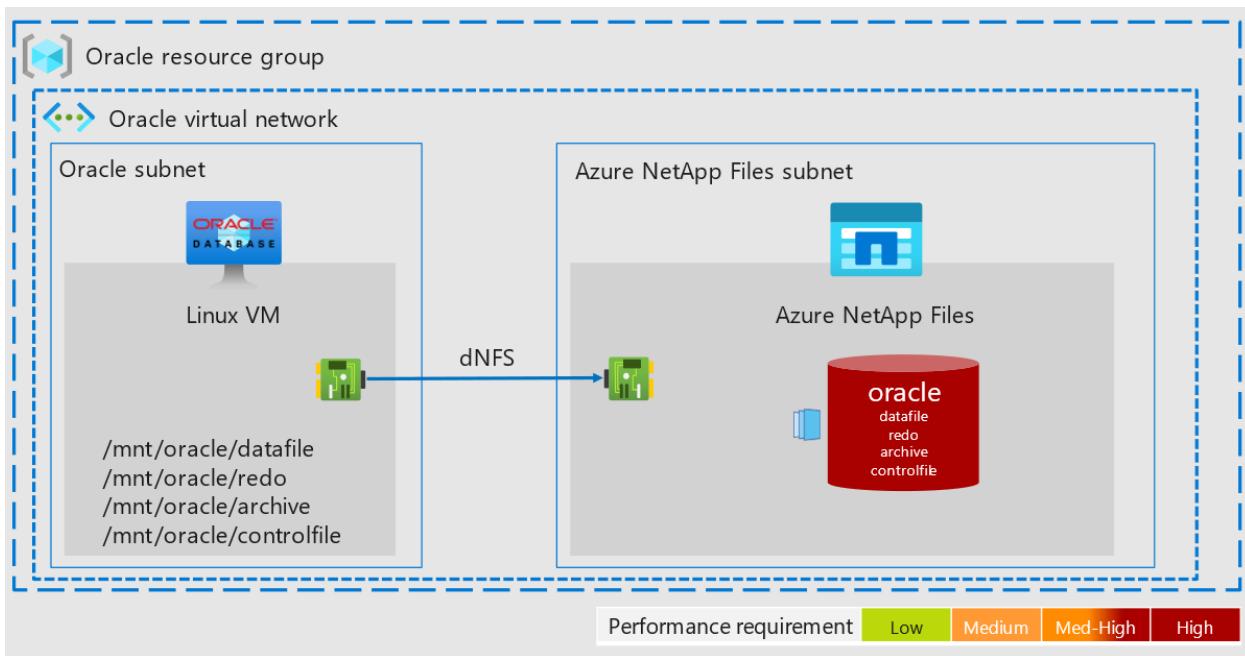
- Disk I/O limits on access rates that apply at the virtual machine (VM) level don't affect Azure NetApp Files. As a result, you can use smaller VMs than you would with disk storage without degrading performance. This approach significantly reduces costs.
- Azure NetApp Files offers flexibility. You can enlarge or reduce deployments on demand to make your configuration cost effective.

## Potential use cases

This solution has many uses:

- Running new Oracle Database instances that require high availability (HA) and have high standards for performance.
- Migrating highly performant, highly available Oracle Database instances from on-premises infrastructure to Azure Virtual Machines.
- Cloning enterprise-scale Oracle Database systems for use in test and development environments. The solution is particularly suited for cases that require advanced data management capabilities. It can help these cases meet aggressive data protection service level agreements (SLAs).
- Migrating Oracle Exadata systems to Azure.
- Implementing Oracle Pacemaker clusters that use NFS shared storage.
- Deploying SAP AnyDB, or Oracle 19c.

## Architecture



[Download an SVG of this architecture.](#)

The components interact in these ways:

- Oracle Database runs on Azure VMs within the Oracle subnet.
- In the Azure NetApp Files subnet, Azure NetApp Files provides NFS access to the data and log files.
- The connection protocol [Oracle Direct NFS \(dNFS\)](#) improves performance and throughput.

## Components

The solution uses the following components:

- [Azure NetApp Files](#) makes it easy to migrate and run file-based applications with no code changes. This shared file-storage service is a joint development from Microsoft and NetApp, a Microsoft partner.
- [Virtual Machines](#) is an infrastructure-as-a-service (IaaS) offer. You can use Virtual Machines to deploy on-demand, scalable computing resources. Virtual Machines provides the flexibility of virtualization but eliminates the maintenance demands of physical hardware. This solution uses [Linux VMs with Oracle Database software](#).
- [Azure Virtual Network](#) is a networking service that manages virtual private networks in Azure. Through Virtual Network, Azure resources like VMs can securely communicate with each other, the internet, and on-premises networks. An Azure virtual network is like a traditional network operating in a datacenter. But an Azure virtual network also provides scalability, availability, isolation, and other benefits of the Azure infrastructure.

- [Oracle Database](#) is a multi-model database management system. It supports various data types and workloads.
- The [dNFS](#) client optimizes I/O paths between Oracle and NFS servers. As a result, it provides better performance than traditional NFS clients.

## Alternatives

This solution uses Oracle Data Guard (ODG) for disaster recovery (DR), and snapshots for local replication. A few options exist, as the following sections explain.

### Cross-region replication

[Cross-region replication](#) provides efficient DR across regions in Azure. Cross-region replication uses storage-based replication. It doesn't use VM resources. For more information, see [Create volume replication for Azure NetApp Files](#).

### Availability sets and availability zones

ODG on Azure Virtual Machines functions like ODG in on-premises systems. But this product relies on its underlying architecture. If you run ODG on Azure VMs, consider also using one of these options to increase redundancy and availability:

- Place the Oracle VMs in the same availability set. This approach provides protection during these events:
  - Outages that equipment failures cause within a datacenter. VMs within an availability set don't share resources.
  - Updates. VMs within an availability set undergo updates at different times.
- Place the Oracle VMs in different availability zones. This approach provides protection against the failure of an entire datacenter. Each zone represents a set of datacenters within a region. If you place resources in different availability zones, datacenter-level outages can't take all your VMs offline.

You can only choose one of these options. An Azure VM can't participate in availability sets and zones at the same time. Each option has advantages:

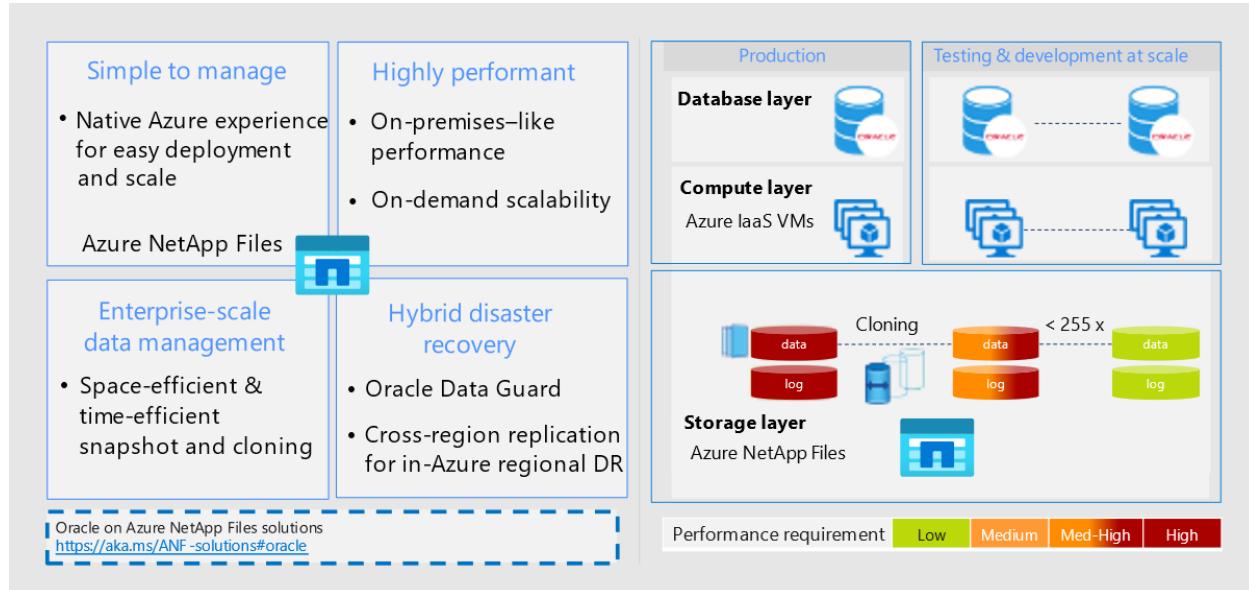
- Availability zones provide better availability than availability sets. See [SLA for Virtual Machines](#) for a comparison.
- You can place VMs that are in the same availability set in a [proximity placement group](#). This configuration minimizes the network latency between the VMs by guaranteeing that they're close to each other. In contrast, VMs that you place in different availability zones have greater network latency between them. It then

takes longer to synchronize data between the primary and secondary replicas. As a result, the primary replica may experience delays. There's also an increased chance of data loss during unplanned failovers.

After you choose a solution, test it under load. Ensure that it meets SLAs for performance and availability.

## Key benefits

This image shows the benefits of using Azure NetApp Files with Oracle Database.



Download an [SVG](#) of this architecture.

## Simple and reliable service

As a simple-to-consume Azure native service, Azure NetApp Files runs within the Azure datacenter environment. You can provision, consume, and scale Azure NetApp Files just like other Azure storage options. Azure NetApp Files uses reliability features that the NetApp data management software ONTAP provides. With this software, you can quickly and reliably provision enterprise-grade NFS volumes for Oracle Database and other enterprise application workloads.

## Highly performant systems

Azure NetApp Files uses a bare-metal fleet of all-flash storage. Besides using shared and highly scalable storage, Azure NetApp Files provides latencies of less than 1 millisecond. These factors make this service well-suited for using the NFS protocol to run Oracle Database workloads over networks.

The Azure DCsv2-series VMs can use high-performance, all-flash NetApp storage systems. These systems are also integrated into the Azure software-defined networking (SDN) and Azure Resource Manager frameworks. As a result, you get high-bandwidth, low-latency shared storage that's comparable to an on-premises solution. The performance of this architecture meets the requirements of the most demanding, business-critical enterprise workloads. For more information on the performance benefits of Azure NetApp Files, see [Benefits of using Azure NetApp Files with Oracle Database](#).

Azure NetApp Files offers on-demand scalability. You can enlarge or reduce deployments to optimize each workload's configuration.

## Enterprise-scale data management

This solution can handle workloads that require advanced data management features. ONTAP provides functionality in this area that's unmatched in the industry:

- Space-efficient, instantaneous cloning enhances development and test environments.
- On-demand capacity and performance scaling makes efficient use of resources.
- Snapshots provide database consistency points and offer these benefits:
  - They're storage efficient. You only need limited capacity to create snapshots.
  - You can quickly create, replicate, restore, or clone them. As a result, they provide backup and recovery solutions that achieve aggressive recovery time objective (RTO) and recovery point objective (RPO) SLAs.
  - They don't affect volume performance.
  - They provide scalability. You can create them frequently and store many simultaneously.

## Hybrid DR

The combination of ODG and Azure NetApp Files provides DR for this architecture. Those DR solutions are appropriate for cloud and hybrid systems. Their plans work across multiple regions and with on-premises datacenters.

## Considerations

The following considerations apply to this solution:

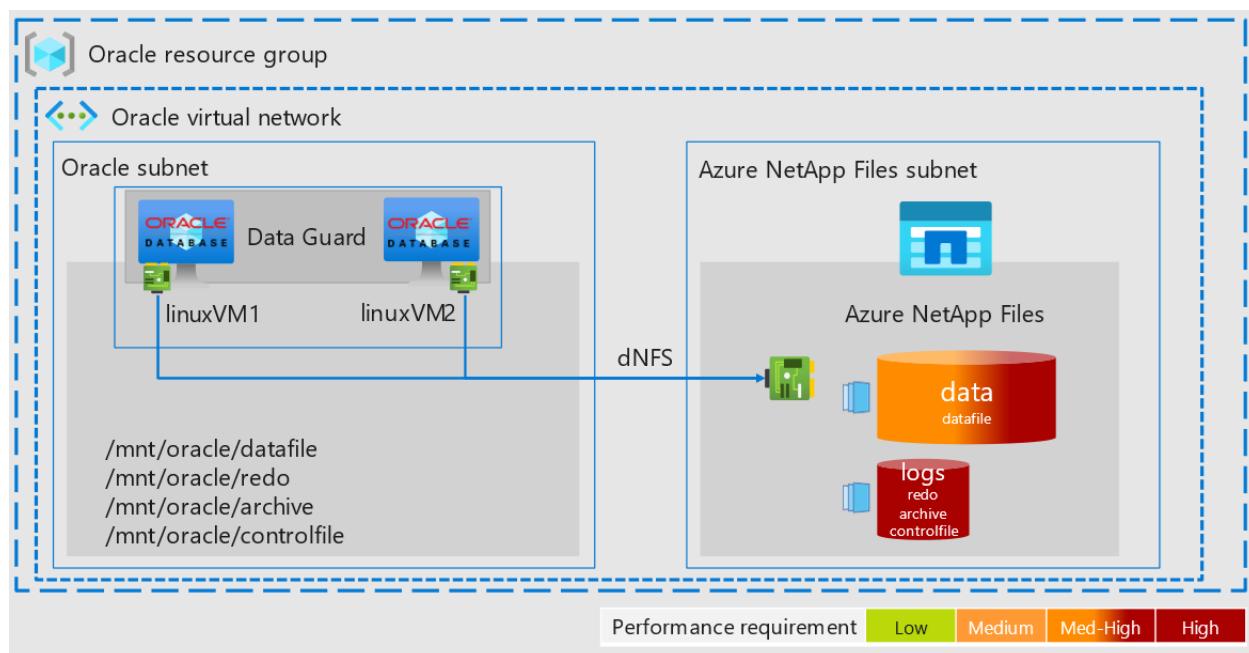
# Availability

For Azure NetApp Files:

- See [SLA for Azure NetApp Files](#) for this service's availability guarantee.
- As [Enterprise-scale data management](#) discusses, you can use snapshots in backup and recovery solutions. Use Oracle hot backup mode and Azure NetApp Files APIs to orchestrate database-consistent snapshots.

When you use Oracle Database in Azure, implement a solution for HA and DR to avoid downtime:

- Use [ODG](#).
- Run the database on one virtual machine.
- Deploy a secondary VM, but only install the binaries on it.
- Put both VMs in the same virtual network. Then they can access each other over the private persistent IP address.



*Download an [SVG](#) of this architecture.*

# Scalability

As [Highly performant systems](#) discusses, Azure NetApp Files provides built-in scalability.

# Security

Azure NetApp Files secures data in many ways. For information about inherent protection, encryption, policy rules, role-based access control features, and activity logs,

see [Security FAQs](#).

## Cost optimization

Using Azure NetApp Files instead of block storage can reduce costs:

- You can make the configuration cost-efficient. Traditional on-premises configurations are sized for maximum workload requirements. Consequently, these configurations are most cost-effective at maximum usage. In contrast, an Azure NetApp Files deployment is scalable. You can optimize the configuration for the current workload requirement to reduce expenses.
- You can use smaller VMs:
  - Azure NetApp Files provides low-latency storage access. With smaller VMs, you get the same performance that larger VMs deliver with ultra disk storage.
  - Cloud resources usually place limits on I/O operations. This practice prevents sudden slowdowns that resource exhaustion or unexpected outages can cause. As a result, VMs have disk throughput limitations and network bandwidth limitations. The network limitations are typically higher than disk throughput limitations. With network-attached storage, only network bandwidth limits are relevant, and they only apply to data egress. In other words, VM-level disk I/O limits don't affect Azure NetApp Files. Because of these factors, network-attached storage can achieve better performance than disk I/O. This fact is true even when Azure NetApp Files runs on smaller VMs.

Smaller VMs offer these pricing advantages over larger ones:

- They cost less.
- They carry a lower Oracle Database license cost, especially when you use smaller, constrained-code SKUs.
- The network-attached storage doesn't have an I/O cost component.

These factors make Azure NetApp Files less costly than disk storage solutions.

## Deploy this scenario

- For resources on deploying Oracle Database on Azure VMs with Azure NetApp Files, see [Solution architectures using Azure NetApp Files](#).
- For information on how to deploy and access Azure NetApp Files volumes, see [Azure NetApp Files documentation](#).
- Consider the database size:

- For small databases, you can deploy all components, such as data files, the redo log, the archive log, and control files, into a single volume. Such simplified configurations are easy to manage.
- For large databases, it's more efficient to configure multiple volumes. You can use [automatic or manual Quality of Service \(QoS\) volumes](#). These volume types provide more granular control over performance requirements.

## Contributors

*This article is maintained by Microsoft. It was originally written by the following contributors.*

Principal author:

- [Deanna Garcia](#) | Principal Program Manager

## Next steps

- [Oracle database performance on Azure NetApp Files single volumes](#)
- [Linux NFS mount options best practices for Azure NetApp Files](#)
- [Azure NetApp Files performance benchmarks for Linux](#)
- [Capacity management FAQs](#)

## Related resources

Fully deployable architectures that use Azure NetApp Files:

- [Run SAP BW/4HANA with Linux virtual machines on Azure](#)
- [Run SAP NetWeaver in Windows on Azure](#)

# Run Oracle databases on Azure

Azure Load Balancer

Azure Application Gateway

## 💡 Solution ideas

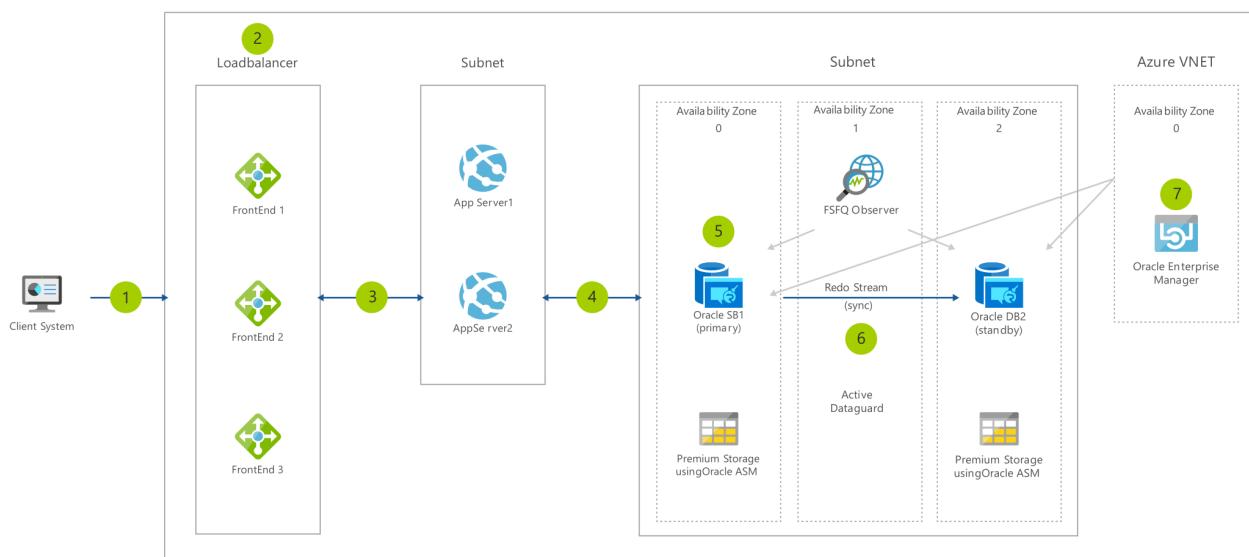
This article is a solution idea. If you'd like us to expand the content with more information, such as potential use cases, alternative services, implementation considerations, or pricing guidance, let us know by providing [GitHub feedback](#).

High availability for your front-end and middle tier can be obtained by using Azure Load Balancers or Application Gateways. An uptime availability of 99.99% for your database tier can be achieved using a combination of Azure Availability Zones and Oracle Active DataGuard with FSFO. For extra availability and/or Disaster Recovery, consider deploying another Database VM in a different Azure region and schedule frequent RMAN backups.

## Potential use cases

This solution idea illustrates a canonical architecture to achieve high availability for your Oracle Database Enterprise Edition in Azure.

## Architecture



Download an [SVG](#) of this architecture.

## Dataflow

1. The client system accesses a custom application with Oracle DB backend via the web.
2. Web front end is configured in a load balancer.
3. Web front end makes a call to the appropriate Application Server to handle the work.
4. Application server queries primary Oracle Database.
5. Oracle Database has been configured using a HyperThreaded Virtual Machine, with multiple Premium storage-based Managed Disks for performance and availability.
6. Oracle databases are replicated with Oracle DataGuard (or Active DataGuard) or Oracle GoldenGate for HA and DR purposes.
7. Oracle databases are monitored for uptime and performance by Oracle Enterprise Manager. OEM also allows you to generate various performance and usage reports.

## Components

Key technologies used to implement this architecture:

- [Application Gateway ↗](#)
- [Load balancing ↗](#)
- [Virtual Network ↗](#)

## Next steps

Product documentation:

- [Design and implement an Oracle database in Azure](#)
- [Overview of Oracle Applications and solutions on Azure](#)
- [What is Azure Application Gateway?](#)
- [What is Azure Load Balancer?](#)
- [What is Azure Virtual Network?](#)

Microsoft Learn modules:

- [Configure Azure Application Gateway](#)
- [Configure Azure Load Balancer](#)
- [Introduction to Azure Virtual Networks](#)

## Related resources

- [Architectures to deploy Oracle applications on Azure](#)

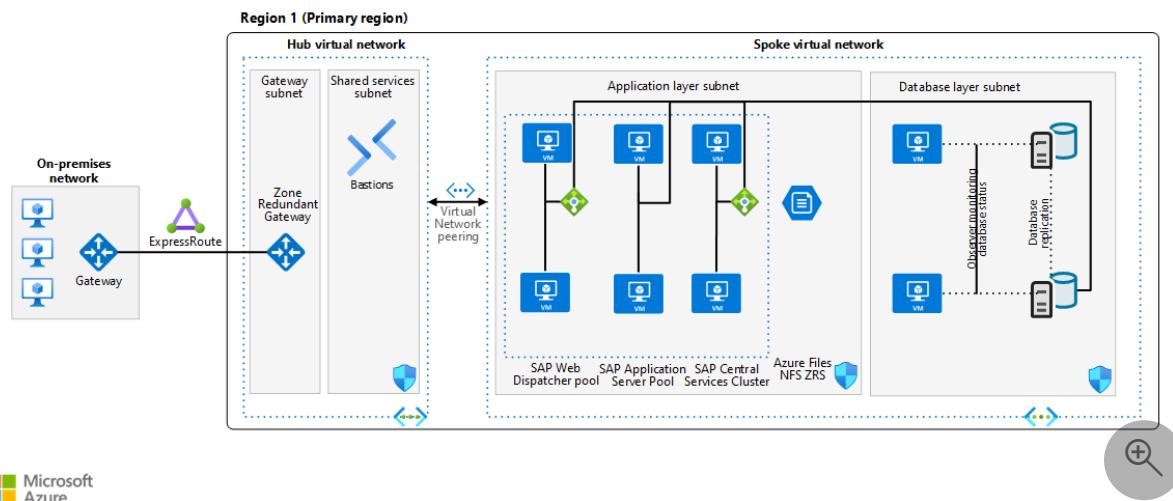
- Oracle application solutions integrating Microsoft Azure and Oracle Cloud Infrastructure
- Oracle Database with Azure NetApp Files
- Reference architectures for Oracle Database Enterprise Edition on Azure

# SAP deployment on Azure using an Oracle database

Azure ExpressRoute    SAP HANA on Azure Large Instances    Azure Virtual Machines    Azure Virtual Network  
Azure NetApp Files

This reference architecture shows a set of proven practices for running a high-availability SAP NetWeaver with Oracle Database on Azure. The architecture principles are OS-agnostic, however, unless otherwise specified, the architecture is assumed to be based on Linux.

The first diagram shows a reference architecture for SAP on Oracle in Azure. The architecture uses availability sets.



Download a [Visio file](#) of this architecture and related architectures.

## ⓘ Note

To deploy this reference architecture, you need the appropriate licensing of SAP products and other non-Microsoft technologies.

## Components

This reference architecture describes a typical SAP production system running on Oracle Database in Azure, in a highly available deployment to maximize system availability. The architecture and its components can be customized based on business requirements.

(RTO, RPO, uptime expectations, system role) and potentially reduced to a single VM. The network layout is simplified to demonstrate the architectural principals of such SAP environment and not intended to describe a full enterprise network.

## Networking

**Virtual networks.** The [Azure Virtual Network](#) service connects Azure resources to each other with enhanced security. In this architecture, the virtual network connects to an on-premises environment via a virtual private network (VPN) gateway deployed in the hub of a [hub-spoke topology](#). SAP applications and database are contained in their own spoke virtual network. The virtual networks are subdivided into separate subnets for each tier: application (SAP NetWeaver), the database, and shared services (like Azure Bastion).

This architecture subdivides the virtual network address space into subnets. Place application servers on a separate subnet and database servers on another. Doing so allows you to secure them more easily by managing the subnet security policies rather than the individual servers and cleanly separates security rules applicable to databases from application servers.

**Virtual network peering.** This architecture uses a hub-and-spoke networking topology with multiple virtual networks that are [peered together](#). This topology provides network segmentation and isolation for services deployed on Azure. Peering enables transparent connectivity between peered virtual networks through the Microsoft backbone network. It doesn't incur a performance penalty if deployed within a single region.

**Zone-redundant gateway.** A gateway connects distinct networks, extending your on-premises network to the Azure virtual network. We recommend that you use [ExpressRoute](#) to create private connections that don't go over the public internet, but you can also use a [site-to-site](#) connection. Azure ExpressRoute or VPN gateways can be deployed across zones to guard against zone failures. See [Zone-redundant virtual network gateways](#) to understand the differences between a zonal deployment and a zone-redundant deployment. It's worth mentioning here that the IP addresses used need to be of Standard SKU for a zone deployment of the gateways.

**Network security groups.** To restrict incoming, outgoing, and intra-subnet traffic in the virtual network, create [network security groups](#) which are in turn assigned to specific subnets. DB and application subnets are secured with workload specific NSGs.

**Application security groups.** To define fine-grained network security policies inside your NSGs based on workloads that are centered on applications, use [application security](#)

[groups](#) instead of explicit IP addresses. They let you group VMs by name and help you secure applications by filtering traffic from trusted segments of your network.

**Network interface cards (NICs).** Network interface cards enable all communication among virtual machines on a virtual network. Traditional on-premises SAP deployments implement multiple NICs per machine to segregate administrative traffic from business traffic.

On Azure, the virtual network is a software-defined network that sends all traffic through the same network fabric. So it's not necessary to use multiple NICs for performance reasons. But if your organization needs to segregate traffic, you can deploy multiple NICs per VM and connect each NIC to a different subnet. You can then use network security groups to enforce different access control policies on each subnet.

Azure NICs support multiple IPs. This support conforms with the SAP recommended practice of using virtual host names for installations. For a complete outline, see [SAP note 962955](#). (To access SAP notes, you need an SAP Service Marketplace account.)

## Virtual machines

This architecture uses virtual machines (VM). For SAP application tier, VMs are deployed for all instance roles - web dispatcher and application servers - both central services SAP (A)SCS and ERS as well as application servers (PAS, AAS). Adjust the number of virtual machines based on your requirements. The [Azure Virtual Machines planning and implementation guide](#) includes details about running SAP NetWeaver on virtual machines.

Similarly for all Oracle purposes virtual machines are used, both for the Oracle Database as well as Oracle observer VMs. Observer VMs in this architecture are smaller compared to actual database servers.

- **Constrained vCPU VMs.** In order to potentially save cost on Oracle licensing, consider utilizing [vCPU constrained VMs](#)
- **Certified VM families for SAP.** For details about SAP support for Azure virtual machine types and throughput metrics (SAPS), see [SAP note 1928533](#). (To access SAP notes, you need an SAP Service Marketplace account.)

**Proximity Placement Groups (PPG).** To optimize network latency, you can use [proximity placement groups](#), which favor collocation, meaning that virtual machines are in the same datacenter to minimize application latency. They can greatly improve the user experience for most SAP applications. Due to potential restrictions with PPGs, adding the database AvSet to the SAP system's PPG should be done sparsely and only when

required for latency between SAP application and database traffic. For more details on the usage scenarios for PPGs see the linked documentation.

**Generation 2 (Gen2) virtual machines.** Azure offers the choice when deploying VMs if they should be generation 1 or 2. [Generation 2 VMs](#) support key features which are not available for generation 1 VMs. Particularly for very large Oracle databases this is of importance since some VM families such as [Mv2](#) or [MdsV2](#) are **only** supported as Gen2 VMs. Similarly, SAP on Azure certification for some newer VMs might require them to be only Gen2 for full support, even if Azure allows both on them. See details in [SAP Note 1928533 - SAP Applications on Microsoft Azure: Supported Products and Azure VM types ↗](#).

Since all other VMs supporting SAP allow the choice of either Gen2 only or Gen1+2 selectively, it is recommended to deploy all SAP VMs as Gen2, even if the memory requirements are very low. Even the smallest VMs once deployed as Gen2 can be scaled up to the largest available with a simple deallocate and resize. Gen1 VMs can only be resized to VM families allowed to run Gen1 VMs.

## Storage

This architecture uses [Azure managed disks](#) for virtual machines and [Azure Files NFS](#) or [Azure NetApp Files](#) for any NFS shared storage requirements such as sapmnt and SAP transport NFS volumes. Guidelines for storage deployment with SAP on Azure are in detail within the [Azure Storage types for SAP workload guide](#)

- **Certified storage for SAP.** Similar to certified VM types for SAP usage, see the details in [SAP note 2015553 ↗](#) and [SAP note 2039619 ↗](#).
- **Storage design for SAP on Oracle.** You can find a recommended storage design for SAP on Oracle in Azure in [Azure Virtual Machines Oracle DBMS deployment for SAP workload](#). This article provides specific guidance on file system layout, disk sizing recommendations, and other storage options.
- **Storing Oracle Database files.** On Linux ext4 or xfs filesystems need to be used for database, NTFS for Windows deployments. [Oracle ASM](#) is also supported for Oracle deployments for Oracle Database 12c Release 2 and higher.
- **Alternatives to managed disks.** You can alternatively use [Azure NetApp Files](#) for the Oracle database. For more information, see [SAP note 2039619 ↗](#) and the [Oracle on Azure](#) documentation. [Azure Files NFS](#) volumes are not intended for storing Oracle Database files, unlike Azure NetApp Files.
- **Azure Premium SSD v2** is designed for performance-critical workloads like SAP. See [Deploy a Premium SSD v2](#) for this storage solution's benefits and its current limitations.

# High availability

The preceding architecture depicts a highly available deployment, with each application layer contained on two or more virtual machines. The following components are used.

On Azure, SAP workload deployment can be either regional or zonal, depending on the availability and resiliency requirements of the SAP applications. Azure provides [different deployment options](#), like Virtual Machine Scale Sets with flexible orchestration (FD=1), availability zones, and availability sets to increase the availability of resources. To get a comprehensive understanding of the deployment options and their applicability across different Azure regions (including across zones, within a single zone, or in a region without zones), see [High-availability architecture and scenarios for SAP NetWeaver](#).

**Load balancers.** [Azure Load Balancer](#) is used to distribute traffic to virtual machines in the SAP subnets. When you incorporate Azure Load Balancer in a zonal deployment of SAP, be sure to select the Standard SKU load balancer. The Basic SKU balancer doesn't support zonal redundancy.

Consider the [decision factors](#) when deploying VMs between availability zones for SAP. Use of [proximity placement groups](#) with an availability zone deployment needs to be evaluated and used only for application tier VMs.

## ⓘ Note

Availability Zones support intra-region high availability, but they aren't effective for DR. The distances between zones are too short. Typical DR regions should be at least 100 miles away from the primary region.

**Oracle-specific components.** Oracle Database VMs are deployed in an availability set or in different availability zones. Each VM contains its own installation of the database software and VM-local database storage. Set up synchronous database replication through Oracle Data Guard between the databases to ensure consistency and allow low RTO & RPO service times in case of individual failures. Besides the database VMs, additional VMs with Oracle Data Guard Observer are needed for an Oracle Data Guard Fast-Start Failover setup. The Oracle observer VMs monitor the database and replication status and facilitate database failover in an automated way, without the need for any cluster manager. Database replication management can be performed then using Oracle Data Guard Broker for ease of use.

For details on Oracle Data Guard deployment see

- [SAP whitepaper - Setting up Oracle 12c Data Guard for SAP Customers ↗](#)

- Oracle Data Guard documentation on Azure

This architecture utilizes native Oracle tooling without any actual cluster setup or the need for a load balancer in the database tier. With Oracle Data Guard Fast-Start Failover and SAP configuration, the failover process is automated and SAP applications re-connect to the new primary database should a failover occur. Various 3rd party cluster solutions exist as an alternative, such as SIOS Protection Suite or Veritas InfoScale, details of which deployment can be found in respective 3rd party vendor's documentation respectively.

**Oracle RAC.** Oracle Real Application Cluster (RAC) is currently [not certified or supported by Oracle in Azure](#). However Oracle Data Guard technologies and architecture for high-availability can provide highly resilient SAP environments with protection against rack, data center, or regional interruptions of service.

**NFS tier.** For all highly available SAP deployments, a resilient NFS tier is required to be used, providing NFS volumes for SAP transport directory, sapmnt volume for SAP binaries as well as further volumes for (A)SCS and ERS instances. Options to provide an NFS tier are

- [Azure Files NFS](#) with zonal redundant storage (ZRS) - [SLES](#) and [RHEL](#) guides
- [Azure NetApp Files](#) deployment of NFS volumes - [SLES](#) and [RHEL](#) guides
- VM based NFS cluster - two additional VMs with local storage, replicated between VMs with DRBD (Distributed Replicated Block Device) - [SLES](#) and [RHEL](#) guides

**SAP central services cluster.** This reference architecture runs Central Services on discrete VMs. Central Services is a potential single point of failure (SPOF) when it's deployed to a single VM. To implement a highly available solution, cluster management software is needed which automates failover of (A)SCS and ERS instances to the respective VM. As this is tied strongly with the chosen NFS solution

Chosen cluster solution requires a mechanism to decide in case of software or infrastructure unavailability which VM should serve the respective service(s). With SAP on Azure, two options are available for Linux based implementation of STONITH - how to deal with unresponsive VM or application

- ***SUSE-Linux-only SBD (STONITH Block Device)*** - using one or three additional VMs which serve as iSCSI exports of a small block device, which is accessed regularly by the actual cluster member VMs, two (A)SCS/ERS VMs in this cluster pool. The VMs use these SBD mounts to cast votes and thus achieve quorum for cluster decisions. The architecture contained on this page does NOT contain the 1 or 3 additional SBD VMs. RedHat does not support any SBD implementations in Azure and thus this option is only available to SUSE SLES operating system.

- **Azure Fence Agent.** Without utilizing additional VMs, Azure management API is used for regular checks for VM availability.

Guides linked within the NFS tier section contain the necessary steps and design for respective cluster choice. Third party Azure certified cluster managers can be also utilized to provide high-availability of the SAP central services.

**SAP application servers pool.** Two or more application servers where high availability is achieved by load-balancing requests through SAP message server or web dispatchers. Each application server is independent and there is no network load balancing required for this pool of VMs.

**SAP web dispatcher pool.** The Web Dispatcher component is used as a load balancer for SAP traffic among the SAP application servers. To achieve [high availability of the SAP Web Dispatcher](#), Azure Load Balancer implements either the failover cluster or the parallel Web Dispatcher setup.

[Embedded Web Dispatcher](#) on (A)SCS is a special option. You should take into account proper sizing because of additional workload on (A)SCS.

For internet-facing communications, we recommend a stand-alone solution in the perimeter network (also known as *DMZ*) to satisfy security concerns.

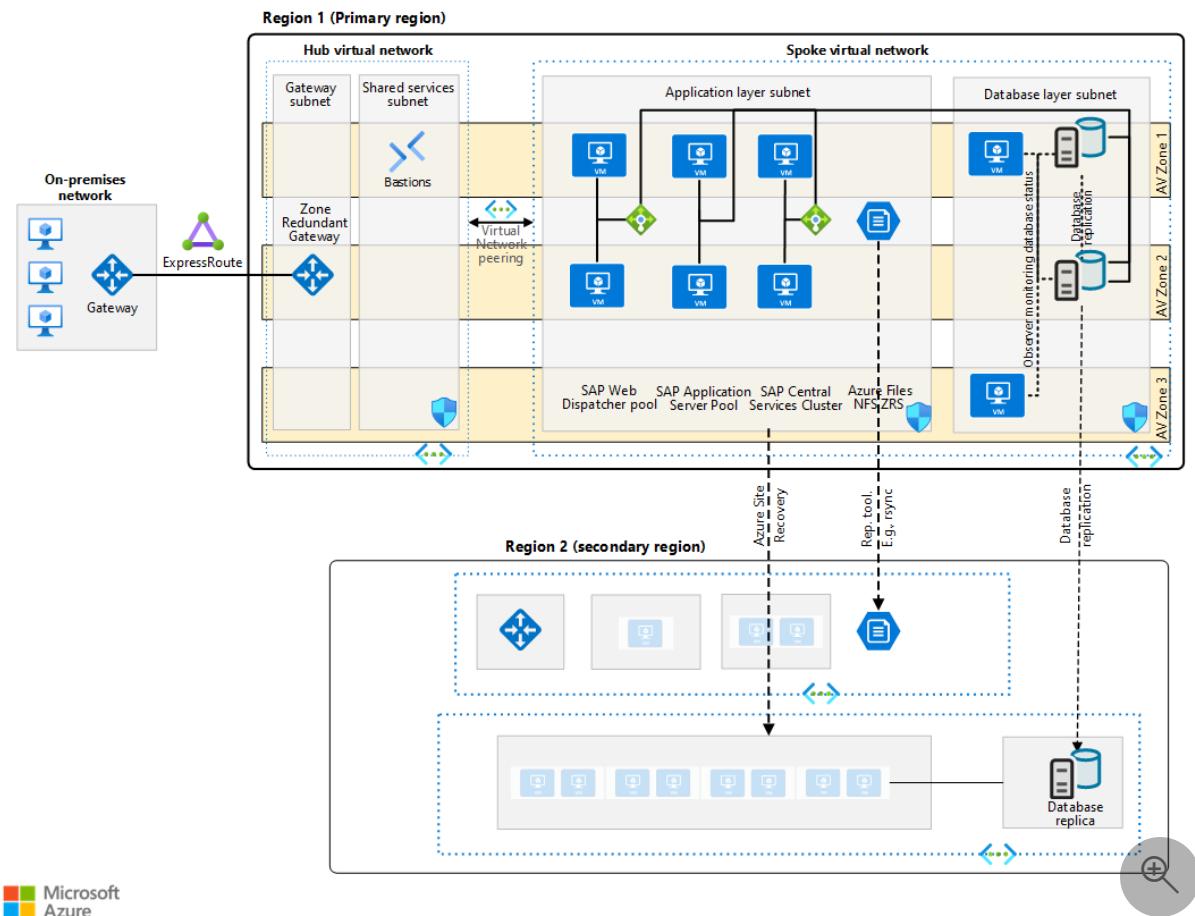
**Windows deployments.** This document, as prefaced in beginning, is focused primarily with Linux based deployments. For usage with Windows, same architectural principles apply and there are no architectural differences with Oracle between Linux and Windows.

For SAP application part, see the details in architecture guide [Run SAP NetWeaver in Windows on Azure](#).

## Considerations

### Disaster recovery

The following diagram shows the architecture of a production SAP system on Oracle in Azure. The architecture provides DR and uses availability zones.



[Download a Visio file](#) of this architecture and related architectures.

Every architectural layer in the SAP application stack uses a different approach to provide DR protection. For DR strategies and implementation details, see [Disaster recovery overview and infrastructure guidelines for SAP workload](#) and [Disaster recovery guidelines for SAP application](#).

## Backup

Backup for Oracle in Azure can be achieved through several means:

- **Azure Backup.** Azure provided and maintained scripts for Oracle Databases, to facilitate Oracle actions pre- and post backup execution.
- **Azure Storage.** Using file-based database backups, for example scheduled with SAP's BR tools, to be stored and versioned as files/directories on Azure Blob NFS, Azure Blob, or Azure Files storage services. See [documented details](#) how to achieve both Oracle data and log backups.
- **3rd party backup solutions.** See architecture of your backup storage provider, supporting Oracle in Azure.

For non-database VMs, [Azure Backup for VM](#) is recommended to protect SAP application VMs and surround infrastructure like SAP Web Dispatcher.

# Contributors

*This article is maintained by Microsoft. It was originally written by the following contributors.*

Principal author:

- [Robert Biro](#) ↗ | Senior Architect

## Next steps

- [High availability for SAP NetWeaver on Azure VMs](#)
- [Azure Virtual Machines planning and implementation for SAP NetWeaver](#)
- [Use Azure to host and run SAP workload scenarios](#)

## Communities

Communities can answer questions and help you set up a successful deployment.

Consider these resources:

- [Running SAP Applications on the Microsoft Platform blog](#) ↗
- [Azure Community Support](#) ↗
- [SAP Community](#) ↗
- [Stack Overflow for SAP](#) ↗

## Related resources

See these articles for more information and for examples of SAP workloads that use some of the same technologies:

- [SAP NetWeaver on SQL Server](#)
- [Run SAP NetWeaver in Windows on Azure](#)
- [Dev/test environments for SAP workloads on Azure](#)

# SAP on Azure architecture center

Azure

The SAP on Azure architecture center provides architectural guidance for SAP workloads running on Azure. The architectural content provides implementation, configuration, and deployment instructions for SAP platforms, workloads, and applications. The guidance applies the principles of the Cloud Adoption Framework (CAF) and the Well-Architected Framework (WAF) to specific SAP use cases. It assumes a familiarity with these principles. We also have guidance specific to SAP that covers the entire cloud adoption journey from migration to workload operations.

## SAP cloud adoption guidance

We recommend using the SAP cloud adoption guidance as you start your cloud adoption journey with SAP. The content outlines processes for preparing, migrating, and modernizing. It outlines integrating an SAP platform into your cloud adoption efforts.

[Cloud Adoption Framework - SAP scenario](#)

## SAP workload best practices

The SAP workload guidance designed to be applied to a single SAP workload. This content gives architecture design recommendations that you should follow throughout the lifecycle of an SAP workload. It encourages the use of assessments and health checks to inform your design decisions and continuously align them to best practices.

[Well-Architected Framework - SAP workload](#)

## SAP architecture guidance

The Azure architecture center has guides, architectures, and solution ideas that provide several use cases that you can use as reference as you build your SAP workload.

Architectures give detailed best practices on how to design SAP platforms, workloads, and applications in Azure.

[Architectures](#)

The guides provide general checklists, configurations, and design guidance. Use these guides to make design decisions for forming the best approach for your SAP solution.

## Guides

Solution ideas are conceptual. They're meant to inspire new ideas and novel approaches to architecture. The goal is to help you extend your SAP solutions to derive more benefits.

## Solution ideas

# RISE with SAP

Integrating SAP RISE environment with Azure resources. Under RISE with SAP solution, the SAP workload is managed by SAP in the partner's own Azure subscription and tenant. Guides to enable use of customer managed Azure services to extend the SAP environment.

## Integrating Azure with SAP RISE

# SAP workloads on Azure: planning and deployment checklist

Article • 06/14/2023

This checklist is designed for customers moving SAP applications to Azure infrastructure as a service. SAP applications in this document represent SAP products running the SAP kernel, including SAP NetWeaver, S/4HANA, BW and BW/4 and others. Throughout the duration of the project, a customer and/or SAP partner should review the checklist. It's important to note that many of the checks are completed at the beginning of the project and during the planning phase. After the deployment is done, straightforward changes on deployed Azure infrastructure or SAP software releases can become complex.

Review the checklist at key milestones during your project. Doing so will enable you to detect small problems before they become large problems. You'll also have enough time to re-engineer and test any necessary changes. Don't consider this checklist complete. Depending on your situation, you might need to perform additional more checks.

The checklist doesn't include tasks that are independent of Azure. For example, SAP application interfaces change during a move to the Azure platform or to a hosting provider. SAP documentation and support notes will also contain further tasks, which are not Azure specific but need to be part of your overall planning checklist.

This checklist can also be used for systems that are already deployed. New features or changed recommendations might apply to your environment. It's useful to review the checklist periodically to ensure you're aware of new features in the Azure platform.

Main content in this document is organized in tabs, in a typical project's chronological order. See content of each tab and consider each next tab to build on top of actions done and learnings obtained in the previous phase. For production migration, the content of **all** tabs should be considered and not just production tab only. To help you map typical project phases with the phase definition used in this article, consult the below table.

Deployment checklist phases	Example project phases or milestones
Preparation and planning phase	Project kick-off / design and definition phase
Pilot phase	Early validation / proof of concept / pilot

<b>Deployment checklist phases</b>	<b>Example project phases or milestones</b>
Non-production phase	Completion of the detailed design / non-production environment builds / testing phase
Production preparation phase	Dress rehearsal / user acceptance testing / mock cut-over / go-live checks
Go-live phase	Production cut-over and go-live
Post-production phase	Hypercare / transition to business as usual

## Planning phase

# Project preparation and planning phase

During this phase, you plan the migration of your SAP workload to the Azure platform. Documents such as [planning guide for SAP](#) in Azure and [Cloud Adoption Framework for SAP](#) cover many topics and help as information in your preparation. At a minimum, during this phase you need to create the following documents, define, and discuss the following elements of the migration:

## High-level design document

This document should contain:

- The current inventory of SAP components and applications, and a target application inventory for Azure.
- A responsibility assignment matrix (RACI) that defines the responsibilities and assignments of the parties involved. Start at a high level, and work to more granular levels throughout planning and the first deployments.
- A high-level solution architecture. Best practices and example architectures from [Azure Architecture Center](#) should be consulted.
- A decision about which Azure regions to deploy to. See the [list of Azure regions](#), and [list of regions with availability zone support](#). To learn which services are available in each region, see [products available by region](#).
- A networking architecture to connect from on-premises to Azure. Start to familiarize yourself with the [Azure enterprise scale landing zone](#) concept.
- Security principles for running high-impact business data in Azure. To learn about data security, start with the Azure security documentation.

- Storage strategy to cover block devices (Managed Disk) and shared filesystems (such as Azure Files or Azure NetApp Files) that should be further refined to file-system sizes and layouts in the technical design document.

## Technical design document

This document should contain:

- A block diagram for the solution showing the SAP and non-SAP applications and services
- An [SAP Quicksizer project](#) based on business document volumes. The output of the Quicksizer is then mapped to compute, storage, and networking components in Azure. Alternatively to SAP Quicksizer, diligent sizing based on current workload of source SAP systems. Taking into account the available information, such as DBMS workload reports, SAP EarlyWatch Reports, compute and storage performance indicators.
- Business continuity and disaster recovery architecture.
- Detailed information about OS, DB, kernel, and SAP support pack versions. It's not necessarily true that every OS release supported by SAP NetWeaver or S/4HANA is supported on Azure VMs. The same is true for DBMS releases. Check the following sources to align and if necessary, upgrade SAP releases, DBMS releases, and OS releases to ensure SAP and Azure support. You need to have release combinations supported by SAP and Azure to get full support from SAP and Microsoft. If necessary, you need to plan for upgrading some software components. More details on supported SAP, OS, and DBMS software are documented here:
  - [What SAP software is supported for Azure deployments](#)
  - [SAP note 1928533 - SAP Applications on Microsoft Azure: Supported Products and Azure VM types](#). This note defines the minimum OS and DBMS releases supported on Azure VMs. Note also provides the SAP sizing for SAP-supported Azure VMs.
  - [SAP note 2015553 - SAP on Microsoft Azure: Support prerequisites](#). This note defines prerequisites around Azure storage, networking, monitoring, and support relationship needed with Microsoft.
  - [SAP note 2039619](#). This note defines the Oracle support matrix for Azure. Oracle supports only Windows and Oracle Linux as guest operating systems on Azure for SAP workloads. This support statement also applies for the SAP application layer that runs SAP instances, as long they contain Oracle Client.
  - SAP HANA-supported Azure VMs are listed on the [SAP website](#). Details for each entry contain specifics and requirements, including supported OS

version. This might not match latest OS version as per [SAP note 2235581](#).

- [SAP Product Availability Matrix](#).

Further included in same technical document(s) should be:

- Storage Architecture high level decisions based on [Azure storage types for SAP workload](#)
  - Managed Disks attached to each VM
  - Filesystem layouts and sizing
  - SMB and/or NFS volume layout and sizes, mount points where applicable
- High availability, backup and disaster recovery architecture
  - Based on RTO and RPO, define what the high availability and disaster recovery architecture needs to look like.
  - Understand the use of [different deployment types](#) for optimal protection.
  - Considerations for Azure Virtual Machines DBMS deployment for SAP workloads and related documents. In Azure, using a shared disk configuration for the DBMS layer as, for example, [described for SQL Server](#), isn't supported. Instead, use solutions like:
    - [SQL Server Always On](#)
    - [HANA System Replication](#)
    - [Oracle Data Guard](#)
    - [IBM Db2 HADR](#)
  - For disaster recovery across Azure regions, review the solutions offered by different DBMS vendors. Most of them support asynchronous replication or log shipping.
  - For the SAP application layer, determine whether you'll run your business regression test systems, which ideally are replicas of your production deployments, in the same Azure region or in your DR region. In the second case, you can target that business regression system as the DR target for your production deployments.
  - Look into Azure Site Recovery as a method for replicating the SAP application layer into the Azure DR region. For more information, see a [set-up disaster recovery for a multi-tier SAP NetWeaver app deployment](#).
  - For projects required to remain in a single region for compliance reasons, consider a combined HADR configuration by using [Azure Availability Zones](#).
- An inventory of all SAP interfaces and the connected systems (SAP and non-SAP).
- Design of foundation services. This design should include the following items, many of which are covered by the [landing zone accelerator for SAP](#):
  - Network topology within Azure and assignment of different SAP environment

- Active Directory and DNS design.
- Identity management solution for both end users and administration
- [Azure role-based access control \(Azure RBAC\)](#) structure for teams that manage infrastructure and SAP applications in Azure.
- Azure resource naming strategy
- Security operations for Azure resources and workloads within
- Security concept for protecting your SAP workload. This should include all aspects – networking and perimeter monitoring, application and database security, operating systems securing, and any infrastructure measures required, such as encryption. Identify the requirements with your compliance and security teams.
- Microsoft recommends either Professional Direct, Premier or Unified Support contract. Identify your escalation paths and contacts for support with Microsoft. For SAP support requirements, see [SAP note 2015553](#).
- The number of Azure subscriptions and core quota for the subscriptions. [Open support requests to increase quotas of Azure subscriptions](#) as needed.
- Data reduction and data migration plan for migrating SAP data into Azure. For SAP NetWeaver systems, SAP has guidelines on how to limit the volume of large amounts of data. See [this SAP guide](#) about data management in SAP ERP systems. Some of the content also applies to NetWeaver and S/4HANA systems in general.
- An automated deployment approach. Many customers start with scripts, using a combination of PowerShell, CLI, Ansible and Terraform. Microsoft developed solutions for SAP deployment automation are:
  - [Azure Center for SAP solutions](#) – Azure service to deploy and operate a SAP system's infrastructure
  - [SAP on Azure Deployment Automation](#), an open-source orchestration tool for deploying and maintaining SAP environments

 **Note**

Define a regular design and deployment review cadence between you as the customer, the system integrator, Microsoft, and other involved parties.

## Automated checks and insights in SAP landscape

Several of the checks above are checked in automated way with [SAP on Azure Quality Check Tool](#). These checks can be executed automated with the provided open-source project. While no automatic remediation of issues found is performed, the tool will warn about configuration against Microsoft recommendations.

### 💡 Tip

Same **quality checks and additional insights** are executed regularly when SAP systems are deployed or registered with **Azure Center for SAP solution** as well and are part of the service.

Further tools to allow easier deployment checks and document findings, plan next remediation steps and generally optimize your SAP on Azure landscape are:

- [Azure Well-Architected Framework review](#) An assessment of your workload focusing on the five main pillars of reliability, security, cost optimization, operation excellence and performance efficiency. Supports SAP workloads and recommended to running a review at start and after every project phase.
- [Azure Inventory Checks for SAP](#) An open source Azure Monitor workbook, which shows your Azure inventory with intelligence to highlight configuration drift and improve quality.

## Next steps

See these articles:

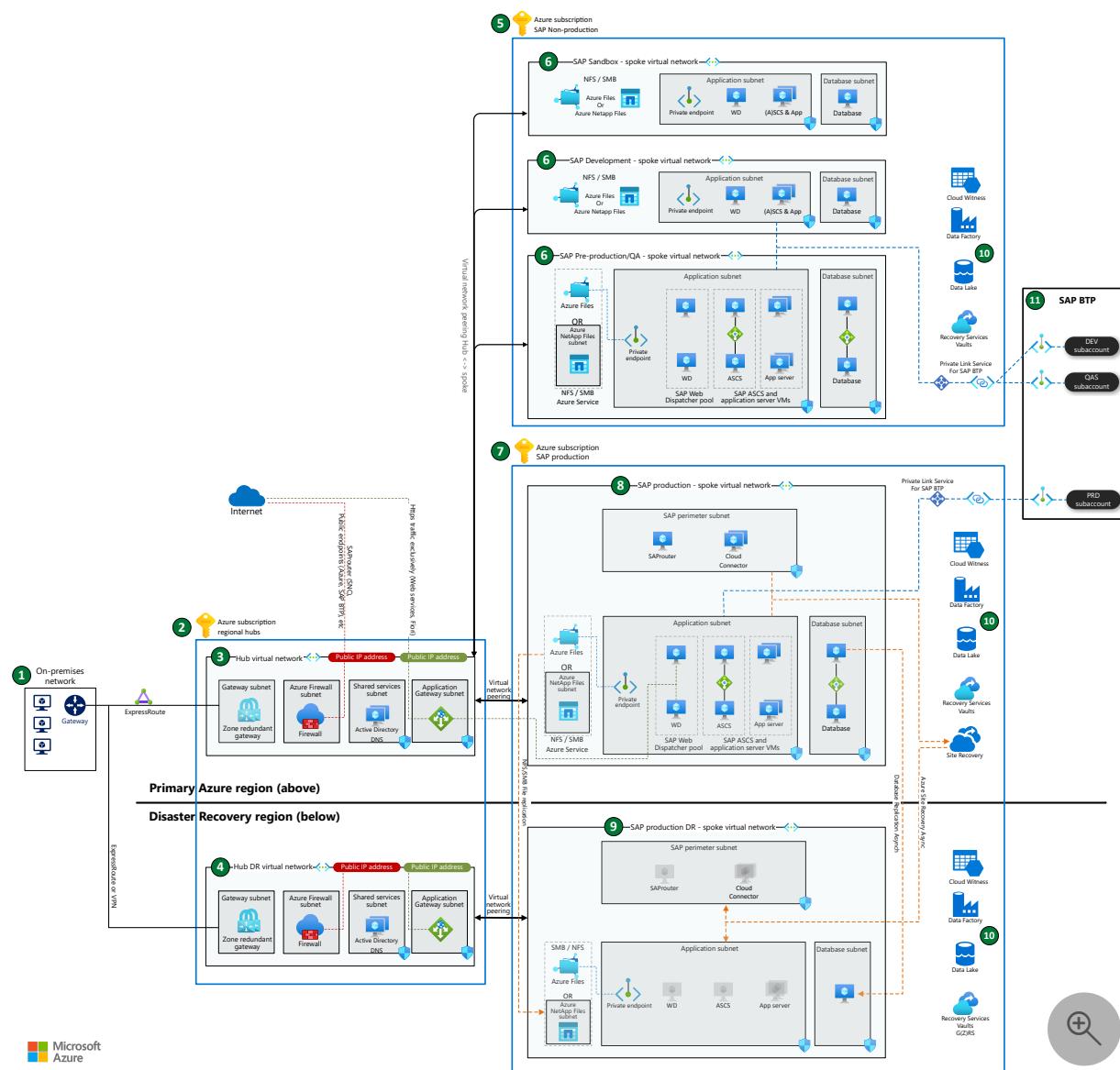
- ✓ [Azure planning and implementation for SAP NetWeaver](#)
- ✓ [Considerations for Azure Virtual Machines DBMS deployment for SAP workloads](#)
- ✓ [Azure Virtual Machines deployment for SAP NetWeaver](#)

# SAP landscape architecture

Azure Virtual Machines   Azure Virtual Network   Azure Files   Azure Load Balancer

This article provides best practices for architecting an entire SAP landscape in Azure. The SAP landscape includes multiple SAP systems across hub, production, non-production, and disaster recovery environments. We provide recommendations that focus on network design and not specific SAP systems. The goal is to provide our recommendations for architecting a secure, high-performing, and resilient SAP landscape.

## Architecture



Download a [Visio file](#) of the architecture.

# Workflow

1. *On-premises network*: ExpressRoute connection from on-premises network to connected Azure regions.
2. *Azure subscription regional hubs*: Azure subscription containing central services for the whole enterprise, not just SAP. The hub subscription provides connectivity by peering to spoke virtual networks containing SAP workloads.
3. *Hub virtual network*: A virtual network spoke for the central hub in the primary region or region A.
4. *Hub disaster recovery (DR) virtual network*: A virtual network spoke for the central hub in disaster recovery region. It mirrors the subnet design of the production virtual network in the primary region.
5. *Azure subscription SAP non-production*: An Azure subscription for all non-production SAP workloads. It includes pre-production, quality assurance, development, and sandbox environments.
6. *SAP non-production spoke virtual networks*: Separate virtual networks for SAP non-production workloads in the primary region. Each SAP environment has its own virtual network and subnets.
7. *Azure subscription SAP production*: An Azure subscription for all production SAP workloads.
8. *SAP production spoke virtual network*: A virtual network for the SAP production environment with multiple subnets. This virtual network is in the primary region.
9. *SAP production disaster recovery (DR) spoke virtual network*: A virtual network for SAP production in the secondary, disaster-recovery region. It mirrors the subnet design of the production virtual network in the primary region.
10. *Azure Services*: A sampling of Azure services that you can connect to the SAP landscape.
11. *SAP Business Technology Platform (BTP)*: The SAP environment accesses the SAP Business Technology Platform through Azure Private Link.

## Azure subscriptions

We recommend using a hub-spoke network design. With a hub-spoke design, you need at least three subscriptions to divide your SAP environments. You should have a subscription for the (1) regional hub virtual networks, (2) non-production virtual networks, and (3) production virtual networks. Subscriptions provide billing, policy, and security boundaries. There's no correct number of subscriptions. The number of subscriptions you use depends on your billing, policy, and security needs. In general, you want to avoid using too many subscriptions. Too many subscriptions can add

unneeded management overhead and networking complexity. For example, you don't need a subscription for each SAP system. Our architecture uses three subscriptions:

- *Regional hubs*: An Azure virtual hub subscription where the hub virtual network exists for the primary and secondary regions. This subscription is for all central services and not just SAP.
- *SAP non-production*: An Azure SAP non-production subscription where non-production systems, including sandbox, development, quality assurance, or pre-production systems, reside.
- *SAP production*: An Azure SAP production subscription where we configured the production and disaster recovery systems.

For more information, see:

- [Limits for each subscription](#)
- [Azure policies](#)
- [Management groups](#)

## Network design

A hub-spoke topology is the recommended network design for an SAP landscape. In this topology, the production hub virtual network acts as a central point of connectivity. It connects to the on-premises network and the various spoke virtual networks and enables users and applications access to the SAP workload. Within this hub-spoke topology, here are our recommendations for SAP network design.

**Use ExpressRoute for on-premises connection.** For SAP workloads, we recommend using ExpressRoute to connect the on-premises network to the Hub virtual network and Hub DR virtual network. You could use Azure virtual WAN topology if you have global locations. Consider setting up a site-to-site (S2S) VPN as a backup to Azure ExpressRoute or any third-party route requirements.

For more information, see:

- [Network topology and connectivity for an SAP migration](#)
- [Hub and spoke architecture](#)
- [Azure virtual WAN](#)
- [S2S VPN as a backup for ExpressRoute private peering](#)

**Use one virtual network per environment.** We recommend using one virtual network per environment (SAP deployment tier). The architecture uses a different virtual network

for production, development, quality assurance, and sandbox. This network design is ideal for large enterprise architectures.

**Use a central firewall.** All the network traffic to the spoke virtual networks, including remote function call (RFC) connections, should pass through a central firewall in the Hub virtual network. Network communication between the spoke virtual networks (spoke-to-spoke communication) passes through the hub virtual network firewall in the Azure Firewall subnet of the Hub virtual network. Similarly, network communication between the spoke virtual networks and on-premises network also pass through the hub virtual network firewall. We used virtual network peering to connect the various spoke virtual networks to the hub virtual network. All the communication between the spoke virtual networks passes through the Hub virtual network firewall. You could also use a network virtual appliance instead of a firewall. For more information, see [network virtual appliance ↗](#).

Network traffic that stays in a virtual network shouldn't pass through a firewall. For example, don't put a firewall between the SAP application subnet and SAP database subnet. You can't place a firewall or network virtual appliances (NVAs) between the SAP application and the database management system (DBMS) layer of SAP systems running the SAP kernel.

**Avoid peering spoke virtual networks.** Virtual network peering between the spoke virtual networks should be avoided if possible. Spoke-to-spoke virtual network peering allows spoke-to-spoke communication to bypass the Hub virtual network firewall. You should only configure spoke-to-spoke virtual network peering when you have high-bandwidth requirements, for example, with database replication between SAP environments. All other network traffic should run through the Hub virtual network and firewall. For more information, see [inbound and outbound internet connections for SAP on Azure](#).

## Subnets

It's a best practice to divide each SAP environment (production, pre-production, development, sandbox) into subnets and to use subnets to group related services. Here are our recommendations for subnetting an SAP landscape.

### Number of subnets

The production virtual network in the architecture has five subnets. This design is ideal for large enterprise solutions. The number of subnets can be less or more. The number of resources in the virtual network should determine the number of subnets in the virtual network.

## Subnet sizing

Ensure the subnets have sufficient network address space. If you use SAP virtual host names, you need more address space in your SAP subnets. Often each SAP instance requires 2-3 IP addresses and includes one IP address for the virtual machine hostname. Other Azure services might require their own dedicated subnet when deployed in the SAP workload virtual networks.

## Application subnet

The application subnet contains virtual machines running SAP application servers, SAP Central Services (ASCS), SAP Enqueue Replication Services (ERS), and SAP Web Dispatcher instances. The subnet also contains a private endpoint to Azure Files. In the diagram, we grouped the virtual machines by role. We recommend using virtual machine scale sets with flexible orchestration, availability zones or availability sets for resilient deployment. For more information, see [next steps](#).

## Database subnet

The database subnet holds virtual machines running databases. In the diagram, a pair of virtual machines with synchronous replication represent all the database virtual machines of one SAP environment.

## Perimeter subnets

Perimeter subnets are internet facing and include an SAP perimeter subnet and an Application Gateway subnet. Here are our recommendations for designing these two subnets.

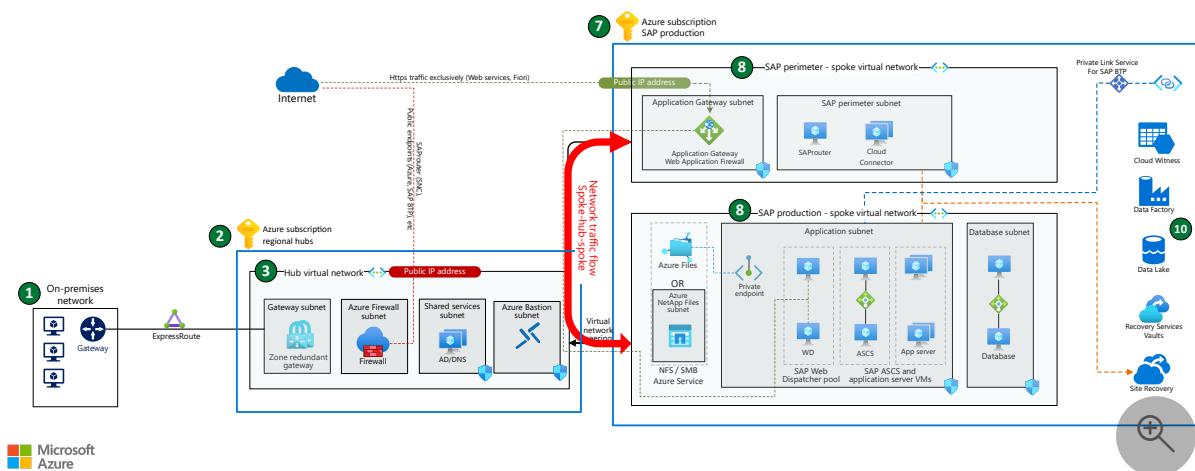
**SAP perimeter subnet:** The SAP perimeter subnet is a perimeter network that contains internet-facing applications such as SAProuter, SAP Cloud Connector, SAP Analytics Cloud Agent, and Application Gateway. These services have dependencies on SAP systems that an SAP team should deploy, manage, and configure. A central IT team shouldn't manage the services in the SAP perimeter subnet. For this reason, you should place these services in the SAP spoke virtual network and not the Hub virtual network. The architecture diagram only shows a production SAP perimeter network. It doesn't have an SAP perimeter subnet in the non-production virtual networks. The workloads in the non-production SAP subscription use the services in the SAP perimeter subnet.

You can create separate set SAP perimeter subnet in the non-production subscription. We only recommend this approach for critical workloads or workloads that change frequently. A dedicated non-production SAP perimeter is helpful for testing and new

feature deployment. Less critical applications or applications that will only have few modifications over time don't need a separate non-production SAP perimeter subnet.

**Application Gateway subnet:** Azure Application Gateway requires its own subnet. Use it to allow traffic from the Internet that SAP services, such as SAP Fiori, can use. The recommended architecture places Azure Application Gateway together with its frontend public IP address in the Hub virtual network. An Azure Application Gateway requires at least a /29 size subnet. We recommend size /27 or larger. You can't use both versions of Application Gateway (v1 and v2) in the same subnet. For more information, see [subnet for Azure Application Gateway](#).

**Place perimeter subnets in a separate virtual network for increased security:** For increased security, you can put the SAP perimeter subnet and Application Gateway subnet in a separate virtual network within the SAP production subscription. The SAP perimeter spoke virtual network is peered with the Hub virtual network, and all network traffic to public networks flows through the perimeter virtual network. This alternative approach shows Azure Application Gateway with its public IP address for inbound connections placed in a spoke virtual network for SAP use exclusively.



Download a [Visio file](#) including this alternative architecture.

This network design provides better incident response capabilities and fine-grained network access control. However, it also increases the management complexity, network latency, and cost of the deployment. Let's discuss each point.

**Better incident response:** The SAP perimeter spoke virtual network allows quick isolation of compromised services if you detect a breach. You can remove virtual network peering from the SAP perimeter spoke virtual network to the hub and immediately isolate the SAP perimeter workloads and SAP application virtual network applications from the internet. You don't want to rely on network security group (NSG) rules changes for

incident response. Changing or removing an NSG rule only affects new connections and won't cut existing malicious connections.

*Fine-grained network access control:* The SAP perimeter virtual network provides more stringent network access control to and from the SAP production spoke virtual network.

*Increased complexity, latency, and cost:* The architecture increases management complexity, cost, and latency. Internet-bound communication from the SAP production virtual network is peered twice, once to the Hub virtual network and again to the SAP perimeter virtual network out to the internet. The firewall in the Hub virtual network has the greatest effect on latency. We recommend measuring the latency to see if your use case can support it.

For more information, see [perimeter network best practices](#).

## Azure NetApp Files subnet

If you're using NetApp Files, you should have a delegated subnet to provide network file system (NFS) or server message block (SMB) file shares for different SAP on Azure scenarios. A /24 subnet is the default size for a NetApp Files subnet, but you can change the size to meet your needs. Use your own requirements to determine the proper sizing. For more information, see [delegated subnet](#).

## Subnet security

Using subnets to group SAP resources that have the same security rule requirements makes it easier to manage the security.

**Network security groups (NSG):** Subnets allow you to implement network security groups at the subnet level. Grouping resources in the same subnet that require different security rules requires network security groups at the subnet level and network-interface level. With this two-level setup, security rules easily conflict and can cause unexpected communication problems that are difficult to troubleshoot. NSG rules also affect traffic [within the subnet](#). For more information on NSGs, see [network security groups](#).

**Application security groups (ASG):** We recommend using application security groups to group virtual machine network interfaces and reference the application security groups in the network-security-group rules. This configuration allows easier rule creation and management for SAP deployments. Each network interface can belong to multiple application security groups with different network-security-group rules. For more information, see [application security groups](#).

## Azure Private Link

We recommend using Azure Private Link to improve the security of network communications. Azure Private Link uses private endpoints with private IP addresses to communicate with Azure services. Azure Private Links avoids sending network communication over the internet to public endpoints. For more information, see [private endpoints on Azure services](#).

**Use private endpoints in the application subnet:** We recommend using private endpoints to connect the application subnet to supported Azure services. In the architecture, there's a private endpoint for Azure Files in the Application subnet of each virtual network. You can extend this concept to any supported Azure service.

**Use Azure Private Link for SAP Business Technology Platform (BTP):** Azure Private Link for SAP Business Technology Platform (BTP) is now generally available. SAP Private Link Service supports connections from SAP BTP, the Cloud Foundry runtime, and other services. Example scenarios include SAP S/4HANA or SAP ERP running on the virtual machine. They can connect to Azure native services such as Azure Database for MariaDB and Azure Database for MySQL.

The architecture depicts an SAP Private Link Service connection from SAP BTP environments. SAP Private Link Service establishes a private connection between specific SAP BTP services and specific services in each network as service provider accounts. Private link allows BTP services to access your SAP environment through private network connections. It improves security by not using the public internet to communicate.

For more information, see:

- [Azure Private Link resources](#) ↗
- [Azure Database for MariaDB](#) ↗
- [Azure Database for MySQL](#) ↗
- [Internet connection for SAP on Azure](#)

## Network file system (NFS) and server message block (SMB) file shares

SAP systems often depend on network file system volumes or server message block shares. These file shares move files between virtual machines or function as a file interface with other applications. We recommend using native Azure services, such as Azure Premium Files and Azure NetApp Files, as your network file system (NFS) and server message block (SMB) file shares. Azure services have better combined availability, resilience, and service level agreements (SLAs) than operating-system-based tools.

For more information, see:

- [Azure Premium Files](#)
- [Azure NetApp Files](#)
- [SAP note 2015553 \(requirements for storage services\)](#) ↗

When architecting your SAP solution, you need to properly size the individual file share volumes and know which SAP system file share connects to. Keep scalability and performance targets of the Azure service in mind during planning. The following table outlines common SAP file shares and gives a brief description and recommended use in a whole SAP environment.

 [Expand table](#)

<b>File share name</b>	<b>Usage</b>	<b>Recommendation</b>
<code>sapmnt</code>	Distributed SAP system, profile, and global directories	Dedicated share for each SAP system, no reuse
<code>cluster</code>	High-availability shares for ASCS, ERS, and database per respective design	Dedicated share for each SAP system, no reuse
<code>saptrans</code>	SAP transport directory	One share for one or few SAP landscapes (ERP, Business Warehouse)
<code>interface</code>	File exchange with non-SAP applications	Customer specific requirements, separate file shares per environment (production, non-production)

You can only share `saptrans` between different SAP environments, and, as such, you should carefully consider its placement. Avoid consolidating too many SAP systems into one `saptrans` share for scalability and performance reasons.

The corporate security policies will drive the architecture and separation of volumes between environments. A transport directory with separation per environment or tier needs RFC communication between SAP environments to allow SAP transport groups or transport domain links. For more information, see:

- [SAP transport groups](#) ↗

- [Transport domain links](#)

## Data services

The architecture contains Azure data services that help you extend and improve your SAP data platform. To help unlock business insights, we recommend you use services such as Azure Synapse Analytics, Azure Data Factory, and Azure Data Lake Storage. These data services help you analyze and visualize SAP data and non-SAP data.

For many data integration scenarios, an integration runtime is required. The Azure integration runtime is the compute infrastructure that Azure Data Factory and Azure Synapse Analytics pipelines use to provide data integration capabilities. We recommend the deployment of runtime virtual machines for these services for each environment separately. For more information, see:

- [Azure integration runtime](#)
- [Set up a self-hosted integration runtime to use in the SAP CDC solution](#)
- [Copy data from SAP HANA](#)
- [Copy data from SAP Business Warehouse via Open Hub](#)

## Shared services

SAP solutions rely on shared services. Load balancer and application gateways are examples of services that multiple SAP systems use. The architecture but your organizational needs should determine how you architect your shared services. Here's general guidance you should follow.

**Load balancers:** We recommend one load balancer per SAP system. This configuration helps minimize complexity. You want to avoid too many pools and rules on a single load balancer. This configuration also ensures naming and placement aligns with the SAP system and resource group. Each SAP system with a clustered high-availability (HA) architecture should have at least one load balancer. The architecture uses one load balancer for the ASCS virtual machines and a second load balancer for the database virtual machines. Some databases might not require load balancers to create a high-availability deployment. SAP HANA does. Check the database-specific documentation for more details.

**Application Gateway:** We recommend at least one application gateway per SAP environment (production, non-production, and sandbox) unless the complexity and number of connected systems is too high. You could use an application gateway for multiple SAP systems to reduce complexity since not all SAP systems in the environment

require public access. A single application gateway could serve multiple web dispatcher ports for a single SAP S/4HANA system or be used by different SAP systems.

**SAP Web Dispatcher virtual machines:** The architecture shows a pool of two or more SAP Web Dispatcher VMs. We recommend that you don't reuse SAP Web Dispatcher virtual machines between different SAP systems. Keeping them separate allows you to size the Web Dispatcher virtual machines to meet the needs of each SAP system. For smaller SAP solutions, we recommend embedding the Web Dispatcher services in the ASCS instance.

**SAP services:** SAP services like SAProuter, Cloud Connector, and Analytics Cloud Agent, are deployed based on application requirements, either centrally or split up. No recommendation on reuse between SAP systems due to diverse customer requirements. Main decision to make is mentioned in networking section, if and when SAP perimeter subnet for non-production should be used. Otherwise with just production perimeter subnet for SAP, the SAP perimeter services are consumed by entire SAP landscape.

## Disaster recovery

Disaster recovery (DR) addresses the requirement for business continuity in case the primary Azure region is unavailable or compromised. From an overall SAP landscape perspective and shown in the diagram, here are our recommendations for disaster recovery design.

**Use different IP address ranges** Virtual networks only span a single Azure region. Any disaster recovery solutions should use a different region. You need to create a different virtual network in the secondary region. The virtual network in the DR environment needs a different IP address range to enable database synchronization through database native technology.

**Central services and connectivity to on-premises:** Connectivity to on-premises and key central services (DNS or firewalls) must be available in the disaster recovery region. Availability and change configuration of the central IT services need to be part your disaster recovery plan. Central IT services are key components for a functioning SAP environment.

**Use Azure Site Recovery** Azure Site Recovery replicates and protects managed disks and virtual machines configurations for application servers to the DR region.

**Ensure file share availability:** SAP depends on availability of key file shares. Backup or continuous file share replication is necessary to provide data on these file shares with minimal data loss in DR scenario.

**Database replication** Azure Site Recovery can't protect SAP database servers due to the high change rate and lack of database support by the service. You need to configure continuous and asynchronous database replication to the disaster recovery region.

For more information, see [disaster recovery overview and infrastructure guidelines for SAP workload](#).

## Smaller SAP architecture

For smaller SAP solutions, it might be beneficial to simplify the network design. Each SAP environment's virtual network would then be subnets inside such combined virtual network. Any simplification of the network and subscription design needs can affect security. You should reevaluate the network routing, access to and from public networks, access to shared services (file shares), and access other Azure services. Here are some options for shrinking the architecture to better meet organizational needs.

**Combine the SAP application and database subnets into one.** You can combine the application and database subnets to create one large SAP network. This network design mirrors many on-premises SAP networks. The combination of these two subnets requires higher attention to subnet security and network-security-group rules. Application security groups are important when using a single subnet for SAP application and database subnets.

**Combine SAP perimeter subnet and application subnet.** You can combine the perimeter subnet and SAP application subnet. A heightened attention must be placed on network-security-group rules and application security group use. We only recommend this simplification approach for small SAP estates.

**Combine SAP spoke virtual networks between different SAP environments** The architecture uses different virtual networks for each SAP environment (hub, production, non-production, and disaster recovery). Depending on the size of your SAP landscape, you can combine the SAP spoke virtual networks into fewer or even only one SAP spoke. You still need to divide between production and non-production environments. Each SAP production environment becomes a subnet in one SAP production virtual network. Each SAP non-production environment becomes a subnet in one SAP non-production virtual network.

## Contributors

*Microsoft maintains this article. It was originally written by the following contributors.*

**Principal authors:**

- [Robert Biro](#) | Senior Architect
- [Pankaj Meshram](#) | Principal Program Manager

## Next steps

- SAP S/4HANA in Linux on Azure
- Run SAP NetWeaver in Windows on Azure
- Run SAP HANA in a scale-up architecture on Azure
- Cloud Adoption Framework - SAP scenario
- In- and Outbound internet connections for SAP on Azure
- SAP on Azure documentation.
- Azure planning and implementation guide for SAP workloads
- SAP workloads on Azure: planning and deployment checklist

# SAP HANA infrastructure configurations and operations on Azure

Article • 11/09/2023

This document provides guidance for configuring Azure infrastructure and operating SAP HANA systems that are deployed on Azure native virtual machines (VMs). The document also includes configuration information for SAP HANA scale-out for the M128s VM SKU. This document isn't intended to replace the standard SAP documentation, which includes the following content:

- [SAP administration guide ↗](#)
- [SAP installation guides ↗](#)
- [SAP notes ↗](#)

## Prerequisites

To use this guide, you need basic knowledge of the following Azure components:

- [Azure virtual machines](#)
- [Azure networking and virtual networks](#)
- [Azure Storage](#)

To learn more about SAP NetWeaver and other SAP components on Azure, see the [SAP on Azure](#) section of the [Azure documentation](#).

## Basic setup considerations

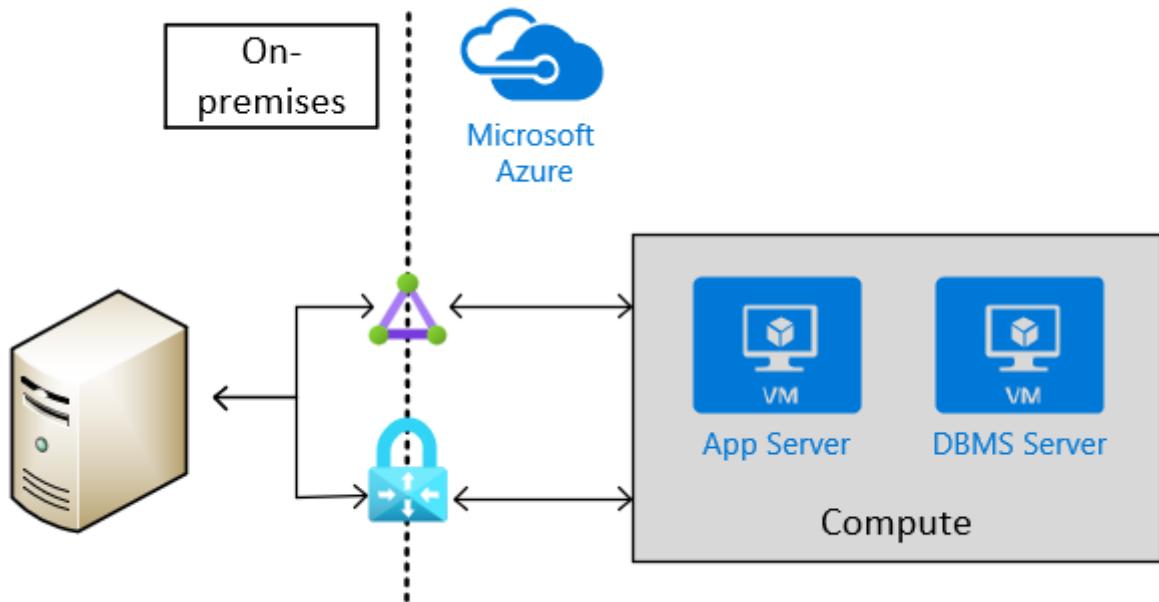
The following sections describe basic setup considerations for deploying SAP HANA systems on Azure VMs.

## Connect into Azure virtual machines

As documented in the [Azure virtual machines planning guide](#), there are two basic methods for connecting into Azure VMs:

- Connect through the internet and public endpoints on a Jump VM or on the VM that is running SAP HANA.
- Connect through a [VPN](#) or [Azure ExpressRoute ↗](#).

Site-to-site connectivity via VPN or ExpressRoute is necessary for production scenarios. This type of connection is also needed for non-production scenarios that feed into production scenarios where SAP software is being used. The following image shows an example of cross-site connectivity:



## Choose Azure VM types

SAP lists which [Azure VM types that you can use for production scenarios](#). For non-production scenarios, a wider variety of native Azure VM types is available.

### ⓘ Note

For non-production scenarios, use the VM types that are listed in the [SAP note #1928533](#). For the usage of Azure VMs for production scenarios, check for SAP HANA certified VMs in the SAP published [Certified IaaS Platforms list](#).

Deploy the VMs in Azure by using:

- The Azure portal.
- Azure PowerShell cmdlets.
- The Azure CLI.

You also can deploy a complete installed SAP HANA platform on the Azure VM services through the [SAP Cloud platform](#). The installation process is described in [Deploy SAP S/4HANA or BW/4HANA on Azure](#).

## Important

In order to use M208xx\_v2 VMs, you need to be careful selecting your Linux image. For more information, see [Memory optimized virtual machine sizes](#).

## Storage configuration for SAP HANA

For storage configurations and storage types to be used with SAP HANA in Azure, read the document [SAP HANA Azure virtual machine storage configurations](#)

## Set up Azure virtual networks

When you have site-to-site connectivity into Azure via VPN or ExpressRoute, you must have at least one Azure virtual network that is connected through a Virtual Gateway to the VPN or ExpressRoute circuit. In simple deployments, the Virtual Gateway can be deployed in a subnet of the Azure virtual network (VNet) that hosts the SAP HANA instances as well. To install SAP HANA, you create two more subnets within the Azure virtual network. One subnet hosts the VMs to run the SAP HANA instances. The other subnet runs Jumpbox or Management VMs to host SAP HANA Studio, other management software, or your application software.

## Important

Out of functionality, but more important out of performance reasons, it is not supported to configure [Azure Network Virtual Appliances](#) in the communication path between the SAP application and the DBMS layer of a SAP NetWeaver, Hybris or S/4HANA based SAP system. The communication between the SAP application layer and the DBMS layer needs to be a direct one. The restriction does not include [Azure ASG and NSG rules](#) as long as those ASG and NSG rules allow a direct communication. Further scenarios where NVAs are not supported are in communication paths between Azure VMs that represent Linux Pacemaker cluster nodes and SBD devices as described in [High availability for SAP NetWeaver on Azure VMs on SUSE Linux Enterprise Server for SAP applications](#). Or in communication paths between Azure VMs and Windows Server SOFS set up as described in [Cluster an SAP ASCS/SCS instance on a Windows failover cluster by using a file share in Azure](#). NVAs in communication paths can easily double the network latency between two communication partners, can restrict throughput in critical paths between the SAP application layer and the DBMS layer. In some scenarios observed with customers, NVAs can cause Pacemaker Linux clusters to fail

in cases where communications between the Linux Pacemaker cluster nodes need to communicate to their SBD device through an NVA.

### ⓘ Important

Another design that is **NOT** supported is the segregation of the SAP application layer and the DBMS layer into different Azure virtual networks that are not **peered** with each other. It is recommended to segregate the SAP application layer and DBMS layer using subnets within an Azure virtual network instead of using different Azure virtual networks. If you decide not to follow the recommendation, and instead segregate the two layers into different virtual network, the two virtual networks need to be **peered**. Be aware that network traffic between two **peered** Azure virtual networks are subject of transfer costs. With the huge data volume in many Terabytes exchanged between the SAP application layer and DBMS layer substantial costs can be accumulated if the SAP application layer and DBMS layer is segregated between two peered Azure virtual networks.

If you deployed Jumpbox or management VMs in a separate subnet, you can define **multiple virtual network interface cards (vNICs)** for the HANA VM, with each vNIC assigned to different subnet. With the ability to have multiple vNICs, you can set up network traffic separation, if necessary. For example, client traffic can be routed through the primary vNIC and admin traffic is routed through a second vNIC. You also assign static private IP addresses that are deployed for both virtual NICs.

### ⓘ Note

You should assign static IP addresses through Azure means to individual vNICs. You should not assign static IP addresses within the guest OS to a vNIC. Some Azure services like Azure Backup Service rely on the fact that at least the primary vNIC is set to DHCP and not to static IP addresses. See also the document **Troubleshoot Azure virtual machine backup**. If you need to assign multiple static IP addresses to a VM, you need to assign multiple vNICs to a VM.

However, for deployments that are enduring, you need to create a virtual datacenter network architecture in Azure. This architecture recommends the separation of the Azure VNet Gateway that connects to on-premises into a separate Azure VNet. This separate VNet should host all the traffic that leaves either to on-premises or to the internet. This approach allows you to deploy software for auditing and logging traffic that enters the virtual datacenter in Azure in this separate hub VNet. So you have one

VNet that hosts all the software and configurations that relate to in- and outgoing traffic to your Azure deployment.

The articles [Azure Virtual Datacenter: A Network Perspective](#) and [Azure Virtual Datacenter and the Enterprise Control Plane](#) give more information on the virtual datacenter approach and related Azure VNet design.

### Note

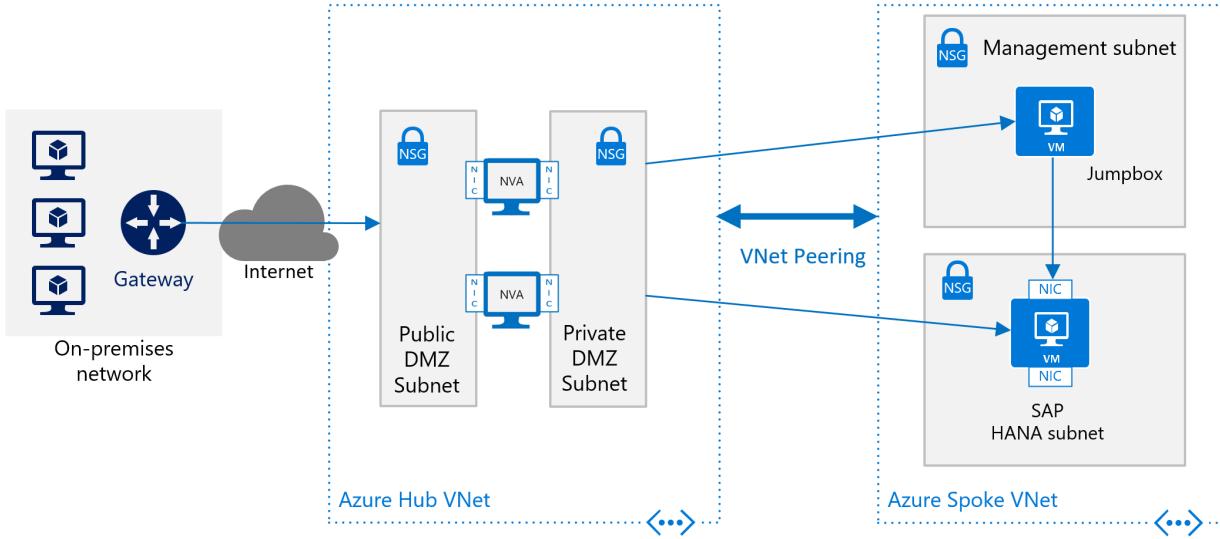
Traffic that flows between a hub VNet and spoke VNet using [Azure VNet peering](#) is subject of additional [costs](#). Based on those costs, you might need to consider making compromises between running a strict hub and spoke network design and running multiple [Azure ExpressRoute Gateways](#) that you connect to 'spokes' in order to bypass VNet peering. However, Azure ExpressRoute Gateways introduce additional [costs](#) as well. You also may encounter additional costs for third-party software you use for network traffic logging, auditing, and monitoring. Dependent on the costs for data exchange through VNet peering on the one side and costs created by additional Azure ExpressRoute Gateways and additional software licenses, you may decide for micro-segmentation within one VNet by using subnets as isolation unit instead of VNets.

For an overview of the different methods for assigning IP addresses, see [IP address types and allocation methods in Azure](#).

For VMs running SAP HANA, you should work with static IP addresses assigned. Reason is that some configuration attributes for HANA reference IP addresses.

[Azure Network Security Groups \(NSGs\)](#) are used to direct traffic that's routed to the SAP HANA instance or the jumpbox. The NSGs and eventually [Application Security Groups](#) are associated to the SAP HANA subnet and the Management subnet.

To deploy SAP HANA in Azure without a site-to-site connection, you still want to shield the SAP HANA instance from the public internet and hide it behind a forward proxy. In this basic scenario, the deployment relies on Azure built-in DNS services to resolve hostnames. In a more complex deployment where public-facing IP addresses are used, Azure built-in DNS services are especially important. Use Azure NSGs and [Azure NVAs](#) to control, monitor the routing from the internet into your Azure VNet architecture in Azure. The following image shows a rough schema for deploying SAP HANA without a site-to-site connection in a hub and spoke VNet architecture:



Another description on how to use Azure NVAs to control and monitor access from Internet without the hub and spoke VNet architecture can be found in the article [Deploy highly available network virtual appliances](#).

## Configuring Azure infrastructure for SAP HANA scale-out

In order to find out the Azure VM types that are certified for either OLAP scale-out or S/4HANA scale-out, check the [SAP HANA hardware directory](#). A checkmark in the column 'Clustering' indicates scale-out support. Application type indicates whether OLAP scale-out or S/4HANA scale-out is supported. For details on nodes certified in scale-out, review the entry for a specific VM SKU listed in the SAP HANA hardware directory.

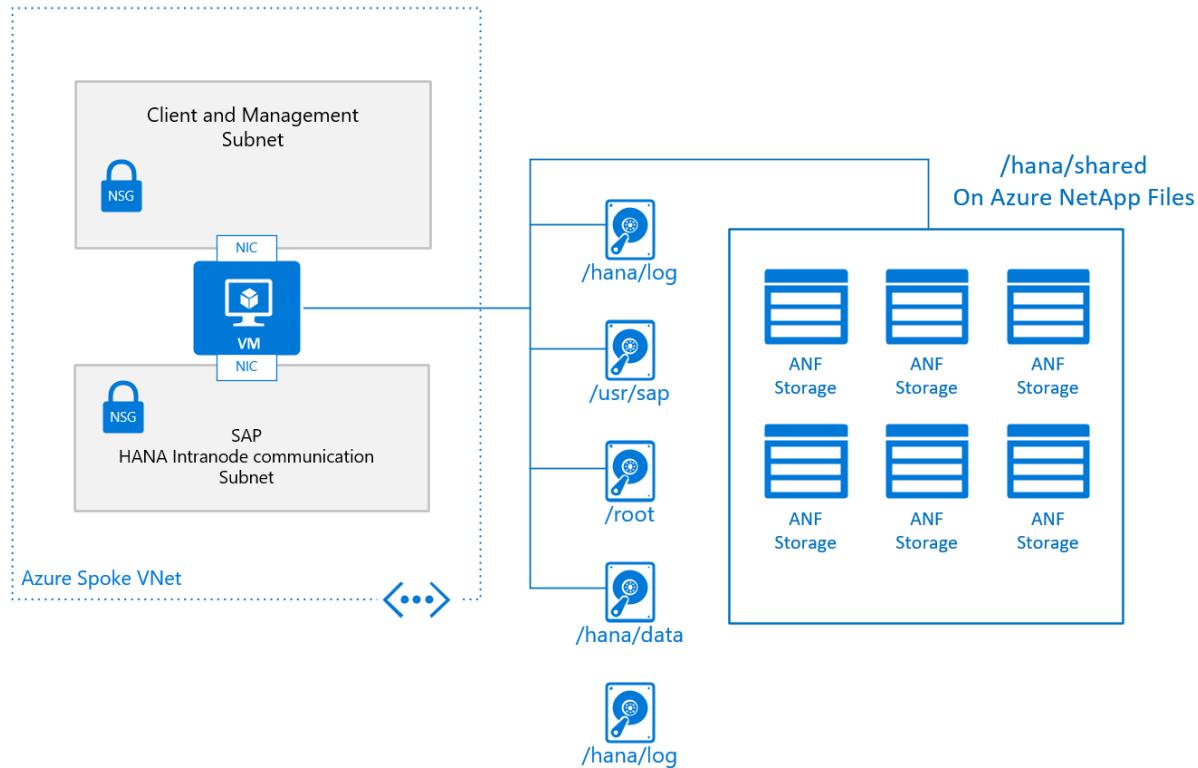
The minimum OS releases for deploying scale-out configurations in Azure VMs, check the details of the entries in the particular VM SKU listed in the SAP HANA hardware directory. Of a n-node OLAP scale-out configuration, one node functions as the main node. The other nodes up to the limit of the certification act as worker node. More standby nodes don't count into the number of certified nodes

### ! Note

Azure VM scale-out deployments of SAP HANA with standby node are only possible using the [Azure NetApp Files](#) storage. No other SAP HANA certified Azure storage allows the configuration of SAP HANA standby nodes

For /hana/shared, we recommend the usage of [Azure NetApp Files](#) or [Azure Files](#).

A typical basic design for a single node in a scale-out configuration, with **/hana/shared** deployed on Azure NetApp Files, looks like:



The basic configuration of a VM node for SAP HANA scale-out looks like:

- For **/hana/shared**, you use the native NFS service provided through Azure NetApp Files or Azure Files.
- All other disk volumes aren't shared among the different nodes and aren't based on NFS. Installation configurations and steps for scale-out HANA installations with non-shared **/hana/data** and **/hana/log** is provided further later in this document. For HANA certified storage that can be used, check the article [SAP HANA Azure virtual machine storage configurations](#).

Sizing the volumes or disks, you need to check the document [SAP HANA TDI Storage Requirements](#), for the size required dependent on the number of worker nodes. The document releases a formula you need to apply to get the required capacity of the volume

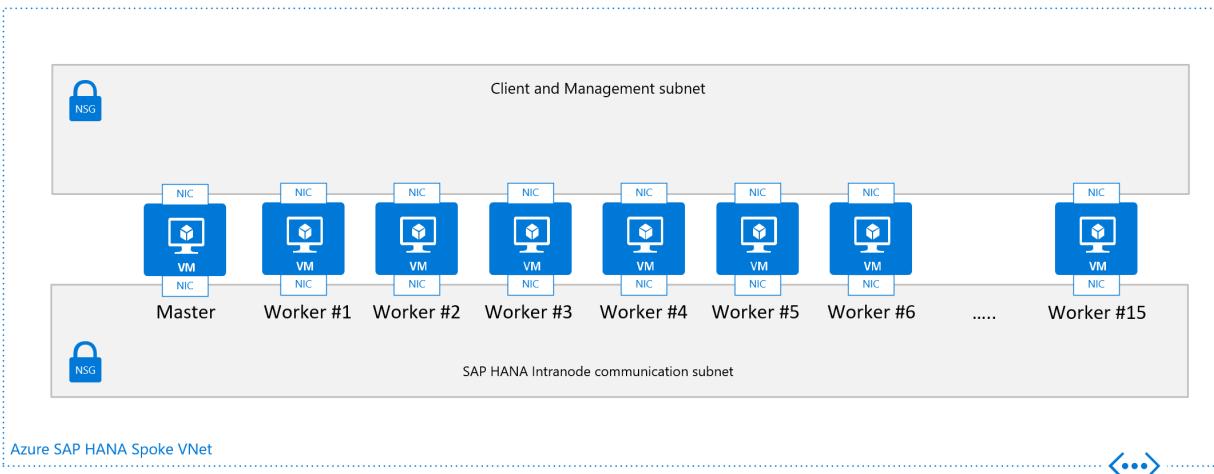
The other design criteria that is displayed in the graphics of the single node configuration for a scale-out SAP HANA VM is the VNet, or better the subnet configuration. SAP highly recommends a separation of the client/application facing traffic from the communications between the HANA nodes. As shown in the graphics, this goal is achieved by having two different vNICs attached to the VM. Both vNICs are in different subnets, have two different IP addresses. You then control the flow of traffic with routing rules using NSGs or user-defined routes.

Particularly in Azure, there are no means and methods to enforce quality of service and quotas on specific vNICs. As a result, the separation of client/application facing and intra-node communication doesn't open any opportunities to prioritize one traffic stream over the other. Instead the separation remains a measure of security in shielding the intra-node communications of the scale-out configurations.

### ⓘ Note

SAP recommends separating network traffic to the client/application side and intra-node traffic as described in this document. Therefore putting an architecture in place as shown in the last graphics is recommended. Also consult your security and compliance team for requirements that deviate from the recommendation

From a networking point of view the minimum required network architecture would look like:



## Installing SAP HANA scale-out n Azure

Installing a scale-out SAP configuration, you need to perform rough steps of:

- Deploying new or adapting an existing Azure VNet infrastructure
- Deploying the new VMs using Azure Managed Premium Storage, Ultra disk volumes, and/or NFS volumes based on ANF
  - Adapt network routing to make sure that, for example, intra-node communication between VMs isn't routed through an [NVA](#).
- Install the SAP HANA main node.
- Adapt configuration parameters of the SAP HANA main node
- Continue with the installation of the SAP HANA worker nodes

## Installation of SAP HANA in scale-out configuration

As your Azure VM infrastructure is deployed, and all other preparations are done, you need to install the SAP HANA scale-out configurations in these steps:

- Install the SAP HANA main node according to SAP's documentation
- When using Azure Premium Storage or Ultra disk storage with non-shared disks of `/hana/data` and `/hana/log`, add the parameter `basepath_shared = no` to the `global.ini` file. This parameter enables SAP HANA to run in scale-out without shared `/hana/data` and `/hana/log` volumes between the nodes. Details are documented in [SAP Note #2080991 ↗](#). If you're using NFS volumes based on ANF for `/hana/data` and `/hana/log`, you don't need to make this change
- After the eventual change in the `global.ini` parameter, restart the SAP HANA instance
- Add more worker nodes. For more information, see [Add Hosts Using the Command-Line Interface ↗](#). Specify the internal network for SAP HANA inter-node communication during the installation or afterwards using, for example, the local `hdblcm`. For more detailed documentation, see [SAP Note #2183363 ↗](#).

To set up an SAP HANA scale-out system with a standby node, see the [SUSE Linux deployment instructions](#) or the [Red Hat deployment instructions](#).

## SAP HANA Dynamic Tiering 2.0 for Azure virtual machines

In addition to the SAP HANA certifications on Azure M-series VMs, SAP HANA Dynamic Tiering 2.0 is also supported on Microsoft Azure. For more information, see [Links to DT 2.0 documentation](#). There's no difference in installing or operating the product. For example, you can install SAP HANA Cockpit inside an Azure VM. However, there are some mandatory requirements, as described in the following section, for official support on Azure. Throughout the article, the abbreviation "DT 2.0" is going to be used instead of the full name Dynamic Tiering 2.0.

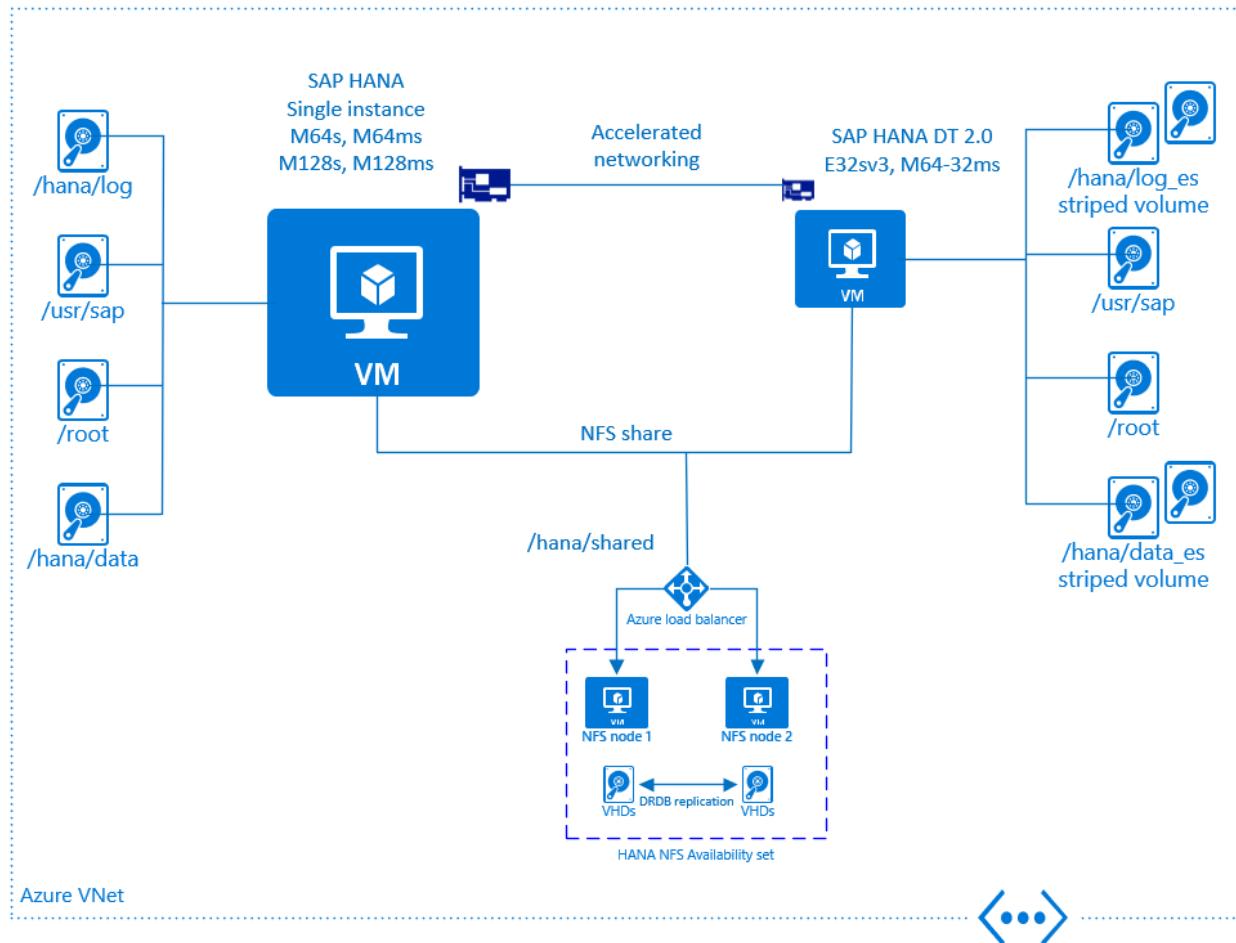
SAP HANA Dynamic Tiering 2.0 isn't supported by SAP BW or S4HANA. Main use cases right now are native HANA applications.

## Overview

The picture below gives an overview regarding DT 2.0 support on Microsoft Azure. There's a set of mandatory requirements, which has to be followed to comply with the official certification:

- DT 2.0 must be installed on a dedicated Azure VM. It may not run on the same VM where SAP HANA runs
- SAP HANA and DT 2.0 VMs must be deployed within the same Azure Vnet
- The SAP HANA and DT 2.0 VMs must be deployed with Azure accelerated networking enabled
- Storage type for the DT 2.0 VMs must be Azure Premium Storage
- Multiple Azure disks must be attached to the DT 2.0 VM
- It's required to create a software raid / striped volume (either via lvm or mdadm) using striping across the Azure disks

More details are going to be explained in the following sections.



## Dedicated Azure VM for SAP HANA DT 2.0

On Azure IaaS, DT 2.0 is only supported on a dedicated VM. It isn't allowed to run DT 2.0 on the same Azure VM where the HANA instance is running. Initially two VM types can be used to run SAP HANA DT 2.0:

- M64-32ms
- E32sv3

For more information on the VM type description, see [Azure VM sizes - Memory](#)

Given the basic idea of DT 2.0, which is about offloading "warm" data in order to save costs it makes sense to use corresponding VM sizes. There's no strict rule though regarding the possible combinations. It depends on the specific customer workload.

Recommended configurations would be:

SAP HANA VM type	DT 2.0 VM type
M128ms	M64-32ms
M128s	M64-32ms
M64ms	E32sv3
M64s	E32sv3

All combinations of SAP HANA-certified M-series VMs with supported DT 2.0 VMs (M64-32ms and E32sv3) are possible.

## Azure networking and SAP HANA DT 2.0

Installing DT 2.0 on a dedicated VM requires network throughput between the DT 2.0 VM and the SAP HANA VM of 10 Gb minimum. Therefore it's mandatory to place all VMs within the same Azure Vnet and enable Azure accelerated networking.

See additional information about Azure accelerated networking [Create an Azure VM with Accelerated Networking using Azure CLI](#)

## VM Storage for SAP HANA DT 2.0

According to DT 2.0 best practice guidance, the disk IO throughput should be minimum 50 MB/sec per physical core.

According to the specifications for the two Azure VM types, which are supported for DT 2.0, the maximum disk IO throughput limit for the VM looks like:

- E32sv3: 768 MB/sec (uncached) which means a ratio of 48 MB/sec per physical core
- M64-32ms: 1000 MB/sec (uncached) which means a ratio of 62.5 MB/sec per physical core

It's required to attach multiple Azure disks to the DT 2.0 VM and create a software raid (striping) on OS level to achieve the max limit of disk throughput per VM. A single Azure

disk can't provide the throughput to reach the max VM limit in this regard. Azure Premium storage is mandatory to run DT 2.0.

- Details about available Azure disk types can be found on the [Select a disk type for Azure IaaS VMs - managed disks](#) page
- Details about creating software raid via mdadm can be found on the [Configure software RAID on a Linux VM](#) page
- Details about configuring LVM to create a striped volume for max throughput can be found on the [Configure LVM on a virtual machine running Linux](#) page

Depending on size requirements, there are different options to reach the max throughput of a VM. Here are possible data volume disk configurations for every DT 2.0 VM type to achieve the upper VM throughput limit. The E32sv3 VM should be considered as an entry level for smaller workloads. In case it should turn out that it's not fast enough it might be necessary to resize the VM to M64-32ms. As the M64-32ms VM has much memory, the IO load might not reach the limit especially for read intensive workloads. Therefore fewer disks in the stripe set might be sufficient depending on the customer specific workload. But to be on the safe side the disk configurations below were chosen to guarantee the maximum throughput:

<b>VM SKU</b>	<b>Disk Config 1</b>	<b>Disk Config 2</b>	<b>Disk Config 3</b>	<b>Disk Config 4</b>	<b>Disk Config 5</b>
M64-32ms	4 x P50 -> 16 TB	4 x P40 -> 8 TB	5 x P30 -> 5 TB	7 x P20 -> 3.5 TB	8 x P15 -> 2 TB
E32sv3	3 x P50 -> 12 TB	3 x P40 -> 6 TB	4 x P30 -> 4 TB	5 x P20 -> 2.5 TB	6 x P15 -> 1.5 TB

Especially in case the workload is read-intense it could boost IO performance to turn on Azure host cache "read-only" as recommended for the data volumes of database software. Whereas for the transaction log Azure host disk cache must be "none".

Regarding the size of the log volume a recommended starting point is a heuristic of 15% of the data size. The creation of the log volume can be accomplished by using different Azure disk types depending on cost and throughput requirements. For the log volume, high I/O throughput is required.

When using the VM type M64-32ms, it's mandatory to enable [Write Accelerator](#). Azure Write Accelerator provides optimal disk write latency for the transaction log (only available for M-series). There are some items to consider though like the maximum number of disks per VM type. Details about Write Accelerator can be found on the [Azure Write Accelerator](#) page

Here are a few examples about sizing the log volume:

data volume size and disk type	log volume and disk type config 1	log volume and disk type config 2
4 x P50 -> 16 TB	5 x P20 -> 2.5 TB	3 x P30 -> 3 TB
6 x P15 -> 1.5 TB	4 x P6 -> 256 GB	1 x P15 -> 256 GB

Like for SAP HANA scale-out, the /hana/shared directory has to be shared between the SAP HANA VM and the DT 2.0 VM. The same architecture as for SAP HANA scale-out using dedicated VMs, which act as a highly available NFS server is recommended. In order to provide a shared backup volume, the identical design can be used. But it's up to the customer if HA would be necessary or if it's sufficient to just use a dedicated VM with enough storage capacity to act as a backup server.

## Links to DT 2.0 documentation

- [SAP HANA Dynamic Tiering installation and update guide ↗](#)
- [SAP HANA Dynamic Tiering tutorials and resources ↗](#)
- [SAP HANA Dynamic Tiering PoC ↗](#)
- [SAP HANA 2.0 SPS 02 dynamic tiering enhancements ↗](#)

## Operations for deploying SAP HANA on Azure VMs

The following sections describe some of the operations related to deploying SAP HANA systems on Azure VMs.

### Back up and restore operations on Azure VMs

The following documents describe how to back up and restore your SAP HANA deployment:

- [SAP HANA backup overview](#)
- [SAP HANA file-level backup](#)
- [SAP HANA storage snapshot benchmark](#)

### Start and restart VMs that contain SAP HANA

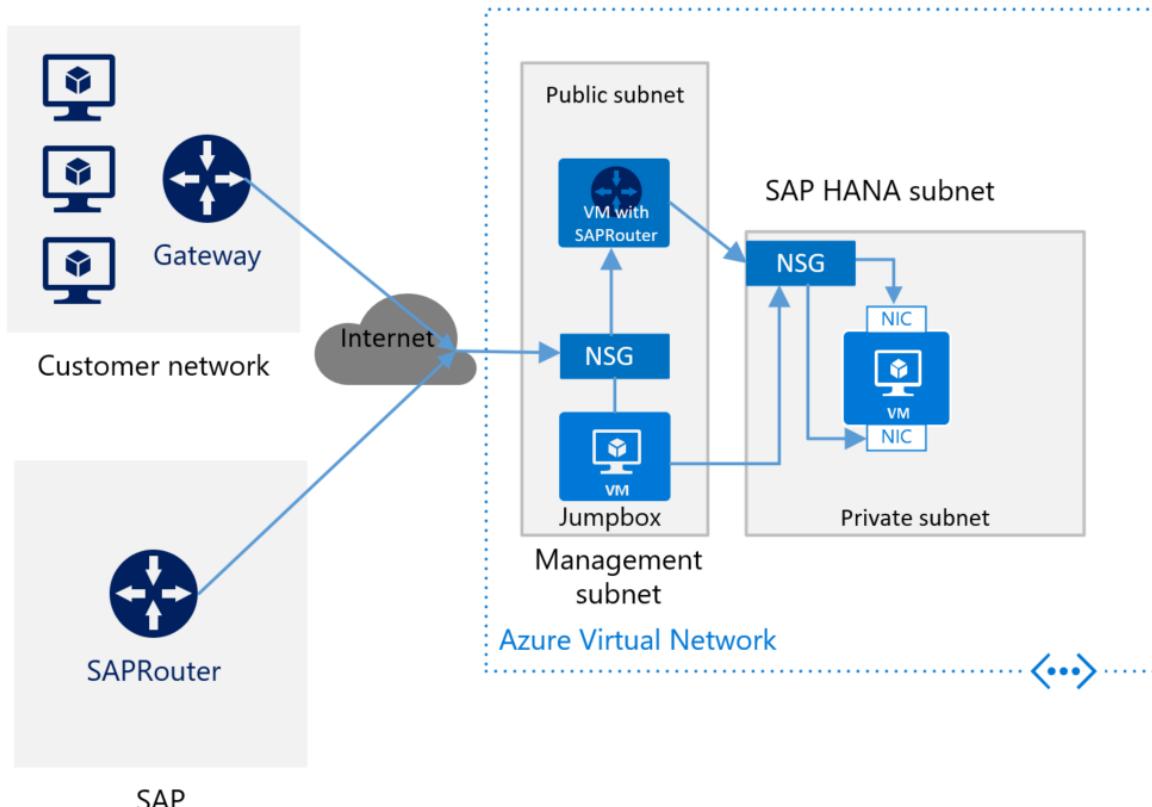
A prominent feature of the Azure public cloud is that you're charged only for your computing minutes. For example, when you shut down a VM that is running SAP HANA, you're billed only for the storage costs during that time. Another feature is available when you specify static IP addresses for your VMs in your initial deployment. When you restart a VM that has SAP HANA, the VM restarts with its prior IP addresses.

## Use SAProuter for SAP remote support

If you have a site-to-site connection between your on-premises locations and Azure, and you're running SAP components, then you're probably already running SAProuter. In this case, complete the following items for remote support:

- Maintain the private and static IP address of the VM that hosts SAP HANA in the SAProuter configuration.
- Configure the NSG of the subnet that hosts the HANA VM to allow traffic through TCP/IP port 3299.

If you're connecting to Azure through the internet, and you don't have an SAP router for the VM with SAP HANA, then you need to install the component. Install SAProuter in a separate VM in the Management subnet. The following image shows a rough schema for deploying SAP HANA without a site-to-site connection and with SAProuter:



Be sure to install SAPProuter in a separate VM and not in your Jumpbox VM. The separate VM must have a static IP address. To connect your SAPProuter to the SAPProuter that is hosted by SAP, contact SAP for an IP address. (The SAPProuter that is hosted by SAP is the counterpart of the SAPProuter instance that you install on your VM.) Use the IP address from SAP to configure your SAPProuter instance. In the configuration settings, the only necessary port is TCP port 3299.

For more information on how to set up and maintain remote support connections through SAPProuter, see the [SAP documentation](#).

## High-availability with SAP HANA on Azure native VMs

If you're running SUSE Linux Enterprise Server or Red Hat, you can establish a Pacemaker cluster with fencing devices. You can use the devices to set up an SAP HANA configuration that uses synchronous replication with HANA System Replication and automatic failover. For more information listed in the 'next steps' section.

## Next Steps

Get familiar with the articles as listed

- [SAP HANA Azure virtual machine storage configurations](#)
- [Deploy a SAP HANA scale-out system with standby node on Azure VMs by using Azure NetApp Files on SUSE Linux Enterprise Server](#)
- [Deploy a SAP HANA scale-out system with standby node on Azure VMs by using Azure NetApp Files on Red Hat Enterprise Linux](#)
- [Deploy a SAP HANA scale-out system with HSR and Pacemaker on Azure VMs on SUSE Linux Enterprise Server](#)
- [Deploy a SAP HANA scale-out system with HSR and PAcemaker on Azure VMs on Red Hat Enterprise Linux](#)
- [High availability of SAP HANA on Azure VMs on SUSE Linux Enterprise Server](#)
- [High availability of SAP HANA on Azure VMs on Red Hat Enterprise Linux](#)

# Inbound and outbound internet connections for SAP on Azure

Azure Virtual Machines

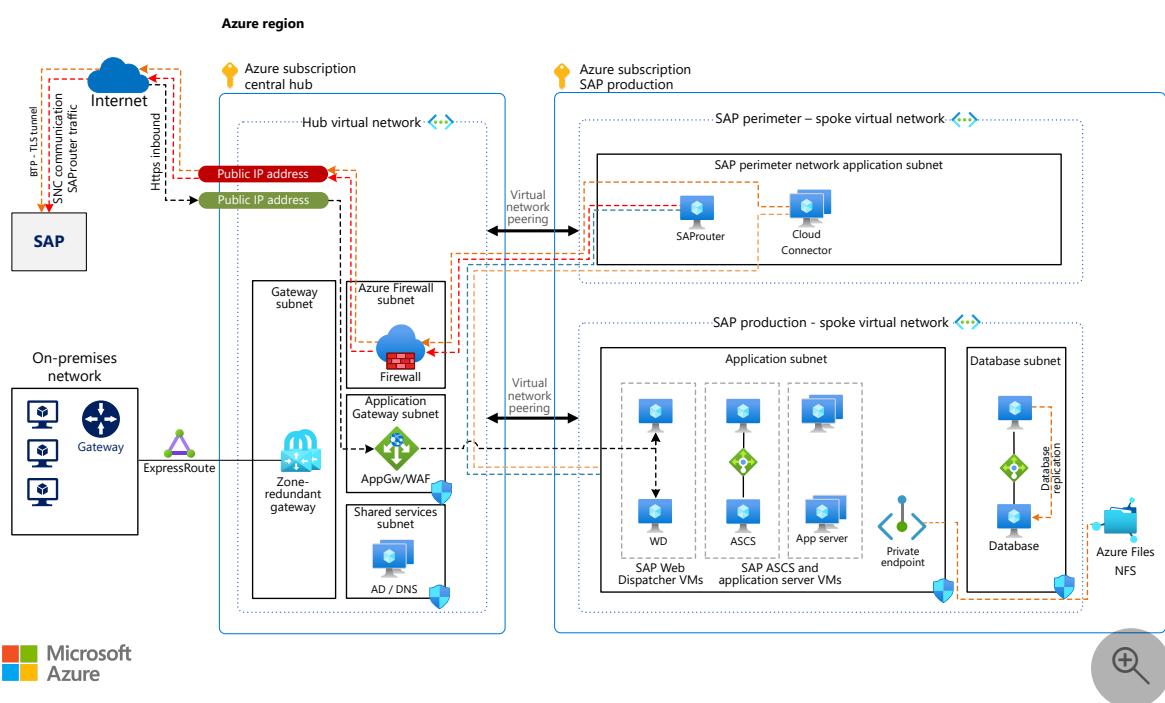
Azure Virtual Network

Azure Application Gateway

Azure Load Balancer

This article provides a set of proven practices for improving the security of inbound and outbound internet connections for your SAP on Azure infrastructure.

## Architecture



Download a [Visio file](#) of the architectures in this article.

This solution illustrates a common production environment. You can reduce the size and scope of the configuration to fit your requirements. This reduction might apply to the SAP landscape: fewer virtual machines (VMs), no high availability, or embedded SAP Web Dispatchers instead of discrete VMs. It can also apply to alternatives to the network design, as described later in this article.

Customer requirements, driven by business policies, will require adaptations to the architecture, particularly to the network design. When possible, we've included alternatives. Many solutions are viable. Choose an approach that's right for your business. It needs to help you secure your Azure resources but still provide a performant solution.

Disaster recovery (DR) isn't covered in this architecture. For the network design, the same principles and design that are valid for primary production regions apply. In your network design, depending on the applications being protected by DR, consider enabling DR in another Azure region. For more information, see the article [Disaster recovery overview and infrastructure guidelines for SAP workload](#)

## Workflow

- The on-premises network connects to a central hub via Azure ExpressRoute. The hub virtual network contains a gateway subnet, an Azure Firewall subnet, a shared services subnet, and an Azure Application Gateway subnet.
- The hub connects to an SAP production subscription via virtual network peering. This subscription contains two spoke virtual networks:
  - The SAP perimeter virtual network contains an SAP perimeter application subnet.
  - The SAP production virtual network contains an application subnet and a database subnet.
- The hub subscription and the SAP production subscription connect to the internet via public IP addresses.

## Components

**Subscriptions.** This architecture implements the Azure [landing zone](#) approach. One Azure subscription is used for each workload. One or more subscriptions are used for central IT services that contain the network hub and central, shared services like firewalls or Active Directory and DNS. Another subscription is used for the SAP production workload. Use the [decision guide](#) in the Cloud Adoption Framework for Azure to determine the best subscription strategy for your scenario.

**Virtual networks.** [Azure Virtual Network](#) connects Azure resources to each other with enhanced security. In this architecture, the virtual network connects to an on-premises environment via an ExpressRoute or virtual private network (VPN) gateway that's deployed in the hub of a [hub-spoke topology](#). The SAP production landscape uses its own spoke virtual networks. Two distinct spoke virtual networks perform different tasks, and subnets provide network segregation.

Separation into subnets by workload makes it easier to use network security groups (NSGs) to set security rules for application VMs or Azure services that are deployed.

**Zone-redundant gateway.** A gateway connects distinct networks, extending your on-premises network to the Azure virtual network. We recommend that you use

[ExpressRoute](#) to create private connections that don't use the public internet. You can also use a [site-to-site](#) connection. You can deploy ExpressRoute or VPN gateways across zones to help avoid zone failures. See [Zone-redundant virtual network gateways](#) for an explanation of the differences between a zonal deployment and a zone-redundant deployment. For a zone deployment of the gateways, you need to use Standard SKU IP addresses.

**NSGs.** To restrict network traffic to and from the virtual network, create [NSGs](#) and assign them to specific subnets. Provide security for individual subnets by using workload-specific NSGs.

**Application security groups.** To define fine-grained network security policies in your NSGs based on workloads that are centered on applications, use [application security groups](#) instead of explicit IP addresses. By using application security groups, you can group VMs by purpose, for example, SAP SID. Application security groups help secure applications by filtering traffic from trusted segments of your network.

**Private endpoint.** Many Azure services operate as public services, by design accessible via the internet. To allow private access via your private network range, you can use private endpoints for some services. [Private endpoints](#) are network interfaces in your virtual network. They effectively bring the service into your private network space.

**Azure Application Gateway.** [Application Gateway](#) is a web-traffic load balancer. With its Web Application Firewall functionality, it's the ideal service to expose web applications to the internet with improved security. Application Gateway can service either public (internet) or private clients, or both, depending on the configuration.

In the architecture, Application Gateway, using a public IP address, allows inbound connections to the SAP landscape over HTTPS. Its back-end pool is two or more SAP Web Dispatcher VMs, accessed round-robin and providing high availability. The application gateway is a reverse proxy and web-traffic load balancer, but it doesn't replace the SAP Web Dispatcher. SAP Web Dispatcher provides application integration with your SAP systems and includes features that Application Gateway by itself doesn't provide. Client authentication, when it reaches the SAP systems, is performed by the SAP application layer natively or via single sign-on. When you enable Azure DDoS protection, consider using [DDoS network protection SKU](#) as you will see discounts for Application Gateway Web Application Firewall.

For optimal performance, enable [HTTP/2 support](#) for Application Gateway, [SAP Web Dispatcher](#), and SAP NetWeaver.

**Azure Load Balancer**. Azure [Standard Load Balancer](#) provides networking elements for the high-availability design of your SAP systems. For clustered systems, Standard

Load Balancer provides the virtual IP address for the cluster service, like ASCS/SCS instances and databases running on VMs. You also can use Standard Load Balancer to provide the IP address for the virtual SAP host name of non-clustered systems when [secondary IPs on Azure network cards](#) aren't an option. The use of Standard Load Balancer instead of Application Gateway to address outbound internet access is covered later in this article.

## Network design

The architecture uses two discrete virtual networks, both spoke virtual networks that are peered to the central hub virtual network. There's no spoke-to-spoke peering. A star topology is used, in which communication passes through the hub. The separation of networks helps to protect the applications from breaches.

An application-specific [perimeter network](#) (also known as a *DMZ*) contains the internet-facing applications, like SAProuter, SAP Cloud Connector, SAP Analytics Cloud Agent, and others. In the architecture diagram, the perimeter network is named *SAP perimeter - spoke virtual network*. Because of dependencies on SAP systems, the SAP team typically does the deployment, configuration, and management of these services. That's why these SAP perimeter services frequently aren't located in a central hub subscription and network. Often organizational challenges are due to such central hub placement of workload specific applications or services.

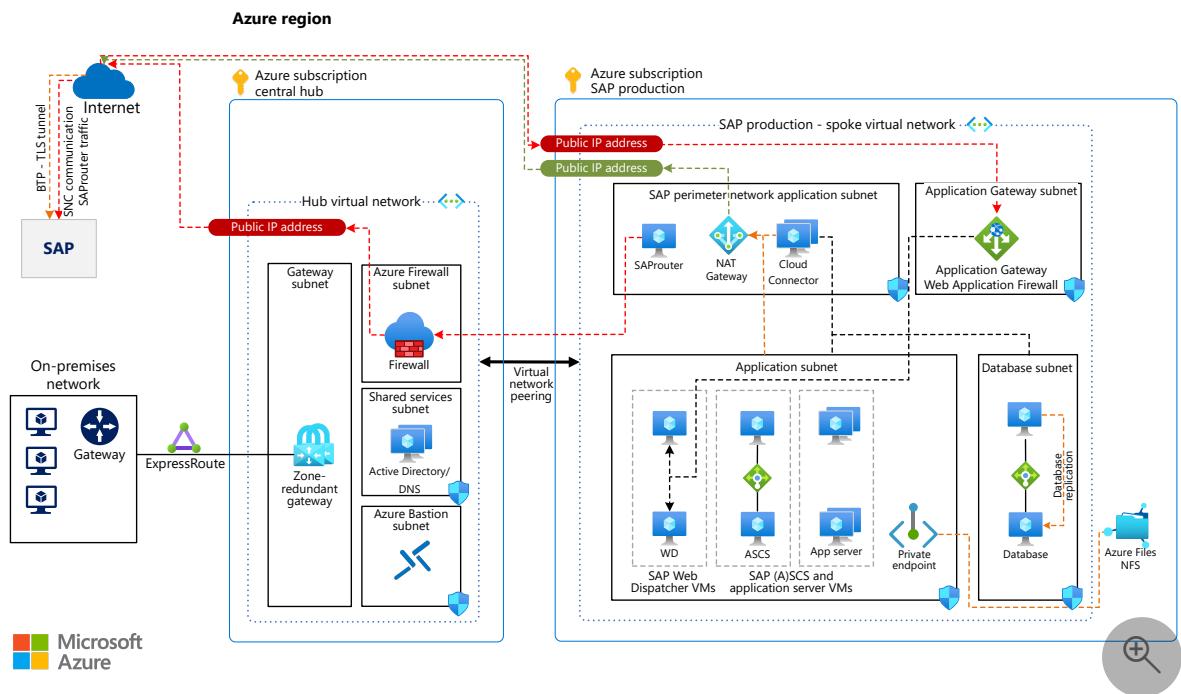
These are some of the benefits of using a separate SAP perimeter virtual network:

- Quick and immediate isolation of compromised services if a breach is detected. Removing virtual network peering from the SAP perimeter to the hub immediately isolates the SAP perimeter workloads and SAP application virtual network applications from the internet. Changing or removing an NSG rule that permits access affects only new connections and doesn't cut existing connections.
- More stringent controls on the virtual network and subnet, with a tight lockdown on communication partners in and out of the SAP perimeter network and SAP application networks. You can extend increased control to authorized users and access methods on SAP perimeter applications, with different authorization back ends, privileged access, or sign-in credentials for SAP perimeter applications.

The drawbacks are increased complexity and extra virtual network peering costs for internet-bound SAP traffic (because communication needs to pass through virtual network peering twice). The latency impact on spoke-hub-spoke peering traffic depends on any firewall that's in place and needs to be measured.

# Simplified architecture

To address the recommendations in this article but limit the drawbacks, you can use a single spoke virtual network for both the perimeter and the SAP applications. The following architecture contains all subnets in a single SAP production virtual network. The benefit of immediate isolation by termination of virtual network peering to the SAP perimeter if it's compromised isn't available. In this scenario, changes to NSGs affect only new connections.



Download a [Visio file](#) of the architectures in this article.

For deployments that are smaller in size and scope, the simplified architecture might be a better fit, and it still adheres to the principles of the more complex architecture. This article, unless otherwise noted, refers to the more complex architecture.

The simplified architecture uses a NAT gateway in the SAP perimeter subnet. This gateway provides outbound connectivity for SAP Cloud Connector and SAP Analytics Cloud Agent and OS updates for the deployed VMs. Because SAProuter requires both incoming and outbound connections, the SAProuter communication path goes through the firewall instead of using the NAT gateway. The simplified architecture also places the Application Gateway with its own designated subnet in the SAP perimeter virtual network, as an alternative approach to hub virtual network.

A [NAT gateway](#) is a service that provides static public IP addresses for outbound connectivity. The NAT gateway is assigned to a subnet. All outbound communications use the NAT gateway's IP addresses for internet access. Inbound connections don't use the NAT gateway. Applications like SAP Cloud Connector, or VM OS update services that

access repositories on the internet, can use the NAT gateway instead of routing all outbound traffic through the central firewall. Frequently, [user-defined rules](#) are implemented on all subnets to force internet-bound traffic from all virtual networks through the central firewall.

Depending on your requirements, you might be able to use the NAT gateway as an alternative to the central firewall, on outbound connections only. By doing so, you can reduce load on the central firewall while communicating with NSG-allowed public endpoints. You also get outbound IP control, because you can configure destination firewall rules on a set IP list of the NAT gateway. Examples include reaching Azure public endpoints that are used by public services, OS patch repositories, or third-party interfaces.

For a high-availability configuration, keep in mind that NAT gateway is deployed in a [specific zone only](#) and isn't currently cross-zone redundant. With a single NAT gateway it's not ideal for SAP deployments that use zone-redundant (cross-zone) deployment for virtual machines.

## Use of network components across an SAP landscape

An architecture document typically depicts only one SAP system or landscape. This makes them easier to understand. The result is that frequently the bigger picture, how the architecture fits into a larger SAP landscape that includes several system tracks and tiers, isn't addressed.

Central networking services, like the firewall, NAT gateway, and proxy servers if they're deployed, are best used across the entire SAP landscape of all tiers: production, pre-production, development, and sandbox. Depending on your requirements, the size of your organization, and business policies, you might want to consider separate implementations per tier, or one production and one sandbox/testing environment.

Services that typically serve an SAP system are best separated as described here:

- **Load balancers** should be dedicated to individual services. Company policy dictates naming and grouping of resources. We recommend one load balancer for ASCS/SCS and ERS and another for database, separated for each SAP SID. Alternatively, a single load balancer for both (A)SCS, ERS and DB clusters of one SAP system is also good design. This configuration helps to ensure that troubleshooting doesn't get complex, with many front-end and back-end pools and load balancing rules all on a single load balancer. A single load balancer per SAP SID also ensures that placement in resource groups matches that of other infrastructure components.

- **Application Gateway**, like a load balancer, allows multiple back ends, front ends, HTTP settings, and rules. The decision to use one application gateway for multiple uses is more common here because not all SAP systems in the environment require public access. Multiple uses in this context include different web dispatcher ports for same SAP S/4HANA systems or different SAP environments. We recommend at least one application gateway per tier (production, non-production, and sandbox) unless the complexity and number of connected systems becomes too high.
- **SAP services**, like SAProuter, Cloud Connector, and Analytics Cloud Agent, are deployed based on application requirements, either centrally or split up. Production and non-production separation is often desired.

## Subnet sizing and design

When you design subnets for your SAP landscape, be sure to follow sizing and design principles:

- Several Azure platform as a service (PaaS) services require their own designated subnets.
- Application Gateway recommends a /24 subnet for scaling. If choosing to limit the Application Gateway scale a smaller subnet could be used, at the [minimum /26 or larger](#). You can't use both versions of Application Gateway (1 and 2) in the same subnet.
- If you use Azure NetApp Files for your NFS/SMB shares or database storage, a designated subnet is required. A /24 subnet is the default. Use your requirements to determine the [proper sizing](#).
- If you use SAP virtual host names, you need more address space in your SAP subnets, including the SAP perimeter.
- Central services like Azure Bastion or Azure Firewall, typically managed by a central IT team, require their own dedicated subnets of sufficient size.

By using dedicated subnets for SAP databases and applications, you can set NSGs to be more strict, which helps to protect both application types with their own sets of rules. You can then limit database access to SAP applications more easily, without needing to resort to application security groups for granular control. Separating your SAP application and database subnets also makes it easier to manage your security rules in NSGs.

## SAP services

## SAProuter

You can use SAProuter to enable third parties like SAP support or your partners to access your SAP system. SAProuter runs on one VM in Azure. Route permissions for using SAProuter are stored in a flat file called *saprouttab*. The *saprouttab* entries allow connection from any TCP/IP port to a network destination behind SAProuter, typically your SAP system VMs. Remote access by SAP support relies on SAProuter. The main architecture uses the design that's described earlier, with a SAProuter VM running within the designated SAP perimeter virtual network. Through virtual network peering, SAProuter then communicates with your SAP servers that run in their own spoke virtual network and subnets.

SAProuter is a tunnel to SAP or to your partners. This architecture describes the use of SAProuter with SNC use to establish an encrypted application tunnel (network layer 7) to SAP/partners. The use of IPSEC based tunnel is not covered in this architecture presently.

The following features help protect the communication path over the internet:

- Azure Firewall or a third-party NVA provides the public IP entry point into your Azure networks. Firewall rules limit communication to only authorized IP addresses. For your connection to SAP support, [SAP note 48243 - Integrating the SAProuter software into a firewall environment](#) documents the IP addresses of SAP routers.
- Like firewall rules, network security rules allow communication on SAProuter's port, typically 3299 with the designated destination.
- You maintain SAProuter allow/deny rules in the *saprouttab* file, specifying who can contact SAProuter and which SAP system can be accessed.
- Further NSG rules are in place on the respective subnets in the SAP production subnet that contains the SAP systems.

For steps for configuring SAProuter with Azure Firewall, see [SAProuter configuration with Azure Firewall](#).

## SAProuter security considerations

Because SAProuter doesn't operate in the same application subnet as your SAP systems, sign-in mechanisms for the OS might be different. Depending on your policies, you can use a separate sign-on domain or entirely host-only user credentials for SAProuter. If there's a security breach, cascading access to the internal SAP systems isn't possible because of the different credential base. Network separation in such a case, as described earlier, can decouple further access from a compromised SAProuter into your

application subnets. You can accomplish this isolation by disconnecting the SAP perimeter virtual network peering.

## SAProuter high availability considerations

Because SAProuter is a simple executable file with a file-based route permission table, it can be easily started. The application has no built-in high availability. If there's a VM or application failure, the service needs to start on another VM. Using a virtual host name for the SAProuter service is ideal. The virtual host name is bound to an IP, which is assigned as a secondary IP config with the VM's NIC or to an internal load balancer that's connected to the VM. In this configuration, if the SAProuter service needs to be moved to another VM, the service virtual host name's IP config can be removed. You then add the virtual host name on another VM without needing to change the route tables or firewall configuration. They're all configured to use the virtual IP address. For more information, see [Use SAP Virtual Host Names with Linux in Azure](#).

## Cascading SAProuters

To implement cascading SAProuters, you can define as many as two SAProuters for SAP support connections. The first SAProuter, running in the SAP perimeter application subnet, provides access from the central firewall and from SAP or partner SAProuters. The only allowed destinations are other SAProuters, running with specific workloads. Cascading SAProuters can use per-tier, per-region, or per-SID separation, depending on your architecture. The second SAProuter accepts only internal connections from the first SAProuter and provides access to individual SAP systems and VMs. This design allows you to separate access and management between different teams if you need to. For an example of a cascading SAProuters, see [SAProuter configuration with Azure Firewall](#).

## SAP Fiori and WebGui

SAP Fiori and other HTTPS front ends for SAP applications are often consumed from outside of the internal corporate network. The need to be available on the internet requires a high-security solution to help protect the SAP application. Application Gateway with [Web Application Firewall](#) is the ideal service for this purpose.

For users who access the public host name of the public IP that's tied to Application Gateway, the HTTPS session is terminated on Application Gateway. A back-end pool of two or more SAP Web Dispatcher VMs gets round-robin sessions from the Application Gateway. This internal traffic application gateway to Web Dispatcher can be either HTTP or HTTPS, depending on requirements. The web application firewall helps to protect the SAP Web Dispatcher from attacks coming over the internet with the [OWASP core rule](#)

set. SAP NetWeaver, often tied to Microsoft Entra ID via [single sign-on \(SSO\)](#), performs user authentication. For the steps needed to configure SSO for Fiori by using Application Gateway, see [Single Sign On Configuration using SAML and Microsoft Entra ID for Public and Internal URLs](#).

Keep in mind that you need to secure the SAP Web Dispatcher in any situation. Even if it's open only internally, open toward Application Gateway via public IP, or accessible through any other network means. For more information, see [Security Information for SAP Web Dispatcher](#).

## Azure Firewall and Application Gateway

All web traffic provided by Application Gateway is HTTPS-based and encrypted with the provided TLS certificate. You can use Azure Firewall as an entry point to the corporate network, via its public IP, and then route SAP Fiori traffic from the firewall to Application Gateway through an internal IP address. For more information, see [Application Gateway after firewall](#). Because the TCP/IP layer-7 encryption is already in place via TLS, there's limited benefit to using a firewall in this scenario, and you can't perform packet inspection. Fiori communicates through the same external IP address for both inbound and outbound traffic, which typically isn't required for SAP Fiori deployments.

There are some benefits of a tandem Application Gateway and layer-4 firewall deployment:

- Possible integration with enterprise-wide security policy management.
- Network traffic that violates [security rules](#) is already discarded, so it doesn't require inspection.

This combined deployment is a good architecture. The method for handling inbound internet traffic depends on your overall enterprise architecture. You also need to consider how the overall network architecture fits with access methods from the internal IP address space, like on-premises clients. This consideration is covered in the next section.

## Application Gateway for internal IP addresses (optional)

This architecture focuses on internet-facing applications. There are various options available for clients that access SAP Fiori, the web UI of an SAP NetWeaver system, or another SAP HTTPS interface through an internal, private IP address. One scenario is treating all access to Fiori as public access, through the public IP. Another option is using direct network access through the private network to the SAP Web Dispatchers, bypassing Application Gateway entirely. A third option is to use [both private and public](#)

IP addresses on Application Gateway, providing access to both the internet and the private network.

You can use a similar configuration with a private IP address on Application Gateway for private-only network access to the SAP landscape. The public IP address in this case is used only for management purposes and doesn't have a listener associated with it.

As an alternative to using Application Gateway, you can use a load balancer internally. You can use a standard internal load balancer with Web Dispatcher VMs configured as a round-robin back end. In this scenario, the standard load balancer is placed with the Web Dispatcher VMs in the SAP production application subnet and provides [active/active](#) load balancing between Web Dispatcher VMs.

For internet-facing deployments, we recommend Application Gateway with Web Application Firewall instead of a load balancer with a public IP.

## SAP Business Technology Platform (BTP)

SAP BTP is a large set of SAP applications, SaaS or PaaS, typically accessed through a public endpoint via the internet. [SAP Cloud Connector](#) is often used to provide communication for applications running in private networks, like an SAP S/4HANA system running on Azure. SAP Cloud Connector runs as an application in a VM. It requires outbound internet access to establish a TLS-encrypted HTTPS tunnel with SAP BTP. It acts as a reverse invoke proxy between the private IP range in your virtual network and SAP BTP applications. Because of this reverse invoke support, there's no need for open firewall ports or other access for inbound connections, because the connection from your virtual network is outbound.

By default, VMs have [outbound internet](#) access natively on Azure. The public IP address that's used for outbound traffic, when there's no dedicated public IP address associated with the virtual machine, is randomly chosen from the pool of public IPs in the specific Azure region. You can't control it. To ensure outbound connections are made through a controlled and identifiable service and IP address, you can use one of the following methods:

- A NAT gateway that's associated with the subnet or load balancer and its public IP address.
- HTTP proxy servers that you operate.
- A [user-defined route](#) that forces the network traffic to flow to a network appliance like a firewall.

The architecture diagram shows the most common scenario: routing internet-bound traffic to the hub virtual network and through the central firewall. You need to configure

further settings [↗](#) in SAP Cloud Connector to connect to your SAP BTP account.

## High availability for SAP Cloud Connector

High availability is built into SAP Cloud Connector. Cloud Connector is installed on two VMs. The main instance is active, and the shadow instance is connected to it. They share configuration and are kept in sync natively. If the main instance isn't available, the secondary VM attempts to take over the main role and re-establish the TLS tunnel to SAP BTP. A high-availability Cloud Connector environment is shown in the architecture. You don't need any other Azure technologies, like a load balancer or cluster software, for the configuration. For details on configuration and operation, see the [SAP documentation ↗](#).

## SAP Analytics Cloud Agent

For some application scenarios, SAP Analytics Cloud Agent is an application that's installed in a VM. It uses SAP Cloud Connector for SAP BTP connectivity. In this architecture, the SAP Analytics Cloud Agent VM runs in the SAP perimeter application subnet, alongside the SAP Cloud Connector VMs. For the traffic flow from private networks like an Azure virtual network to SAP BTP via SAP Analytics Cloud Agent, see the [SAP documentation ↗](#).

## SAP Private Link service on Azure

SAP provides [Private Link service ↗](#) for SAP BTP. It enables private connections between selected SAP BTP services and selected services in your Azure subscription and virtual network. When you use Private Link service, the communication isn't routed through the public internet. It remains on the high-security Azure global network backbone. Communication to Azure services occurs via a private address space. Improved data exfiltration protection is built in when you use Private Link service, because the private endpoint maps the specific Azure service to an IP address. Access is limited to the mapped Azure service.

For some SAP BTP integration scenarios, the Private Link service approach is preferred. For others, SAP Cloud Connector is better. For information to help you decide which to use, see [Running Cloud Connector and SAP Private Link side-by-side ↗](#).

## SAP RISE/ECS

If SAP operates your SAP system under an SAP RISE/ECS contract, SAP is the managed service partner. The SAP environment is deployed by SAP. On SAP's architecture, the

architecture shown here doesn't apply to your systems that run in RISE with SAP/ECS. For information about integrating this type of SAP landscape with Azure services and your network, see the [Azure documentation](#).

## Other SAP communication requirements

Additional considerations regarding internet-bound communications might apply to an SAP landscape operating on Azure. Traffic flow in this architecture uses a central Azure firewall for this outbound traffic. User-defined rules in the spoke virtual networks route internet-bound traffic requests to the firewall. Alternatively, you can use NAT gateways on specific subnets, [default Azure outbound](#) communication, public IP addresses on VMs (not recommended), or a public load balancer with outbound rules.

For VMs that are behind a standard internal load balancer, like those in clustered environments, keep in mind that Standard Load Balancer modifies the behavior for public connectivity. For more information, see article [Public endpoint connectivity for VMs using Azure Standard Load Balancer in SAP high-availability scenarios](#).

## Operating system updates

Operating system updates are often located behind a public endpoint and accessed via the internet. If no enterprise repository and update management are in place, mirroring OS updates from vendors on private IP addresses / VMs, your SAP workload needs to access the update repositories of the vendors.

For Linux operating systems, you can access the following repositories if you obtain the OS license from Azure. If you purchase licenses directly and bring them to Azure (BYOS), contact the OS vendor about ways to connect to OS repositories and their respective IP address ranges.

- For SUSE Enterprise Linux, [SUSE maintains](#) a list of servers in each Azure region.
- For Red Hat Enterprise Linux, [Red Hat Update Infrastructure is documented here](#).
- For Windows, Windows Update is available via [FQDN tags](#) for Azure Firewall.

## High-availability cluster management

Highly available systems like clustered SAP ASCS/SCS or databases might use a cluster manager with Azure fence agent as a STONITH device. These systems depend on reaching Azure Resource Manager. Resource Manager is used for status queries about Azure resources and for operations to stop and start VMs. Because Resource Manager is a public endpoint, available under [management.azure.com](https://management.azure.com), VM outbound

communication needs to be able to reach it. This architecture relies on a central firewall with user-defined rules routing traffic from SAP virtual networks. For alternatives, see the preceding sections.

## Contributors

*This article is maintained by Microsoft. It was originally written by the following contributors.*

Principal authors:

- [Robert Biro](#) | Senior Architect
- [Dennis Padia](#) | Senior SAP Architect

Other contributor:

- [Mick Alberts](#) | Technical Writer

*To see non-public LinkedIn profiles, sign in to LinkedIn.*

## Communities

Consider using these communities to get answers to questions and for help with setting up a deployment:

- [Azure Community Support](#)
- [SAP Community](#)
- [Stack Overflow SAP](#)

## Next steps

- [SAP Blogs | SAP on Azure: Azure Application Gateway Web Application Firewall v2 Setup for Internet-facing SAP Fiori Apps](#)
- [SAP Blogs | Getting Started with BTP Private Link Service for Azure](#)
- [SAP Blogs | BTP private linky swear with Azure – running Cloud Connector and SAP Private Link side-by-side](#)
- [SAP on Azure Tech Community | SAProuter configuration with Azure Firewall](#)
- [SAP on Azure Tech Community | Use SAP Virtual Host Names with Linux in Azure](#)
- [SAP Documentation | What is Cloud Connector?](#)
- [SAP Documentation | What is SAP Analytics Cloud Agent?](#)
- [Default outbound access in Azure](#)

- Public endpoint connectivity for virtual machines using Azure Standard Load Balancer in SAP high-availability scenarios
- Subscription decision guide
- YouTube | Deploying Fiori at Scale ↗

## Related resources

- SAP workloads on Azure: planning and deployment checklist
- Run SAP NetWeaver in Windows on Azure
- SAP S/4HANA in Linux on Azure

# SAP workload configurations with Azure Availability Zones

Article • 06/01/2023

Additionally to the deployment of the different SAP architecture layers in Azure availability sets, [Azure Availability Zones](#) can be used for SAP workload deployments as well. An Azure Availability Zone is defined as: "Unique physical locations within a region. Each zone is made up of one or more datacenters equipped with independent power, cooling, and networking". Azure Availability Zones aren't available in all regions. For Azure regions that provide Availability Zones, check the [Azure region map](#). This map is going to show you which regions provide or are announced to provide Availability Zones.

As of the typical SAP NetWeaver or S/4HANA architecture, you need to protect three different layers:

- SAP application layer, which can be one to a few dozen VMs. You want to minimize the chance of VMs getting deployed on the same host server. You also want those VMs in an acceptable proximity to the DBMS layer to keep network latency in an acceptable window
- SAP ASCS/SCS layer that is representing a single point of failure in the SAP NetWeaver and S/4HANA architecture. You usually look at two VMs that you want to cover with a failover framework. Therefore, these VMs should be allocated in different infrastructure fault domains
- SAP DBMS layer, which represents a single point of failure as well. In the usual cases, it consists out of two VMs that are covered by a failover framework. Therefore, these VMs should be allocated in different infrastructure fault domains. Exceptions are SAP HANA scale-out deployments where more than two VMs are can be used

The major differences between deploying your critical VMs through availability sets or Availability Zones are:

- Deploying with an availability set is lining up the VMs within the set in a single zone or datacenter (whatever applies for the specific region). As a result the deployment through the availability set isn't protected by power, cooling or networking issues that affect the datacenter(s) of the zone as a whole. On the plus side, the VMs are aligned with update and fault domains within that zone or datacenter. Specifically for the SAP ASCS or DBMS layer where we protect two VMs

per availability set, the alignment with fault domains prevents that both VMs are ending up on the same host hardware.

- On deploying VMs through Azure Availability Zones and choosing different zones (maximum of three possible), is going to deploy the VMs across the different physical locations and with that adds protection from power, cooling or networking issues that affect the datacenter(s) of the zone as a whole. However, as you deploy more than one VM of the same VM family into the same Availability Zone, there's no protection from those VMs ending up on the same host or same fault domain. As a result, deploying through Availability Zones is ideal for the SAP ASCS and DBMS layer where we usually look at two VMs each. For the SAP application layer, which can be drastically more than two VMs, you might need to fall back to a different deployment model (see later)

Your motivation for a deployment across Azure Availability Zones should be that you, on top of covering failure of a single critical VM or ability to reduce downtime for software patching within a critical, want to protect from larger infrastructure issues that might affect the availability of one or multiple Azure datacenters.

As another resiliency deployment functionality, Azure introduced [Virtual machine scale sets with flexible orchestration](#) for SAP workload. Virtual machine scale set provides logical grouping of platform managed virtual machines. The flexible orchestration of virtual machine scale set provides the option to create the scale set within a region or span it across availability zones. On creating, the flexible scale set within a region with `platformFaultDomainCount>1` ( $FD > 1$ ), the VMs deployed in the scale set would be distributed across specified number of fault domains in the same region. On the other hand, creating the flexible scale set across availability zones with `platformFaultDomainCount=1` ( $FD = 1$ ) would distribute the virtual machines across different zones and the scale set would also distribute VMs across different fault domains within each zone on a best effort basis. **For SAP workload only flexible scale set with  $FD=1$  is supported.** The advantage of using flexible scale sets with  $FD=1$  for cross zonal deployment, instead of traditional availability zone deployment is that the VMs deployed with the scale set would be distributed across different fault domains within the zone in a best-effort manner. For more information, see [deployment guide of flexible scale set for SAP workload](#).

## Considerations for deploying across Availability Zones

Consider the following when you use Availability Zones:

- The maximum network roundtrip latency between Azure Availability Zones is stated in the document [Regions and availability zones](#).
- The experienced network roundtrip latency isn't necessarily indicative to the real geographical distance of the datacenters that form the different zones. The network roundtrip latency is also influenced by the cable connectivities and the routing of the cables between these different datacenters.
- Availability Zones aren't an ideal DR solution. Natural disasters can cause widespread damage in world regions, including heavy damage to power infrastructures. The distances between various zones might not be large enough to constitute a proper DR solution.
- The network latency across Availability Zones isn't the same in all Azure regions. In some cases, you can deploy and run the SAP application layer across different zones because the network latency from one zone to the active DBMS VM is acceptable. But in some Azure regions, the latency between the active DBMS VM and the SAP application instance, when deployed in different zones, might not be acceptable for SAP business processes. In these cases, the deployment architecture needs to be different, with an active/active architecture for the application, or an active/passive architecture where cross-zone network latency is too high.
- When deciding where to use Availability Zones, base your decision on the network latency between the zones. Network latency plays an important role in two areas:
  - Latency between the two DBMS instances that need to have synchronous replication. The higher the network latency, the more likely it affects the scalability of your workload.
  - The difference in network latency between a VM running an SAP dialog instance in-zone with the active DBMS instance and a similar VM in another zone. As this difference increases, the influence on the running time of business processes and batch jobs also increases, dependent on whether they run in-zone with the DBMS or in a different zone.

When you deploy Azure VMs across Availability Zones and establish failover solutions within the same Azure region, some restrictions apply:

- You must use [Azure Managed Disks](#) when you deploy to Azure Availability Zones.
- The mapping of zone enumerations to the physical zones is fixed on an Azure subscription basis. If you're using different subscriptions to deploy your SAP systems, you need to define the ideal zones for each subscription. If you want to compare the logical mapping of your different subscriptions, consider the [Avzone-Mapping script](#)
- You can't deploy Azure availability sets within an Azure Availability Zone unless you use [Azure Proximity Placement Group](#). The way how you can deploy the SAP DBMS

layer and the central services across zones and at the same time deploy the SAP application layer using availability sets and still achieve close proximity of the VMs is documented in the article [Azure Proximity Placement Groups for optimal network latency with SAP applications](#). If you aren't using Azure proximity placement groups, you need to choose one or the other as a deployment framework for virtual machines.

- You can't use an [Azure Basic Load Balancer](#) to create failover cluster solutions based on Windows Server Failover Clustering or Linux Pacemaker. Instead, you need to use the [Azure Standard Load Balancer SKU](#).

## The ideal Availability Zones combination

If you want to deploy an SAP NetWeaver or S/4HANA system across zones, there are two architecture patterns you can deploy:

- Active/active: The pair of VMs running ASCS/SCS and the pair of VMS running the DBMS layer are distributed across two zones. The number of VMs running the SAP application layer are deployed in even numbers across the same two zones. If a DBMS or ASCS/SCS VM is failing over, some of the open and active transactions might be rolled back. But users are remaining logged in. It doesn't really matter in which of the zones the active DBMS VM and the application instances run. This architecture is the preferred architecture to deploy across zones.
- Active/passive: The pair of VMs running ASCS/SCS and the pair of VMS running the DBMS layer are distributed across two zones. The number of VMs running the SAP application layer are deployed into one of the Availability Zones. You run the application layer in the same zone as the active ASCS/SCS and DBMS instance. You use this deployment architecture if the network latency across the different zones is too high to run the application layer distributed across the zones. Instead the SAP application layer needs to run in the same zone as the active ASCS/SCS and/or DBMS instance. If an ASCS/SCS or DBMS VM fails over to the secondary zone, you might encounter higher network latency and with that a reduction of throughput. And you're required to fail back the previously failed over VM as soon as possible to get back to the previous throughput levels. If a zonal outage occurs, the application layer needs to be failed over to the secondary zone. An activity that users experience as complete system shutdown. In some of the Azure regions, this architecture is the only viable architecture when you want to use Availability Zones. If you can't accept the potential impact of an ASCS/SCS or DBMS VMS failing over to the secondary zone, you might be better off staying with availability set deployments

So before you decide how to use Availability Zones, you need to determine:

- The network latency among the three zones of an Azure region. Knowing the network latency between the zones of a region is going to enable you to choose the zones with the least network latency in cross-zone network traffic.
- The difference between VM-to-VM latency within one of the zones, of your choosing, and the network latency across two zones of your choosing.
- A determination of whether the VM types that you need to deploy are available in the two zones that you selected. With some VMs SKUs, you might encounter situations in which some SKUs are available in only two of the three zones.

## Network latency between and within zones

To determine the latency between the different zones, you need to:

- Deploy the VM SKU you want to use for your DBMS instance in all three zones. Make sure [Azure Accelerated Networking](#) is enabled when you take this measurement. Accelerated Networking is the default setting since a few years. Nevertheless, check whether it's enabled and working
- When you find the two zones with the least network latency, deploy another three VMs of the VM SKU that you want to use as the application layer VM across the three Availability Zones. Measure the network latency against the two DBMS VMs in the two DBMS zones that you selected.
- Use `niping` as a measuring tool. This tool, from SAP, is described in SAP support notes [#500235](#) and [#1100926](#). Focus on the commands documented for latency measurements. Because `ping` doesn't work through the Azure Accelerated Networking code paths, we don't recommend that you use it.

You don't need to perform these tests manually. You can find a PowerShell procedure [Availability Zone Latency Test](#) that automates the latency tests described.

Based on your measurements and the availability of your VM SKUs in the Availability Zones, you need to make some decisions:

- Define the ideal zones for the DBMS layer.
- Determine whether you want to distribute your active SAP application layer across one, two, or all three zones, based on differences of network latency in-zone versus across zones.
- Determine whether you want to deploy an active/passive configuration or an active/active configuration, from an application point of view. (These configurations are explained later in this article.)

In making these decisions, also take into account SAP's network latency recommendations, as documented in SAP note [#1100926](#).

### **ⓘ Important**

The measurements and decisions you make are valid for the Azure subscription you used when you took the measurements. If you use another Azure subscription, the mapping of enumerated zones might be different for another Azure subscription. As a result, you need to repeat the measurements or find out the mapping of the new subscription relative to the old subscription the tool [Avzone-Mapping script](#).

### **ⓘ Important**

It's expected that the measurements described earlier will provide different results in every Azure region that supports [Availability Zones](#). Even if your network latency requirements are the same, you might need to adopt different deployment strategies in different Azure regions because the network latency between zones can be different. In some Azure regions, the network latency among the three different zones can be vastly different. In other regions, the network latency among the three different zones might be more uniform. The claim that there's always a network latency between 1 and 2 milliseconds isn't correct. The network latency across Availability Zones in Azure regions can't be generalized.

## **Active/Active deployment**

This deployment architecture is called active/active because you deploy your active SAP application servers across two or three zones. The SAP Central Services instance that uses enqueue replication will be deployed between two zones. The same is true for the DBMS layer, which will be deployed across the same zones as SAP Central Service. When considering this configuration, you need to find the two Availability Zones in your region that offer cross-zone network latency that's acceptable for your workload and your synchronous DBMS replication. You also want to be sure the delta between network latency within the zones you selected and the cross-zone network latency isn't too large.

Nature of the SAP architecture is that, unless you configure it differently, users and batch jobs can be executed in the different application instances. The side effect of this fact with the active/active deployment is that batch jobs might be executed by any SAP application instances independent on whether those run in the same zone with the active DBMS or not. If the difference in network latency between the difference zones is small compared to network latency within a zone, the difference in run times of batch

jobs might not be significant. However, the larger the difference of network latency within a zone, compared to across zone network traffic is, the run time of batch jobs can be impacted more if the job got executed in a zone where the DBMS instance isn't active. It's on you as a customer to decide what acceptable differences in run time are. And with that what the tolerable network latency for cross zones traffic is for your workload.

Azure regions where such an active/active deployment could be possible without significant large differences in run time and throughput within the application layer deployed across different Availability Zones, list like:

- Australia East (two of the three zones)
- Brazil South (all three zones)
- Central India (all three zones)
- Central US (all three zones)
- East Asia (all three zones)
- East US (two of the three zones)
- East US2 (all three zones)
- Germany West Central (all three zones)
- Israel Central (all three zones)
- Italy North (two of the three zones)
- Korea Central (all three zones)
- Poland Central (all three zones)
- Qatar Central (all three zones)
- North Europe (all three zones)
- Norway East (two of the three zones)
- South Africa North (two of the three)
- South Central US (all three zones)
- Southeast Asia (all three zones)
- Sweden Central (all three zones)
- Switzerland North (all three zones)
- UAE North (all three zones)
- UK South (two of the three zones)
- West Europe (two of the three zones)
- West US2 (all three zones)
- West US3 (all three zones)

The region list provided doesn't relief you as a customer to test your workload to decide whether an active/active deployment architecture is possible.

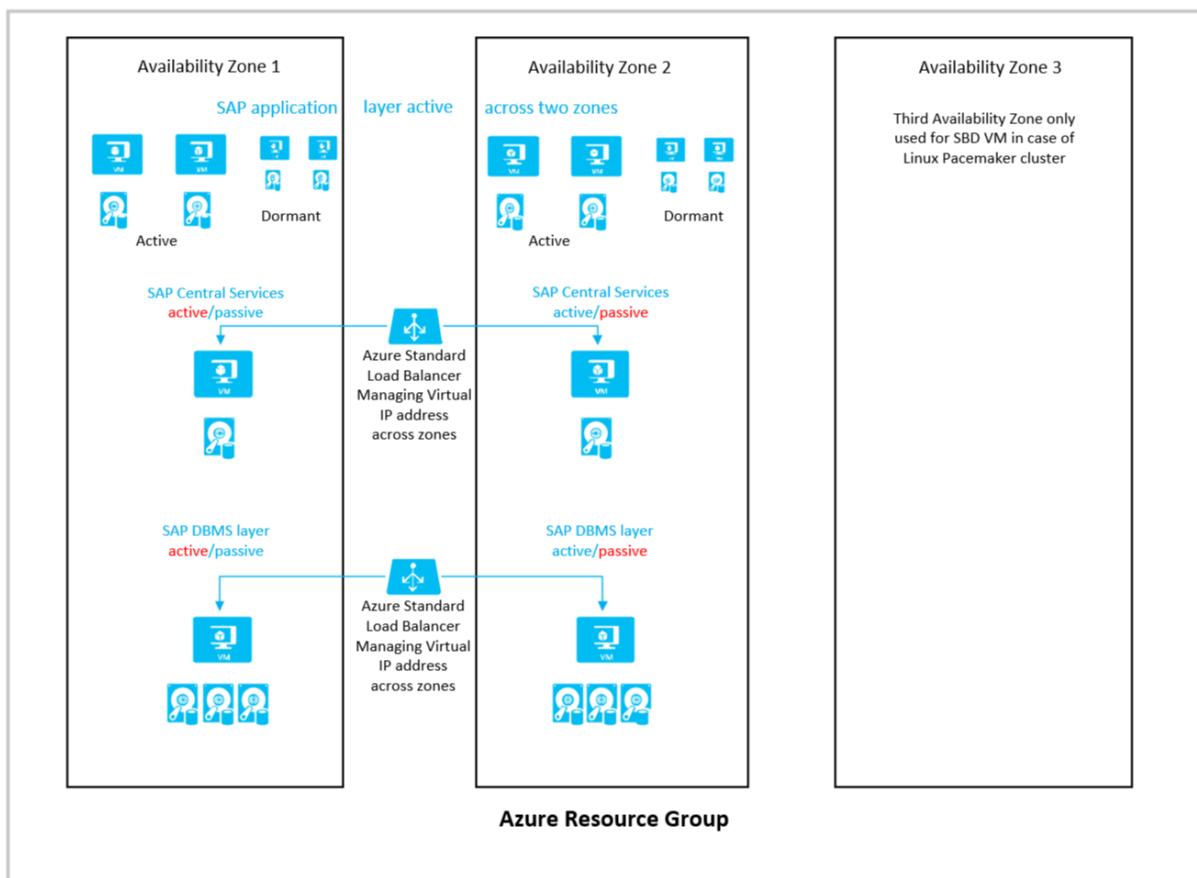
Azure regions where the active/active SAP deployment architecture across zones might not be possible, list like:

- Canada Central
- France Central
- Japan East

Though for your individual workload, it might work. Therefore, you should test before you decide for an architecture. Azure is constantly working to improve quality and latency of its networks. Measurements conducted years back might not reflect current conditions anymore.

Dependent on what you're willing to tolerate on run time differences other regions not listed could qualify as well.

A simplified schema of an active/active deployment across two zones could look like this:



The following considerations apply for this configuration:

- Not using [Azure Proximity Placement Group](#), you treat the Azure Availability Zones as fault domains for all the VMs because availability sets can't be deployed in Azure Availability Zones.
- If you want to combine zonal deployments for the DBMS layer and central services, but want to use Azure availability sets for the application layer, you need to use Azure proximity groups as described in the article [Azure Proximity Placement Groups for optimal network latency with SAP applications](#).

- For the load balancers of the failover clusters of SAP Central Services and the DBMS layer, you need to use the [Standard SKU Azure Load Balancer](#). The Basic Load Balancer won't work across zones.
- The Azure virtual network that you deployed to host the SAP system, together with its subnets, is stretched across zones. You don't need separate virtual networks and subnets for each zone.
- For all virtual machines you deploy, you need to use [Azure Managed Disks](#). Unmanaged disks aren't supported for zonal deployments.
- Azure Premium Storage, [Ultra SSD storage](#), or ANF don't support any type of storage replication across zones. For DBMS deployments, we rely on database methods to replicate data across zones
- For SMB and NFS shares based on [Azure Premium Files](#), zonal redundancy with synchronous replication is offered. Check [this document](#) for availability of ZRS for Azure Premium Files in the region you want to deploy into. The usage of zonal replicated NFS and SMB shares is fully supported with SAP application layer deployments and high availability failover clusters for NetWeaver or S/4HANA central services. Documents that cover these cases are:
  - [High availability for SAP NetWeaver on Azure VMs on SUSE Linux Enterprise Server with NFS on Azure Files](#)
  - [Azure Virtual Machines high availability for SAP NetWeaver on Red Hat Enterprise Linux with Azure NetApp Files for SAP applications](#)
  - [High availability for SAP NetWeaver on Azure VMs on Windows with Azure Files Premium SMB for SAP applications](#)
- The third zone is used to host the SBD device if you build a [SUSE Linux Pacemaker cluster](#) and use SBD devices instead of the Azure Fencing Agent. Or for more application instances.
- To achieve run time consistency for critical business processes, you can try to direct certain batch jobs and users to application instances that are in-zone with the active DBMS instance by using SAP batch server groups, SAP logon groups, or RFC groups. However, in zonal failover process, you would need to manually move these groups to instances running on VMs that are in-zone with the active DB VM.
- You might want to deploy dormant dialog instances in each of the zones.

 **Important**

In this active/active scenario charges for cross zone traffic apply. Check the document [Bandwidth Pricing Details](#). The data transfer between the SAP application layer and SAP DBMS layer is quite intensive. Therefore the active/active scenario can contribute to costs.

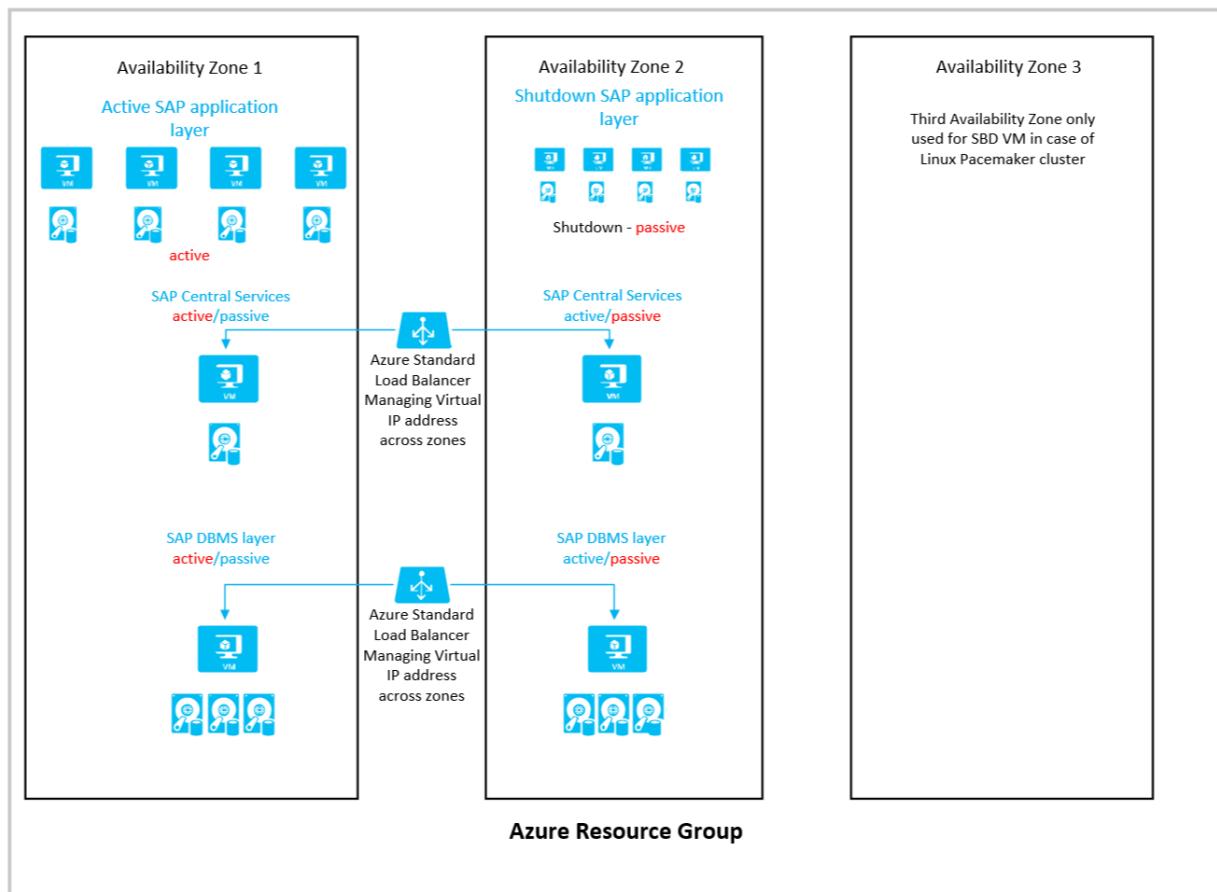
# Active/Passive deployment

If you can't find an acceptable delta between the network latency within one zone and the latency of cross-zone network traffic, you can deploy an architecture that has an active/passive character from the SAP application layer point of view. You define an *active* zone, which is the zone where you deploy the complete application layer and where you attempt to run both the active DBMS and the SAP Central Services instance. With such a configuration, you need to make sure you don't have extreme run time variations, depending on whether a job runs in-zone with the active DBMS instance or not, in business transactions and batch jobs.

Azure regions where this type of deployment architecture across different zones could be preferable are:

- Canada Central
- France Central
- Japan East
- Norway East
- South Africa North

The basic layout of the architecture looks like this:



The following considerations apply for this configuration:

- Availability sets can't be deployed in Azure Availability Zones. To mitigate, you can use Azure proximity placement groups as documented in the article [Azure Proximity Placement Groups for optimal network latency with SAP applications](#).
- When you use this architecture, you need to monitor the status closely and try to keep the active DBMS and SAP Central Services instances in the same zone as your deployed application layer. If there was a failover of SAP Central Service or the DBMS instance, you want to make sure that you can manually fail back into the zone with the SAP application layer deployed as quickly as possible.
- For the load balancers of the failover clusters of SAP Central Services and the DBMS layer, you need to use the [Standard SKU Azure Load Balancer](#). The Basic Load Balancer won't work across zones.
- The Azure virtual network that you deployed to host the SAP system, together with its subnets, is stretched across zones. You don't need separate virtual networks for each zone.
- For all virtual machines you deploy, you need to use [Azure Managed Disks](#). Unmanaged disks aren't supported for zonal deployments.
- Azure Premium Storage, [Ultra SSD storage](#), or ANF don't support any type of storage replication across zones. For DBMS deployments, we rely on database methods to replicate data across zones
- For SMB and NFS shares based on [Azure Premium Files](#), zonal redundancy with synchronous replication is offered. Check [this document](#) for availability of ZRS for Azure Premium Files in the region you want to deploy into. The usage of zonal replicated NFS and SMB shares is fully supported with SAP application layer deployments and high availability failover clusters for NetWeaver or S/4HANA centrals services. Documents that cover these cases are:
  - [High availability for SAP NetWeaver on Azure VMs on SUSE Linux Enterprise Server with NFS on Azure Files](#)
  - [Azure Virtual Machines high availability for SAP NetWeaver on Red Hat Enterprise Linux with Azure NetApp Files for SAP applications](#)
  - [High availability for SAP NetWeaver on Azure VMs on Windows with Azure Files Premium SMB for SAP applications](#)
- The third zone is used to host the SBD device if you build a [SUSE Linux Pacemaker cluster](#) and use SBD devices instead of the Azure Fencing Agent. Or for additional application instances.
- You should deploy dormant VMs in the passive zone (from a DBMS point of view) so you can start application resources for the case of a zone failure. Another possibility could be to use [Azure Site Recovery](#), which is able to replicate active VMs to dormant VMs between zones.
- You should invest in automation that allows you to automatically start the SAP application layer in the second zone if a zonal outage occurs.

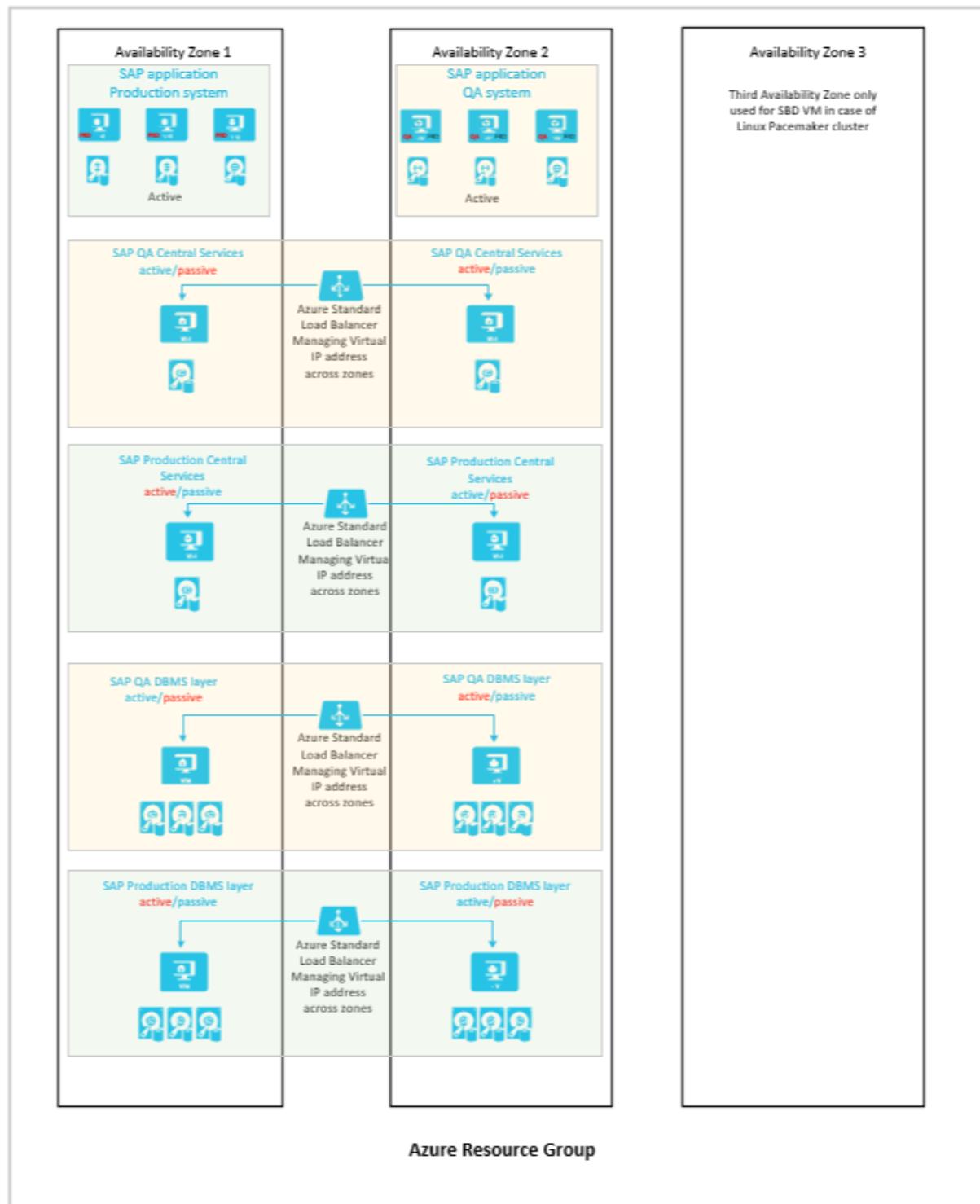
# Combined high availability and disaster recovery configuration

Microsoft doesn't share any information about geographical distances between the facilities that host different Azure Availability Zones in an Azure region. Still, some customers are using zones for a combined HA and DR configuration that promises a recovery point objective (RPO) of zero. An RPO of zero means that you shouldn't lose any committed database transactions even in disaster recovery cases.

## ⓘ Note

We recommend that you use a configuration like this only in certain circumstances. For example, you might use it when data can't leave the Azure region for security or compliance reasons.

Here's one example of how such a configuration might look:



The following considerations apply for this configuration:

- You're either assuming that there's a significant distance between the facilities hosting an Availability Zone or you're forced to stay within a certain Azure region. Availability sets can't be deployed in Azure Availability Zones. To compensate for that, you can use Azure proximity placement groups as documented in the article [Azure Proximity Placement Groups for optimal network latency with SAP applications](#).
- When you use this architecture, you need to monitor the status closely, and try to keep the active DBMS and SAP Central Services instances in the same zone as your

deployed application layer. If there was a failover of SAP Central Service or the DBMS instance, you want to make sure that you can manually fail back into the zone with the SAP application layer deployed as quickly as possible.

- You should have production application instances preinstalled in the VMs that run the active QA application instances.
- In a zonal failure case, shut down the QA application instances and start the production instances instead. You need to use virtual names for the application instances to make this work.
- For the load balancers of the failover clusters of SAP Central Services and the DBMS layer, you need to use the [Standard SKU Azure Load Balancer](#). The Basic Load Balancer won't work across zones.
- The Azure virtual network that you deployed to host the SAP system, together with its subnets, is stretched across zones. You don't need separate virtual networks for each zone.
- For all virtual machines you deploy, you need to use [Azure Managed Disks](#). Unmanaged disks aren't supported for zonal deployments.
- Azure Premium Storage, [Ultra SSD storage](#), or ANF don't support any type of storage replication across zones. For DBMS deployments, we rely on database methods to replicate data across zones
- For SMB and NFS shares based on [Azure Premium Files](#), zonal redundancy with synchronous replication is offered. Check [this document](#) for availability of ZRS for Azure Premium Files in the region you want to deploy into. The usage of zonal replicated NFS and SMB shares is fully supported with SAP application layer deployments and high availability failover clusters for NetWeaver or S/4HANA central services. Documents that cover these cases are:
  - [High availability for SAP NetWeaver on Azure VMs on SUSE Linux Enterprise Server with NFS on Azure Files](#)
  - [Azure Virtual Machines high availability for SAP NetWeaver on Red Hat Enterprise Linux with Azure NetApp Files for SAP applications](#)
  - [High availability for SAP NetWeaver on Azure VMs on Windows with Azure Files Premium SMB for SAP applications](#)
- The third zone is used to host the SBD device if you build a [SUSE Linux Pacemaker cluster](#) and use SBD devices instead of the Azure Fencing Agent.

## Next steps

Here are some next steps for deploying across Azure Availability Zones:

- Cluster an SAP ASCS/SCS instance on a Windows failover cluster by using a cluster shared disk in Azure

- Prepare Azure infrastructure for SAP high availability by using a Windows failover cluster and file share for SAP ASCS/SCS instances

# Supported scenarios for HANA Large Instances

Article • 02/10/2023

This article describes the supported scenarios and architectural details for HANA Large Instances (HLI).

## ⓘ Note

If your scenario isn't mentioned in this article, contact the Microsoft Service Management team to assess your requirements. Before you set up the HLI unit, validate the design with SAP or your service implementation partner.

## Terms and definitions

Let's understand the terms and definitions used in this article:

- **SID:** A system identifier for the HANA system.
- **HLI:** Hana Large Instances.
- **DR:** Disaster recovery (DR).
- **Normal DR:** A system setup with a dedicated resource for DR purposes only.
- **Multipurpose DR:** A DR site system that's configured to use a non-production environment alongside a production instance that's configured for a DR event.
- **Single-SID:** A system with one instance installed.
- **Multi-SID:** A system with multiple instances configured; also called an MCOS environment.
- **HSR:** SAP HANA system replication.

## Overview

HANA Large Instances support various architectures to help you accomplish your business requirements. The following sections cover the architectural scenarios and their configuration details.

The derived architectural designs are purely from an infrastructure perspective. Consult SAP or your implementation partners for the HANA deployment. If your scenarios aren't listed in this article, contact the Microsoft account team to review the architecture and derive a solution for you.

## Note

These architectures are fully compliant with Tailored Data Integration (TDI) design and are supported by SAP.

This article describes the details of the two components in each supported architecture:

- Ethernet
- Storage

## Ethernet

Each provisioned server comes preconfigured with sets of Ethernet interfaces. The Ethernet interfaces configured on each HLI unit are categorized into four types:

- **A:** Used for or by client access.
- **B:** Used for node-to-node communication. This interface is configured on all servers no matter what topology you request. However, it's used only for scale-out scenarios.
- **C:** Used for node-to-storage connectivity.
- **D:** Used for node-to-iSCSI device connection for fencing setup. This interface is configured only when an HSR setup is requested.

 Expand table

NIC logical interface	SKU type	Name with SUSE OS	Name with RHEL OS	Use case
A	TYPE I	eth0.tenant	eno1.tenant	Client-to-HLI
B	TYPE I	eth2.tenant	eno3.tenant	Node-to-node
C	TYPE I	eth1.tenant	eno2.tenant	Node-to-storage
D	TYPE I	eth4.tenant	eno4.tenant	Fencing
A	TYPE II	vlan<tenantNo>	team0.tenant	Client-to-HLI
B	TYPE II	vlan<tenantNo+2>	team0.tenant+2	Node-to-node
C	TYPE II	vlan<tenantNo+1>	team0.tenant+1	Node-to-storage
D	TYPE II	vlan<tenantNo+3>	team0.tenant+3	Fencing

You choose the interface based on the topology that's configured on the HLI unit. For example, interface "B" is set up for node-to-node communication, which is useful when you have a scale-out topology configured. This interface isn't used for single node scale-up configurations. For more information about interface usage, review your required scenarios (later in this article).

If necessary, you can define more NIC cards on your own. However, the configurations of existing NICs *can't* be changed.

 **Note**

You might find additional interfaces that are physical interfaces or bonding. Consider only the previously mentioned interfaces for your use case. Ignore any others.

The distribution for units with two assigned IP addresses should look as follows:

- Ethernet "A" should have an assigned IP address that's within the server IP pool address range that you submitted to Microsoft. This IP address should be maintained in the */etc/hosts* directory of the operating system (OS).
- Ethernet "C" should have an assigned IP address that's used for communication to NFS. You *don't* need to maintain this address in the *etc/hosts* directory to allow instance-to-instance traffic within the tenant.

For HANA system replication or HANA scale-out deployment, a blade configuration with two assigned IP addresses isn't suitable. If you have only two assigned IP addresses, and you want to deploy such a configuration, contact SAP HANA on Azure Service Management. They can assign you a third IP address in a third VLAN. For HANA Large Instances with three assigned IP addresses on three NIC ports, the following usage rules apply:

- Ethernet "A" should have an assigned IP address that's outside of the server IP pool address range that you submitted to Microsoft. This IP address shouldn't be maintained in the *etc/hosts* directory of the OS.
- Ethernet "B" should be maintained exclusively in the *etc/hosts* directory for communication between the various instances. Maintain these IP addresses in scale-out HANA configurations as the IP addresses that HANA uses for the inter-node configuration.
- Ethernet "C" should have an assigned IP address that's used for communication to NFS storage. This type of address shouldn't be maintained in the *etc/hosts*

directory.

- Ethernet “D” should be used exclusively for access to fencing devices for Pacemaker. This interface is required when you configure HANA system replication and want to achieve auto failover of the operating system by using an SBD-based device.

## Storage

Storage is preconfigured based on the requested topology. The volume sizes and mount points vary depending on the number of servers and SKUs, and the configured topology. For more information, review your required scenarios (later in this article). If you require more storage, you can purchase it in 1-TB increments.

 **Note**

The mount point /usr/sap/<SID> is a symbolic link to the /hana/shared mount point.

## Supported scenarios

The architectural diagrams in the next sections use the following notations:

	Ethernet
	HDI Server
	Storage and mountpoints for active instance
	Storage and mountpoint for storage-based replication at the DR site
	Storage volumes and mountpoints for the HANA instance installed on the DR node or a standby node in HA scenario
	Gateway in Azure
	Express route

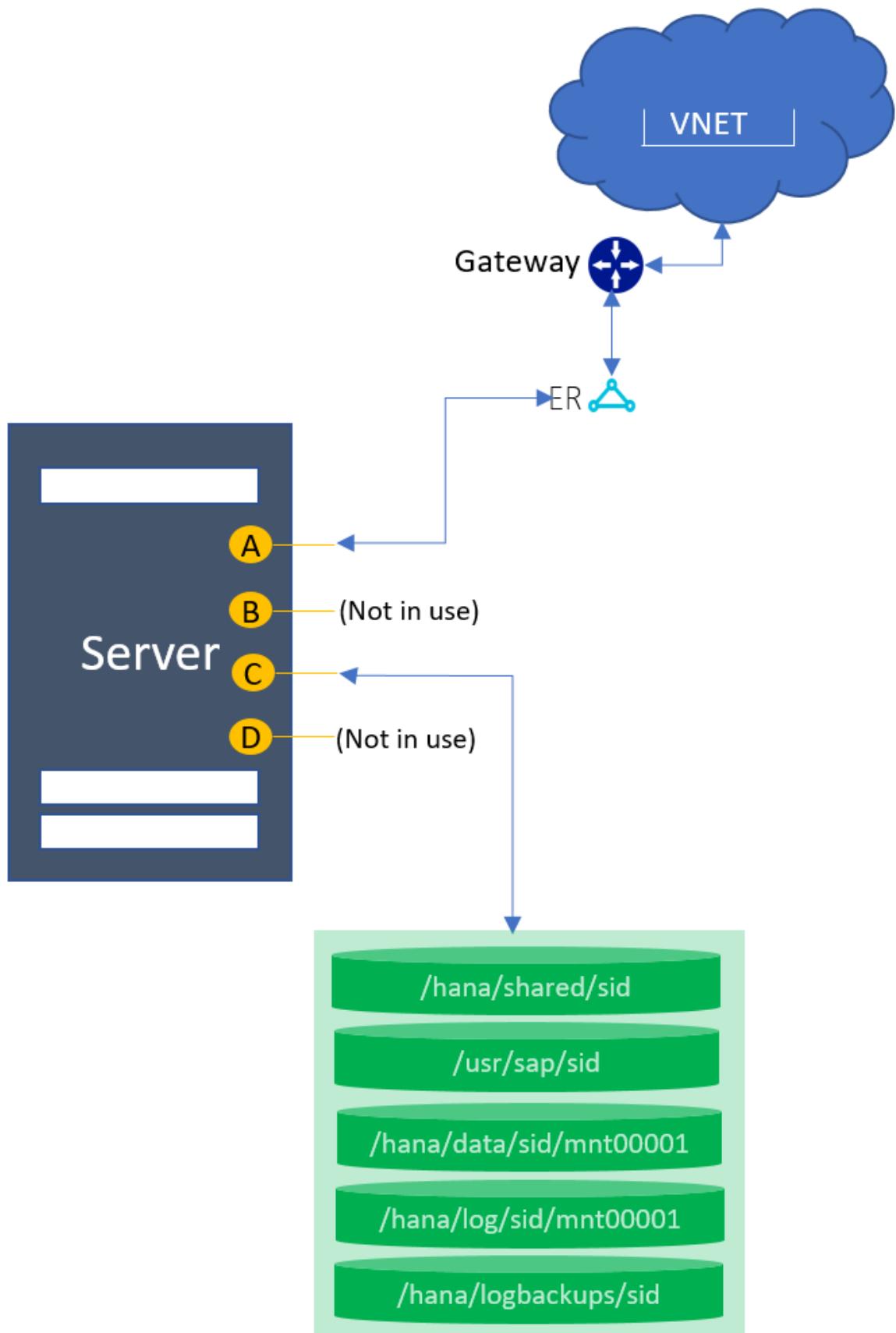
Here are the supported scenarios:

- Single node with one SID
- Single node MCOS
- Single node with DR (normal)
- Single node with DR (multipurpose)
- HSR with fencing
- HSR with DR (normal/multipurpose)
- Host auto failover (1+1)
- Scale-out with standby
- Scale-out without standby
- Scale-out with DR

## Single node with one SID

This topology supports one node in a scale-up configuration with one SID.

## Architecture diagram



## Ethernet

The following network interfaces are preconfigured:

[\[\] Expand table](#)

<b>NIC logical interface</b>	<b>SKU type</b>	<b>Name with SUSE OS</b>	<b>Name with RHEL OS</b>	<b>Use case</b>
A	TYPE I	eth0.tenant	eno1.tenant	Client-to-HLI
B	TYPE I	eth2.tenant	eno3.tenant	Configured but not in use
C	TYPE I	eth1.tenant	eno2.tenant	Node-to-storage
D	TYPE I	eth4.tenant	eno4.tenant	Configured but not in use
A	TYPE II	vlan<tenantNo>	team0.tenant	Client-to-HLI
B	TYPE II	vlan<tenantNo+2>	team0.tenant+2	Configured but not in use
C	TYPE II	vlan<tenantNo+1>	team0.tenant+1	Node-to-storage
D	TYPE II	vlan<tenantNo+3>	team0.tenant+3	Configured but not in use

## Storage

The following mount points are preconfigured:

[\[\] Expand table](#)

<b>Mount point</b>	<b>Use case</b>
/hana/shared/SID	HANA installation
/hana/data/SID/mnt00001	Data files installation
/hana/log/SID/mnt00001	Log files installation
/hana/logbackups/SID	Redo logs

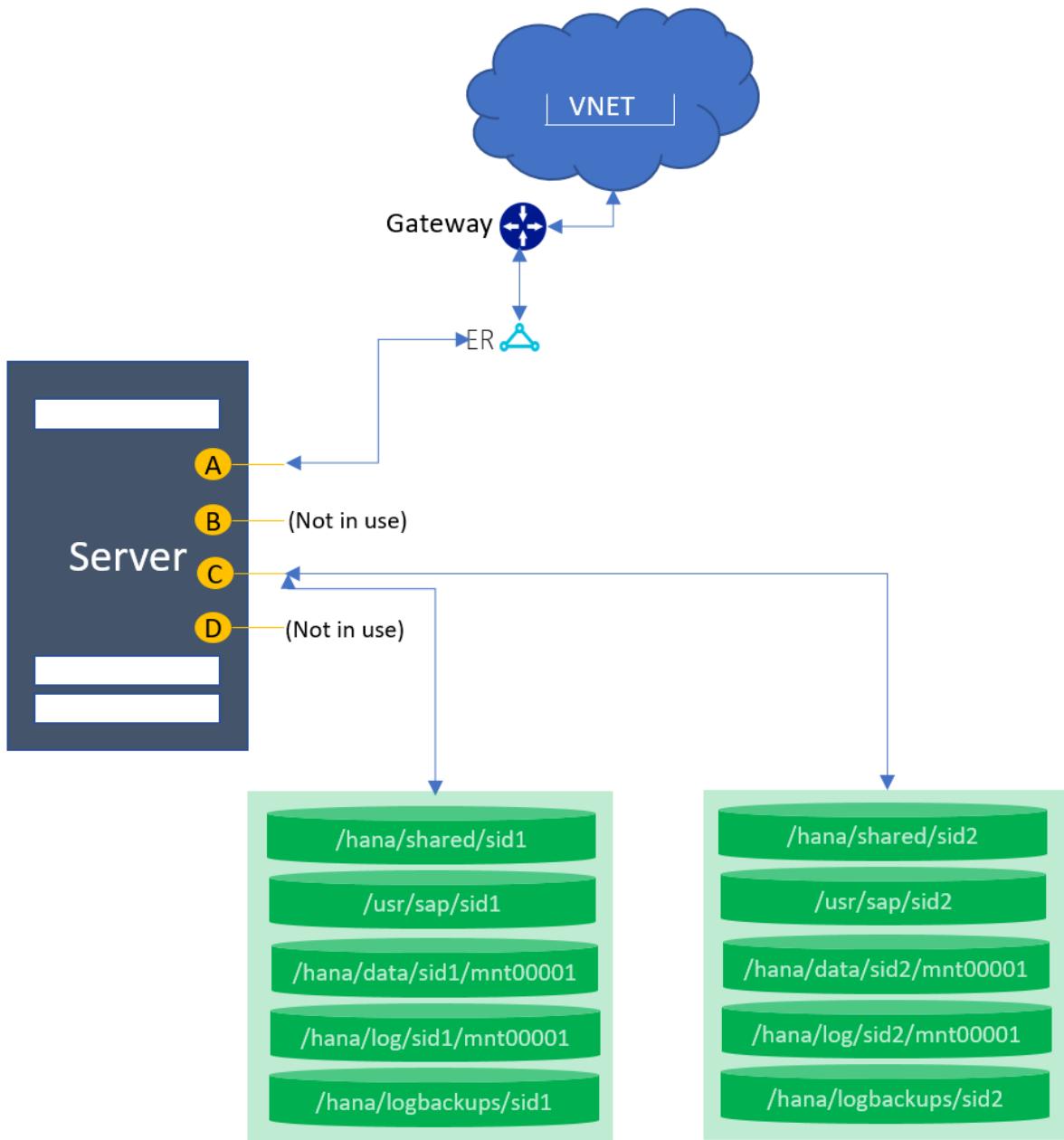
## Key considerations

- /usr/sap/SID is a symbolic link to /hana/shared/SID.

## Single node MCOS

This topology supports one node in a scale-up configuration with multiple SIDs.

## Architecture diagram



## Ethernet

The following network interfaces are preconfigured:

[] Expand table

NIC logical interface	SKU type	Name with SUSE OS	Name with RHEL OS	Use case
A	TYPE I	eth0.tenant	eno1.tenant	Client-to-HLI

<b>NIC logical interface</b>	<b>SKU type</b>	<b>Name with SUSE OS</b>	<b>Name with RHEL OS</b>	<b>Use case</b>
B	TYPE I	eth2.tenant	eno3.tenant	Configured but not in use
C	TYPE I	eth1.tenant	eno2.tenant	Node-to-storage
D	TYPE I	eth4.tenant	eno4.tenant	Configured but not in use
A	TYPE II	vlan<tenantNo>	team0.tenant	Client-to-HLI
B	TYPE II	vlan<tenantNo+2>	team0.tenant+2	Configured but not in use
C	TYPE II	vlan<tenantNo+1>	team0.tenant+1	Node-to-storage
D	TYPE II	vlan<tenantNo+3>	team0.tenant+3	Configured but not in use

## Storage

The following mount points are preconfigured:

[\[\] Expand table](#)

<b>Mount point</b>	<b>Use case</b>
/hana/shared/SID1	HANA installation for SID1
/hana/data/SID1/mnt00001	Data files installation for SID1
/hana/log/SID1/mnt00001	Log files installation for SID1
/hana/logbackups/SID1	Redo logs for SID1
/hana/shared/SID2	HANA installation for SID2
/hana/data/SID2/mnt00001	Data files installation for SID2
/hana/log/SID2/mnt00001	Log files installation for SID2
/hana/logbackups/SID2	Redo logs for SID2

## Key considerations

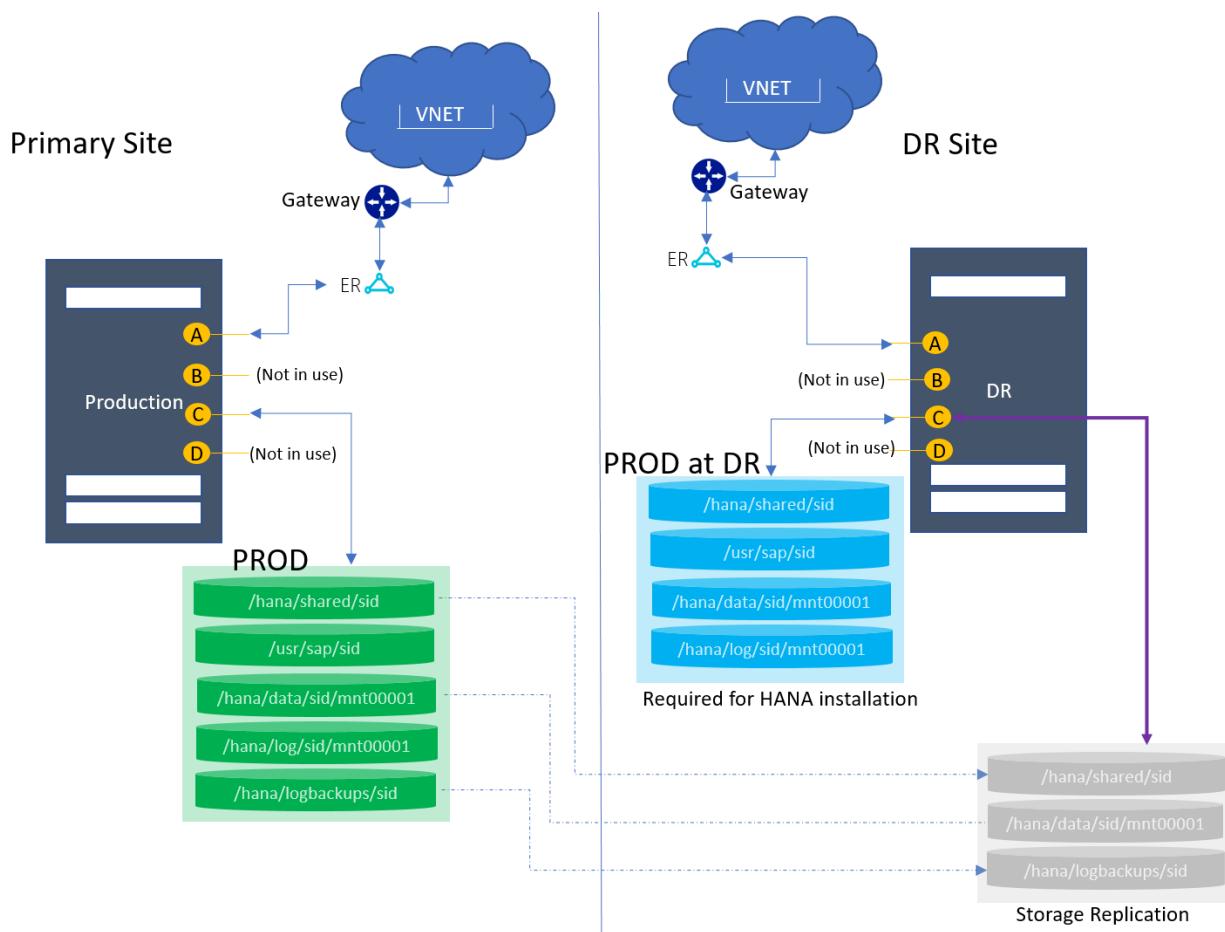
- /usr/sap/SID is a symbolic link to /hana/shared/SID.

- Volume size distribution is based on the database size in memory. To learn what database sizes in memory are supported in a multi-SID environment, see [Overview and architecture](#).

## Single node with DR using storage replication

This topology supports one node in a scale-up configuration with one or multiple SIDs. Storage-based replication to the DR site is used for a primary SID. In the diagram, only a single-SID system is shown at the primary site, but MCOS systems are supported as well.

### Architecture diagram



## Ethernet

The following network interfaces are preconfigured:

[ ] [Expand table](#)

NIC logical interface	SKU type	Name with SUSE OS	Name with RHEL OS	Use case
A	TYPE I	eth0.tenant	eno1.tenant	Client-to-HLI

<b>NIC logical interface</b>	<b>SKPE I type</b>	<b>Name with SUSE OS</b>	<b>Name with RHEL OS</b>	<b>Use case</b>
C	TYPE I	eth1.tenant	eno2.tenant	Node-to-storage
D	TYPE I	eth4.tenant	eno4.tenant	Configured but not in use
A	TYPE II	vlan<tenantNo>	team0.tenant	Client-to-HLI
B	TYPE II	vlan<tenantNo+2>	team0.tenant+2	Configured but not in use
C	TYPE II	vlan<tenantNo+1>	team0.tenant+1	Node-to-storage
D	TYPE II	vlan<tenantNo+3>	team0.tenant+3	Configured but not in use

## Storage

The following mount points are preconfigured:

[\[+\] Expand table](#)

<b>Mount point</b>	<b>Use case</b>
/hana/shared/SID	HANA installation for SID
/hana/data/SID/mnt00001	Data files installation for SID
/hana/log/SID/mnt00001	Log files installation for SID
/hana/logbackups/SID	Redo logs for SID

## Key considerations

- /usr/sap/SID is a symbolic link to /hana/shared/SID.
- For MCOS: Volume size distribution is based on the database size in memory. To learn what database sizes in memory are supported in a multi-SID environment, see [Overview and architecture](#).
- At the DR site: The volumes and mount points are configured (marked as "Required for HANA installation") for the production HANA instance installation at the DR HLI unit.
- At the DR site: The data, log backups, and shared volumes (marked as "Storage Replication") are replicated via snapshot from the production site. These volumes

are mounted during failover only. For more information, see [Disaster recovery failover procedure](#).

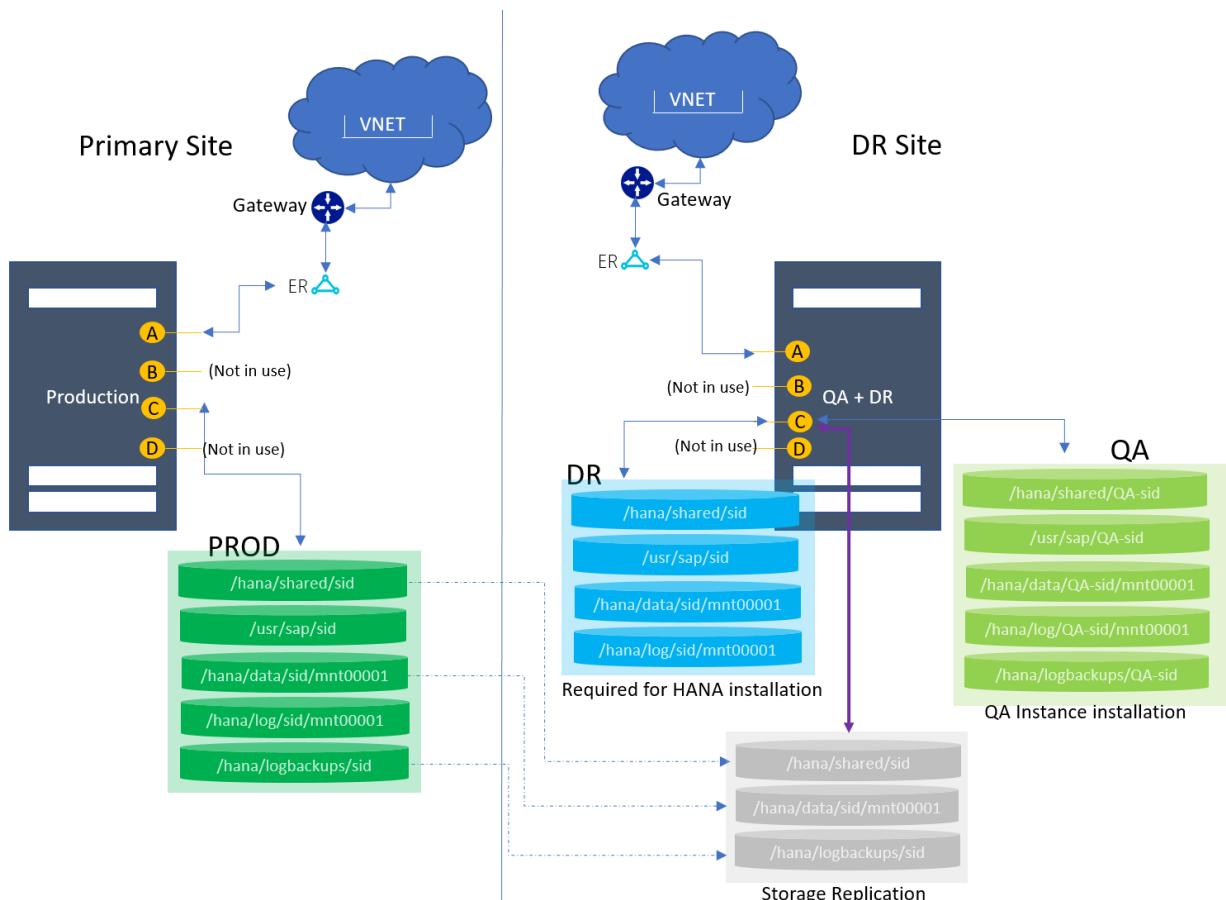
- The boot volume for *SKU Type I class* is replicated to the DR node.

## Single node with DR (multipurpose) using storage replication

This topology supports one node in a scale-up configuration with one or multiple SIDs. Storage-based replication to the DR site is used for a primary SID.

In the diagram, only a single-SID system is shown at the primary site, but multi-SID (MCOS) systems are supported as well. At the DR site, the HLI unit is used for the QA instance. Production operations run from the primary site. During DR failover (or failover test), the QA instance at the DR site is taken down.

## Architecture diagram



## Ethernet

The following network interfaces are preconfigured:

[\[\] Expand table](#)

<b>NIC logical interface</b>	<b>SKU type</b>	<b>Name with SUSE OS</b>	<b>Name with RHEL OS</b>	<b>Use case</b>
A	TYPE I	eth0.tenant	eno1.tenant	Client-to-HLI
B	TYPE I	eth2.tenant	eno3.tenant	Configured but not in use
C	TYPE I	eth1.tenant	eno2.tenant	Node-to-storage
D	TYPE I	eth4.tenant	eno4.tenant	Configured but not in use
A	TYPE II	vlan<tenantNo>	team0.tenant	Client-to-HLI
B	TYPE II	vlan<tenantNo+2>	team0.tenant+2	Configured but not in use
C	TYPE II	vlan<tenantNo+1>	team0.tenant+1	Node-to-storage
D	TYPE II	vlan<tenantNo+3>	team0.tenant+3	Configured but not in use

## Storage

The following mount points are preconfigured:

[\[\] Expand table](#)

<b>Mount point</b>	<b>Use case</b>
<b>At the primary site</b>	
/hana/shared/SID	HANA installation for production SID
/hana/data/SID/mnt00001	Data files installation for production SID
/hana/log/SID/mnt00001	Log files installation for production SID
/hana/logbackups/SID	Redo logs for production SID
<b>At the DR site</b>	
/hana/shared/SID	HANA installation for production SID
/hana/data/SID/mnt00001	Data files installation for production SID
/hana/log/SID/mnt00001	Log files installation for production SID

Mount point	Use case
/hana/shared/QA-SID	HANA installation for QA SID
/hana/data/QA-SID/mnt00001	Data files installation for QA SID
/hana/log/QA-SID/mnt00001	Log files installation for QA SID
/hana/logbackups/QA-SID	Redo logs for QA SID

## Key considerations

- /usr/sap/SID is a symbolic link to /hana/shared/SID.
- For MCOS: Volume size distribution is based on the database size in memory. To learn what database sizes in memory are supported in a multi-SID environment, see [Overview and architecture](#).
- At the DR site: The volumes and mount points are configured (marked as "Required for HANA installation") for the production HANA instance installation at the DR HLI unit.
- At the DR site: The data, log backups, and shared volumes (marked as "Storage Replication") are replicated via snapshot from the production site. These volumes are mounted during failover only. For more information, see [Disaster recovery failover procedure](#).
- At the DR site: The data, log backups, log, and shared volumes for QA (marked as "QA instance installation") are configured for the QA instance installation.
- The boot volume for *SKU Type I class* is replicated to the DR node.

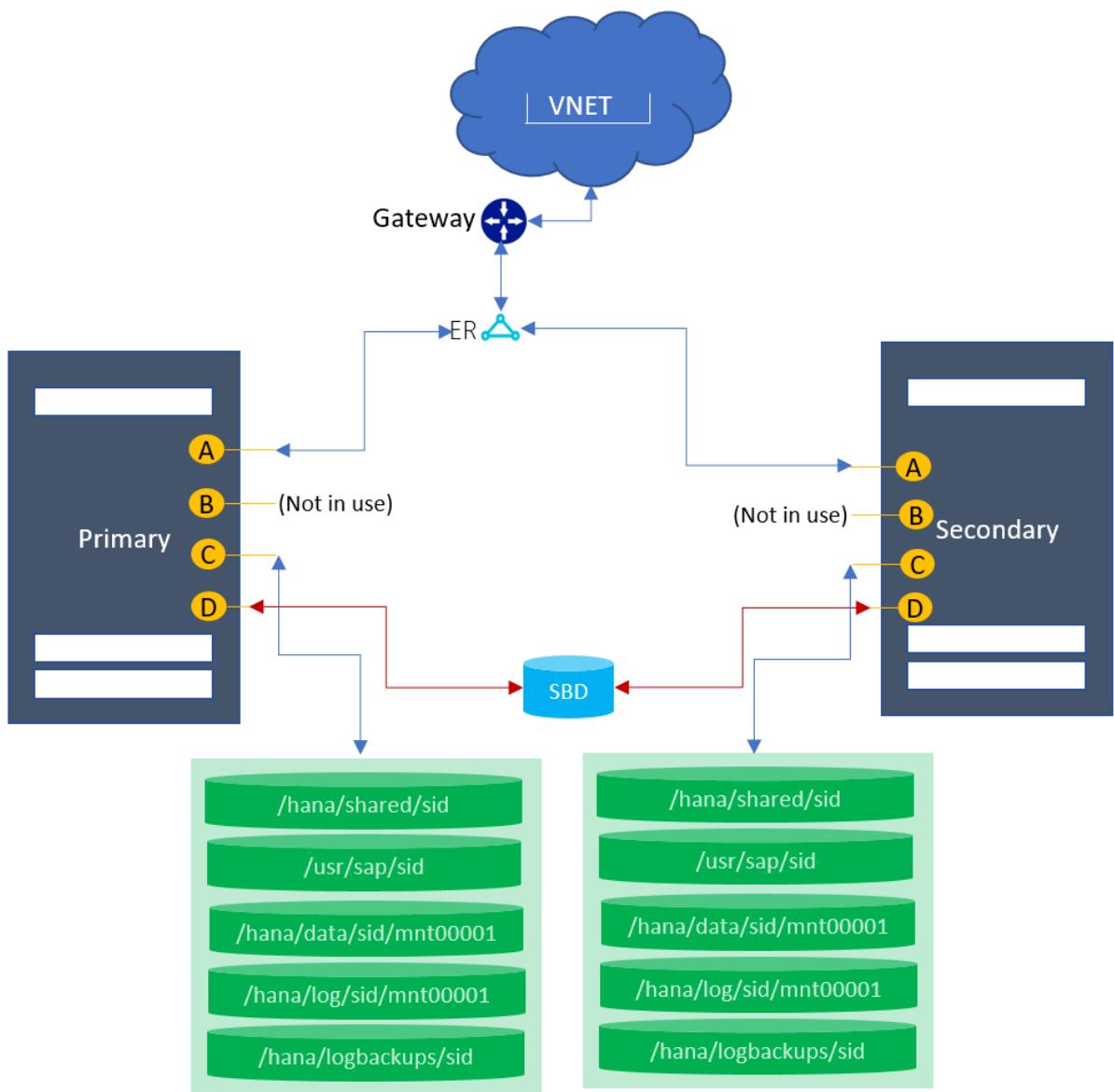
## HSR with fencing for high availability

This topology supports two nodes for the HANA system replication configuration. This configuration is supported only for single HANA instances on a node. MCOS scenarios aren't supported.

 **Note**

As of December 2019, this architecture is supported only for the SUSE operating system.

## Architecture diagram



## Ethernet

The following network interfaces are preconfigured:

[\[+\] Expand table](#)

NIC logical interface	SKU type	Name with SUSE OS	Name with RHEL OS	Use case
A	TYPE I	eth0.tenant	eno1.tenant	Client-to-HLI
B	TYPE I	eth2.tenant	eno3.tenant	Configured but not in use
C	TYPE I	eth1.tenant	eno2.tenant	Node-to-storage
D	TYPE I	eth4.tenant	eno4.tenant	Used for fencing

<b>NIC logical interface</b>	<b>SKU type</b>	<b>Name with SUSE OS</b>	<b>Name with RHEL OS</b>	<b>Use case</b>
A	TYPE II	vlan<tenantNo>	team0.tenant	Client-to-HLI
B	TYPE II	vlan<tenantNo+2>	team0.tenant+2	Configured but not in use
C	TYPE II	vlan<tenantNo+1>	team0.tenant+1	Node-to-storage
D	TYPE II	vlan<tenantNo+3>	team0.tenant+3	Used for fencing

## Storage

The following mount points are preconfigured:

[ ] Expand table

<b>Mount point</b>	<b>Use case</b>
<b>On the primary node</b>	
/hana/shared/SID	HANA installation for production SID
/hana/data/SID/mnt00001	Data files installation for production SID
/hana/log/SID/mnt00001	Log files installation for production SID
/hana/logbackups/SID	Redo logs for production SID
<b>On the secondary node</b>	
/hana/shared/SID	HANA installation for secondary SID
/hana/data/SID/mnt00001	Data files installation for secondary SID
/hana/log/SID/mnt00001	Log files installation for secondary SID
/hana/logbackups/SID	Redo logs for secondary SID

## Key considerations

- /usr/sap/SID is a symbolic link to /hana/shared/SID.
- For MCOS: Volume size distribution is based on the database size in memory. To learn what database sizes in memory are supported in a multi-SID environment, see [Overview and architecture](#).

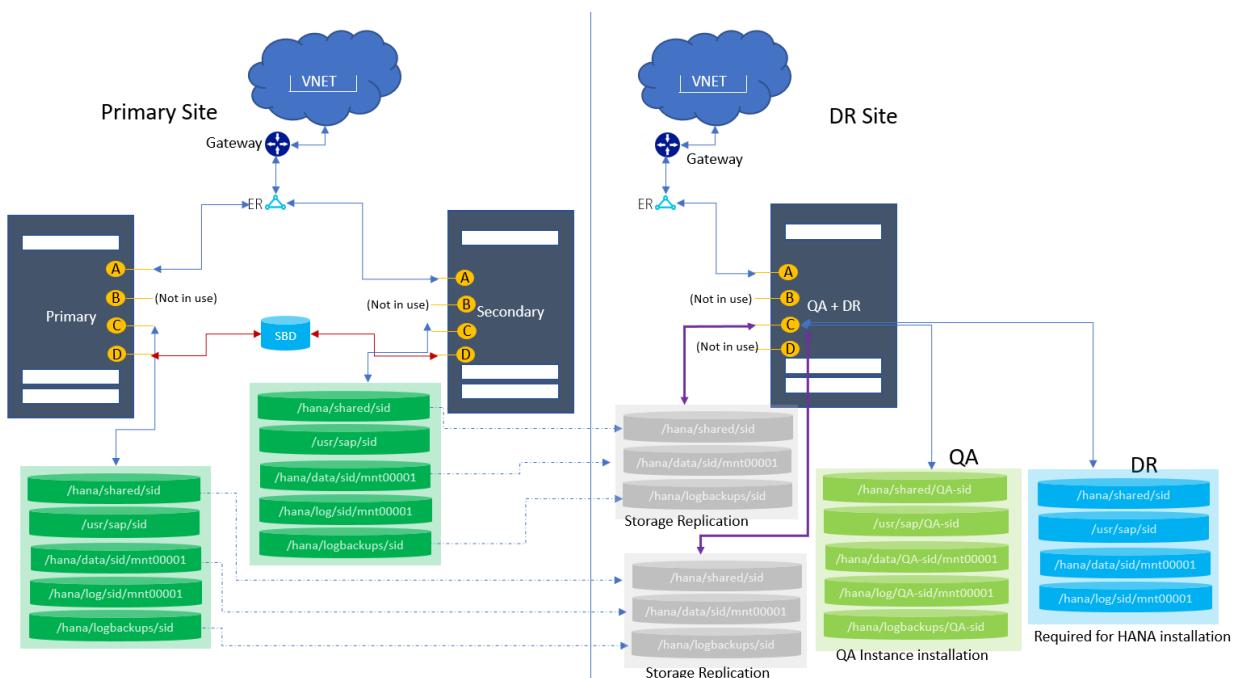
- Fencing: An SBD is configured for the fencing device setup. However, the use of fencing is optional.

# High availability with HSR and DR with storage replication

This topology supports two nodes for the HANA system replication configuration. Both normal and multipurpose DRs are supported. These configurations are supported only for single HANA instances on a node. MCOS scenarios aren't supported with these configurations.

In the diagram, a multipurpose scenario is shown at the DR site, where the HLI unit is used for the QA instance. Production operations run from the primary site. During DR failover (or failover test), the QA instance at the DR site is taken down.

## Architecture diagram



## Ethernet

The following network interfaces are preconfigured:

[Expand table](#)

NIC logical interface	SKU type	Name with SUSE OS	Name with RHEL OS	Use case
A	TYPE I	eth0.tenant	eno1.tenant	Client-to-HLI

<b>NIC logical interface</b>	<b>SKU type</b>	<b>Name with SUSE OS</b>	<b>Name with RHEL OS</b>	<b>Use case</b>
B	TYPE I	eth2.tenant	eno3.tenant	Configured but not in use
C	TYPE I	eth1.tenant	eno2.tenant	Node-to-storage
D	TYPE I	eth4.tenant	eno4.tenant	Used for fencing
A	TYPE II	vlan<tenantNo>	team0.tenant	Client-to-HLI
B	TYPE II	vlan<tenantNo+2>	team0.tenant+2	Configured but not in use
C	TYPE II	vlan<tenantNo+1>	team0.tenant+1	Node-to-storage
D	TYPE II	vlan<tenantNo+3>	team0.tenant+3	Used for fencing

## Storage

The following mount points are preconfigured:

[\[+\] Expand table](#)

<b>Mount point</b>	<b>Use case</b>
<b>On the primary node at the primary site</b>	
/hana/shared/SID	HANA installation for production SID
/hana/data/SID/mnt00001	Data files installation for production SID
/hana/log/SID/mnt00001	Log files installation for production SID
/hana/logbackups/SID	Redo logs for production SID
<b>On the secondary node at the primary site</b>	
/hana/shared/SID	HANA installation for secondary SID
/hana/data/SID/mnt00001	Data files installation for secondary SID
/hana/log/SID/mnt00001	Log files installation for secondary SID
/hana/logbackups/SID	Redo logs for secondary SID
<b>At the DR site</b>	
/hana/shared/SID	HANA installation for production SID
/hana/data/SID/mnt00001	Data files installation for production SID

Mount point	Use case
/hana/log/SID/mnt00001	Log files installation for production SID
/hana/shared/QA-SID	HANA installation for QA SID
/hana/data/QA-SID/mnt00001	Data files installation for QA SID
/hana/log/QA-SID/mnt00001	Log files installation for QA SID
/hana/logbackups/QA-SID	Redo logs for QA SID

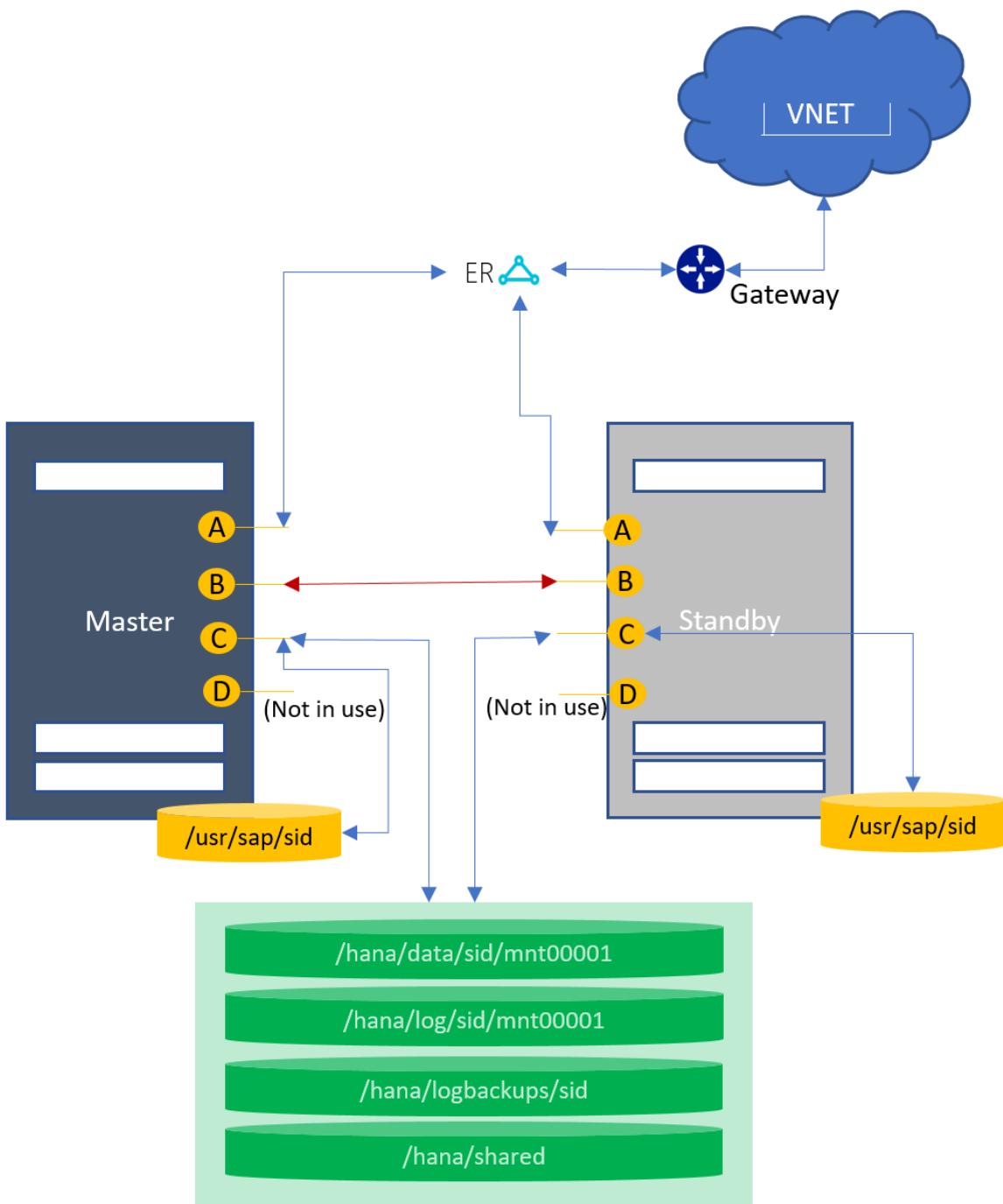
## Key considerations

- /usr/sap/SID is a symbolic link to /hana/shared/SID.
- For MCOS: Volume size distribution is based on the database size in memory. To learn what database sizes in memory are supported in a multi-SID environment, see [Overview and architecture](#).
- Fencing: An SBD is configured for the fencing setup. However, the use of fencing is optional.
- At the DR site: *Two sets of storage volumes are required* for primary and secondary node replication.
- At the DR site: The volumes and mount points are configured (marked as "Required for HANA installation") for the production HANA instance installation at the DR HLI unit.
- At the DR site: The data, log backups, and shared volumes (marked as "Storage Replication") are replicated via snapshot from the production site. These volumes are mounted during failover only. For more information, see [Disaster recovery failover procedure](#).
- At the DR site: The data, log backups, log, and shared volumes for QA (marked as "QA instance installation") are configured for the QA instance installation.
- The boot volume for *SKU Type I class* is replicated to the DR node.

## Host auto failover (1+1)

This topology supports two nodes in a host auto failover configuration. There's one node with a primary/worker role and another as a standby. *SAP supports this scenario only for S/4 HANA*. For more information, see [OSS note 2408419 - SAP S/4HANA - Multi-Node Support](#).

## Architecture diagram



## Ethernet

The following network interfaces are preconfigured:

[\[+\] Expand table](#)

NIC logical interface	SKU type	Name with SUSE OS	Name with RHEL OS	Use case
A	TYPE I	eth0.tenant	eno1.tenant	Client-to-HLI

<b>NIC logical interface</b>	<b>SKU type</b>	<b>Name with SUSE OS</b>	<b>Name with RHEL OS</b>	<b>Use case</b>
B	TYPE I	eth2.tenant	eno3.tenant	Node-to-node communication
C	TYPE I	eth1.tenant	eno2.tenant	Node-to-storage
D	TYPE I	eth4.tenant	eno4.tenant	Configured but not in use
A	TYPE II	vlan<tenantNo>	team0.tenant	Client-to-HLI
B	TYPE II	vlan<tenantNo+2>	team0.tenant+2	Node-to-node communication
C	TYPE II	vlan<tenantNo+1>	team0.tenant+1	Node-to-storage
D	TYPE II	vlan<tenantNo+3>	team0.tenant+3	Configured but not in use

## Storage

The following mount points are preconfigured:

[\[+\] Expand table](#)

<b>Mount point</b>	<b>Use case</b>
On the primary and standby nodes	
/hana/shared	HANA installation for production SID
/hana/data/SID/mnt00001	Data files installation for production SID
/hana/log/SID/mnt00001	Log files installation for production SID
/hana/logbackups/SID	Redo logs for production SID

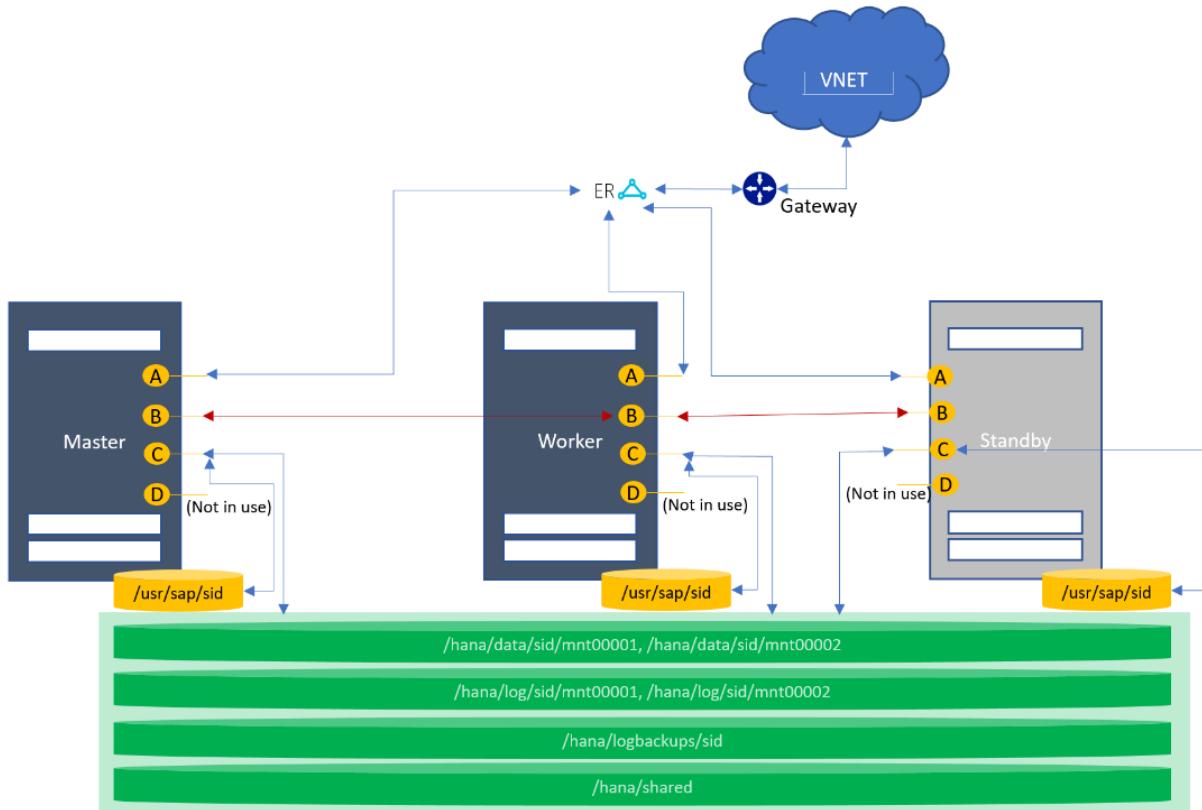
## Key considerations

- /usr/sap/SID is a symbolic link to /hana/shared/SID.
- On standby: The volumes and mount points are configured (marked as "Required for HANA installation") for the HANA instance installation at the standby unit.

## Scale-out with standby

This topology supports multiple nodes in a scale-out configuration. There's one node with a primary role, one or more nodes with a worker role, and one or more nodes as standby. However, there can be only one primary node at any given time.

## Architecture diagram



## Ethernet

The following network interfaces are preconfigured:

[Expand table](#)

NIC logical interface	SKU type	Name with SUSE OS	Name with RHEL OS	Use case
A	TYPE I	eth0.tenant	eno1.tenant	Client-to-HLI
B	TYPE I	eth2.tenant	eno3.tenant	Node-to-node communication
C	TYPE I	eth1.tenant	eno2.tenant	Node-to-storage
D	TYPE I	eth4.tenant	eno4.tenant	Configured but not in use

<b>NIC logical interface</b>	<b>SKU type</b>	<b>Name with SUSE OS</b>	<b>Name with RHEL OS</b>	<b>Use case</b>
A	TYPE II	vlan<tenantNo>	team0.tenant	Client-to-HLI
B	TYPE II	vlan<tenantNo+2>	team0.tenant+2	Node-to-node communication
C	TYPE II	vlan<tenantNo+1>	team0.tenant+1	Node-to-storage
D	TYPE II	vlan<tenantNo+3>	team0.tenant+3	Configured but not in use

## Storage

The following mount points are preconfigured:

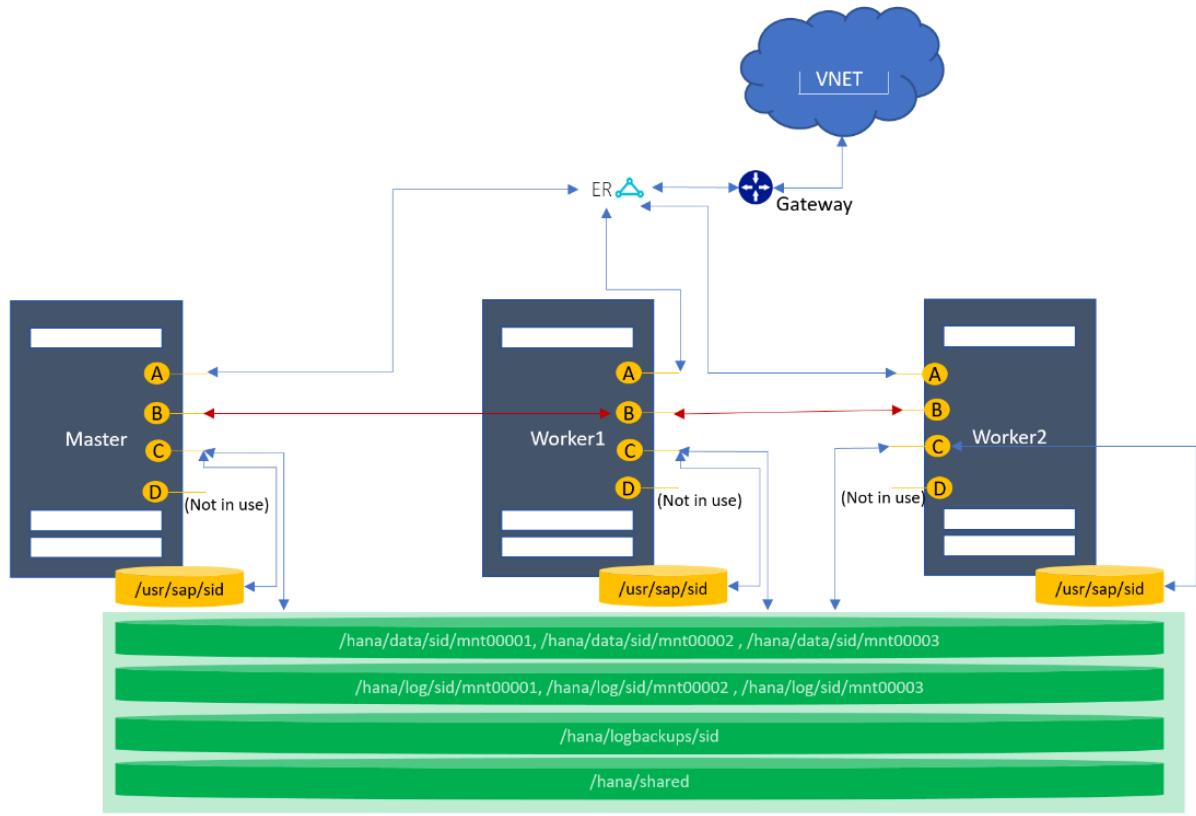
[\[+\] Expand table](#)

<b>Mount point</b>	<b>Use case</b>
<b>On the primary, worker, and standby nodes</b>	
/hana/shared	HANA installation for production SID
/hana/data/SID/mnt00001	Data files installation for production SID
/hana/log/SID/mnt00001	Log files installation for production SID
/hana/logbackups/SID	Redo logs for production SID

## Scale-out without standby

This topology supports multiple nodes in a scale-out configuration. There's one node with a primary role, and one or more nodes with a worker role. However, there can be only one primary node at any given time.

## Architecture diagram



## Ethernet

The following network interfaces are preconfigured:

[\[ \] Expand table](#)

NIC logical interface	SKU type	Name with SUSE OS	Name with RHEL OS	Use case
A	TYPE I	eth0.tenant	eno1.tenant	Client-to-HLI
B	TYPE I	eth2.tenant	eno3.tenant	Node-to-node communication
C	TYPE I	eth1.tenant	eno2.tenant	Node-to-storage
D	TYPE I	eth4.tenant	eno4.tenant	Configured but not in use
A	TYPE II	vlan<tenantNo>	team0.tenant	Client-to-HLI
B	TYPE II	vlan<tenantNo+2>	team0.tenant+2	Node-to-node communication
C	TYPE II	vlan<tenantNo+1>	team0.tenant+1	Node-to-storage

<b>NIC logical interface</b>	<b>SKU type</b>	<b>Name with SUSE OS</b>	<b>Name with RHEL OS</b>	<b>Use case</b>
D	TYPE II	vlan<tenantNo+3>	team0.tenant+3	Configured but not in use

## Storage

The following mount points are preconfigured:

[Expand table](#)

<b>Mount point</b>	<b>Use case</b>
<b>On the primary and worker nodes</b>	
/hana/shared	HANA installation for production SID
/hana/data/SID/mnt00001	Data files installation for production SID
/hana/log/SID/mnt00001	Log files installation for production SID
/hana/logbackups/SID	Redo logs for production SID

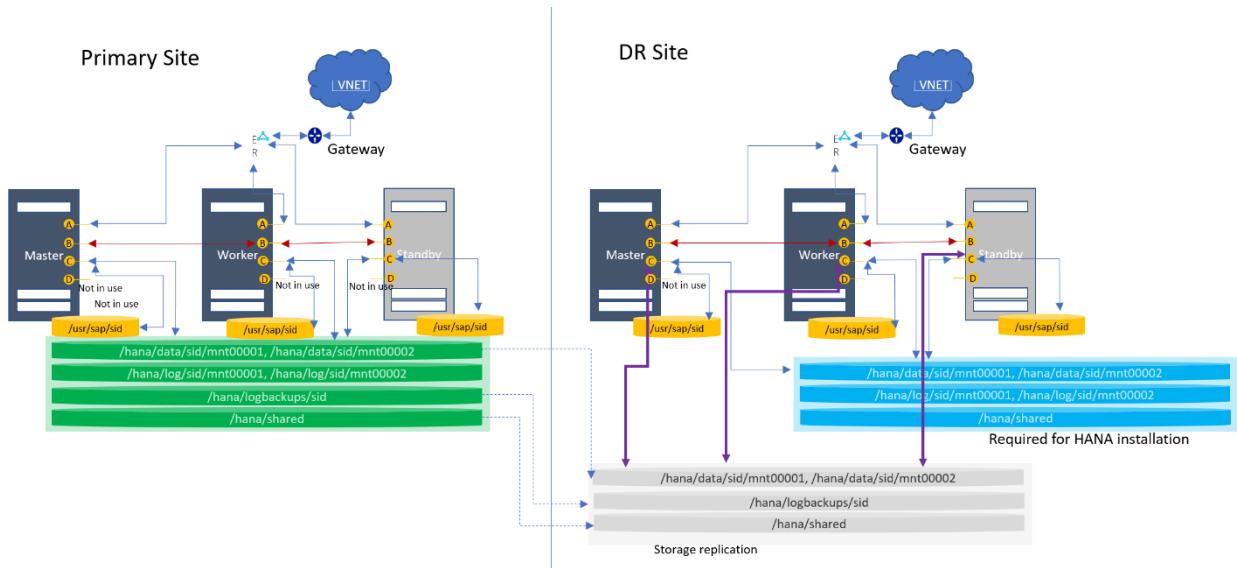
## Key considerations

- /usr/sap/SID is a symbolic link to /hana/shared/SID.

## Scale-out with DR using storage replication

This topology supports multiple nodes in a scale-out with a DR. Both normal and multipurpose DRs are supported. In the diagram, only the single purpose DR is shown. You can request this topology with or without the standby node.

## Architecture diagram



## Ethernet

The following network interfaces are preconfigured:

[Expand table](#)

NIC logical interface	SKU type	Name with SUSE OS	Name with RHEL OS	Use case
A	TYPE I	eth0.tenant	eno1.tenant	Client-to-HLI
B	TYPE I	eth2.tenant	eno3.tenant	Node-to-node communication
C	TYPE I	eth1.tenant	eno2.tenant	Node-to-storage
D	TYPE I	eth4.tenant	eno4.tenant	Configured but not in use
A	TYPE II	vlan<tenantNo>	team0.tenant	Client-to-HLI
B	TYPE II	vlan<tenantNo+2>	team0.tenant+2	Node-to-node communication
C	TYPE II	vlan<tenantNo+1>	team0.tenant+1	Node-to-storage
D	TYPE II	vlan<tenantNo+3>	team0.tenant+3	Configured but not in use

## Storage

The following mount points are preconfigured:

Mount point	Use case
<b>On the primary node</b>	
/hana/shared	HANA installation for production SID
/hana/data/SID/mnt00001	Data files installation for production SID
/hana/log/SID/mnt00001	Log files installation for production SID
/hana/logbackups/SID	Redo logs for production SID
<b>On the DR node</b>	
/hana/shared	HANA installation for production SID
/hana/data/SID/mnt00001	Data files installation for production SID
/hana/log/SID/mnt00001	Log files installation for production SID

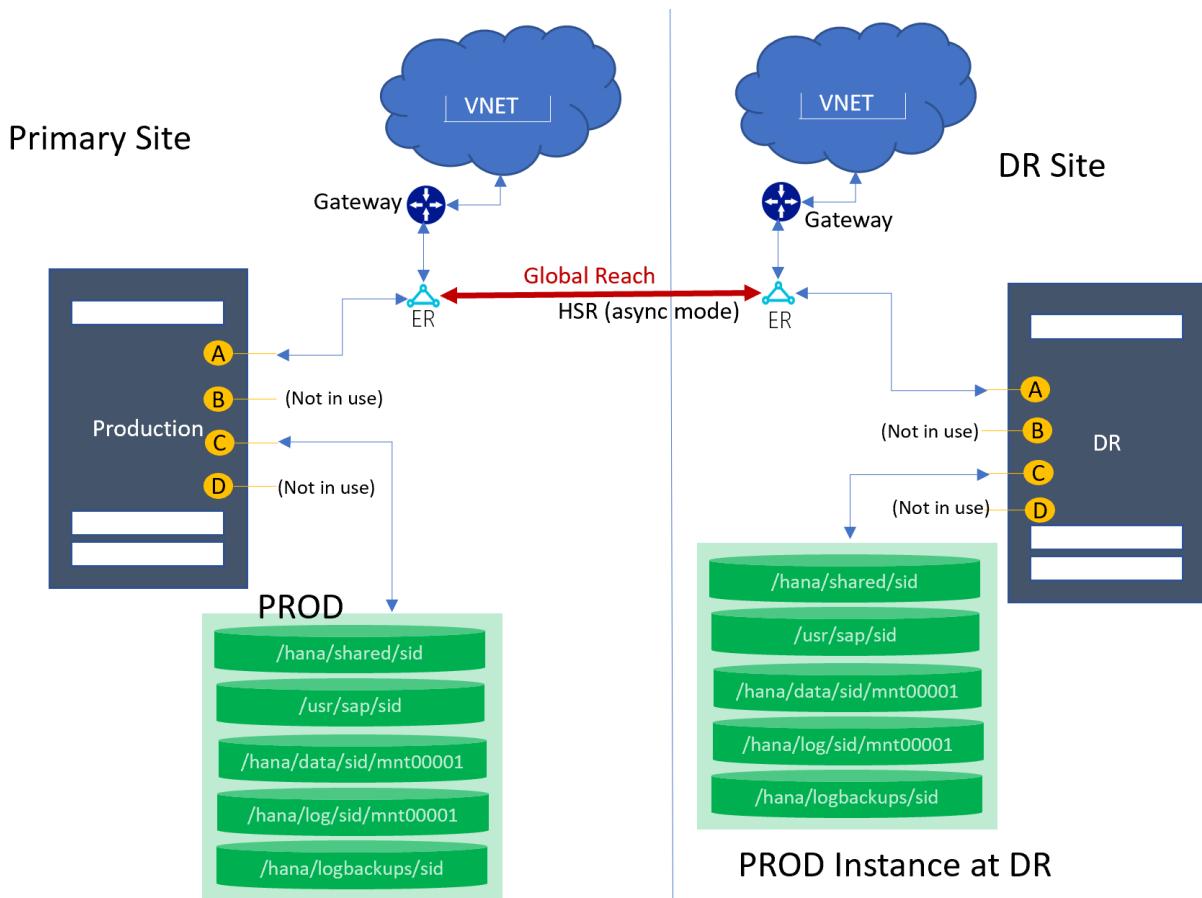
## Key considerations

- /usr/sap/SID is a symbolic link to /hana/shared/SID.
- At the DR site: The volumes and mount points are configured (marked as "Required for HANA installation") for the production HANA instance installation at the DR HLI unit.
- At the DR site: The data, log backups, and shared volumes (marked as "Storage Replication") are replicated via snapshot from the production site. These volumes are mounted during failover only. For more information, see [Disaster recovery failover procedure](#).
- The boot volume for *SKU Type I class* is replicated to the DR node.

## Single node with DR using HSR

This topology supports one node in a scale-up configuration with one SID, with HANA system replication to the DR site for a primary SID. In the diagram, only a single-SID system is shown at the primary site, but multi-SID (MCOS) systems are supported as well.

## Architecture diagram



## Ethernet

The following network interfaces are preconfigured:

[\[ \] Expand table](#)

NIC logical interface	SKU type	Name with SUSE OS	Name with RHEL OS	Use case
A	TYPE I	eth0.tenant	eno1.tenant	Client-to-HLI/HSR
B	TYPE I	eth2.tenant	eno3.tenant	Configured but not in use
C	TYPE I	eth1.tenant	eno2.tenant	Node-to-storage
D	TYPE I	eth4.tenant	eno4.tenant	Configured but not in use
A	TYPE II	vlan<tenantNo>	team0.tenant	Client-to-HLI/HSR
B	TYPE II	vlan<tenantNo+2>	team0.tenant+2	Configured but not in use
C	TYPE II	vlan<tenantNo+1>	team0.tenant+1	Node-to-storage

NIC logical interface	SKU type	Name with SUSE OS	Name with RHEL OS	Use case
D	TYPE II	vlan<tenantNo+3>	team0.tenant+3	Configured but not in use

## Storage

The following mount points are preconfigured on both HLI units (Primary and DR):

[Expand table](#)

Mount point	Use case
/hana/shared/SID	HANA installation for SID
/hana/data/SID/mnt00001	Data files installation for SID
/hana/log/SID/mnt00001	Log files installation for SID
/hana/logbackups/SID	Redo logs for SID

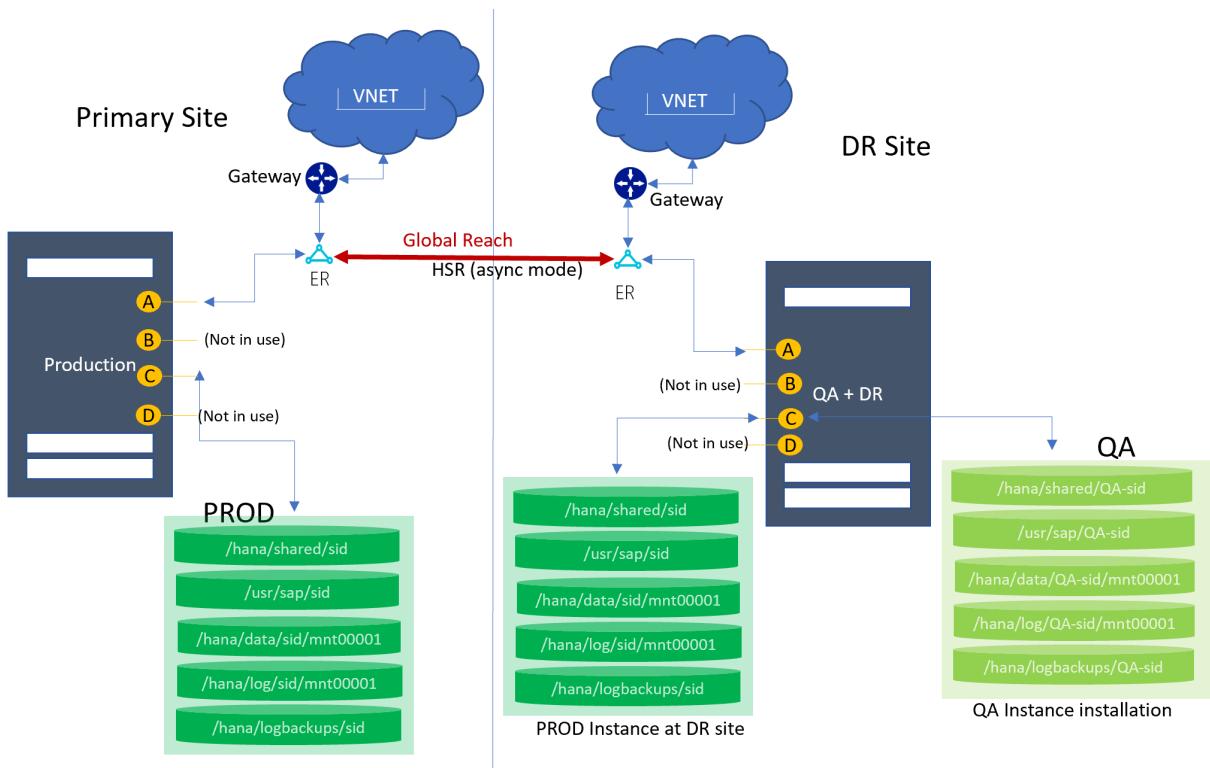
## Key considerations

- /usr/sap/SID is a symbolic link to /hana/shared/SID.
- For MCOS: Volume size distribution is based on the database size in memory. To learn what database sizes in memory are supported in a multi-SID environment, see [Overview and architecture](#).
- The primary node syncs with the DR node by using HANA system replication.
- [Global Reach](#) is used to link the ExpressRoute circuits together to make a private network between your regional networks.

## Single node HSR to DR (cost optimized)

This topology supports one node in a scale-up configuration with one SID. HANA system replication to the DR site is used for a primary SID. In the diagram, only a single-SID system is shown at the primary site, but multi-SID (MCOS) systems are supported as well. At the DR site, an HLI unit is used for the QA instance. Production operations run from the primary site. During DR failover (or failover test), the QA instance at the DR site is taken down.

## Architecture diagram



## Ethernet

The following network interfaces are preconfigured:

[Expand table](#)

NIC logical interface	SKU type	Name with SUSE OS	Name with RHEL OS	Use case
A	TYPE I	eth0.tenant	eno1.tenant	Client-to-HLI/HSR
B	TYPE I	eth2.tenant	eno3.tenant	Configured but not in use
C	TYPE I	eth1.tenant	eno2.tenant	Node-to-storage
D	TYPE I	eth4.tenant	eno4.tenant	Configured but not in use
A	TYPE II	vlan<tenantNo>	team0.tenant	Client-to-HLI/HSR
B	TYPE II	vlan<tenantNo+2>	team0.tenant+2	Configured but not in use
C	TYPE II	vlan<tenantNo+1>	team0.tenant+1	Node-to-storage
D	TYPE II	vlan<tenantNo+3>	team0.tenant+3	Configured but not in use

# Storage

The following mount points are preconfigured:

[+] Expand table

Mount point	Use case
<b>At the primary site</b>	
/hana/shared/SID	HANA installation for production SID
/hana/data/SID/mnt00001	Data files installation for production SID
/hana/log/SID/mnt00001	Log files installation for production SID
/hana/logbackups/SID	Redo logs for production SID
<b>At the DR site</b>	
/hana/shared/SID	HANA installation for production SID
/hana/data/SID/mnt00001	Data files installation for production SID
/hana/log/SID/mnt00001	Log files installation for production SID
/hana/logbackups/SID	Redo logs for production SID
/hana/shared/QA-SID	HANA installation for QA SID
/hana/data/QA-SID/mnt00001	Data files installation for QA SID
/hana/log/QA-SID/mnt00001	Log files installation for QA SID
/hana/logbackups/QA-SID	Redo logs for QA SID

## Key considerations

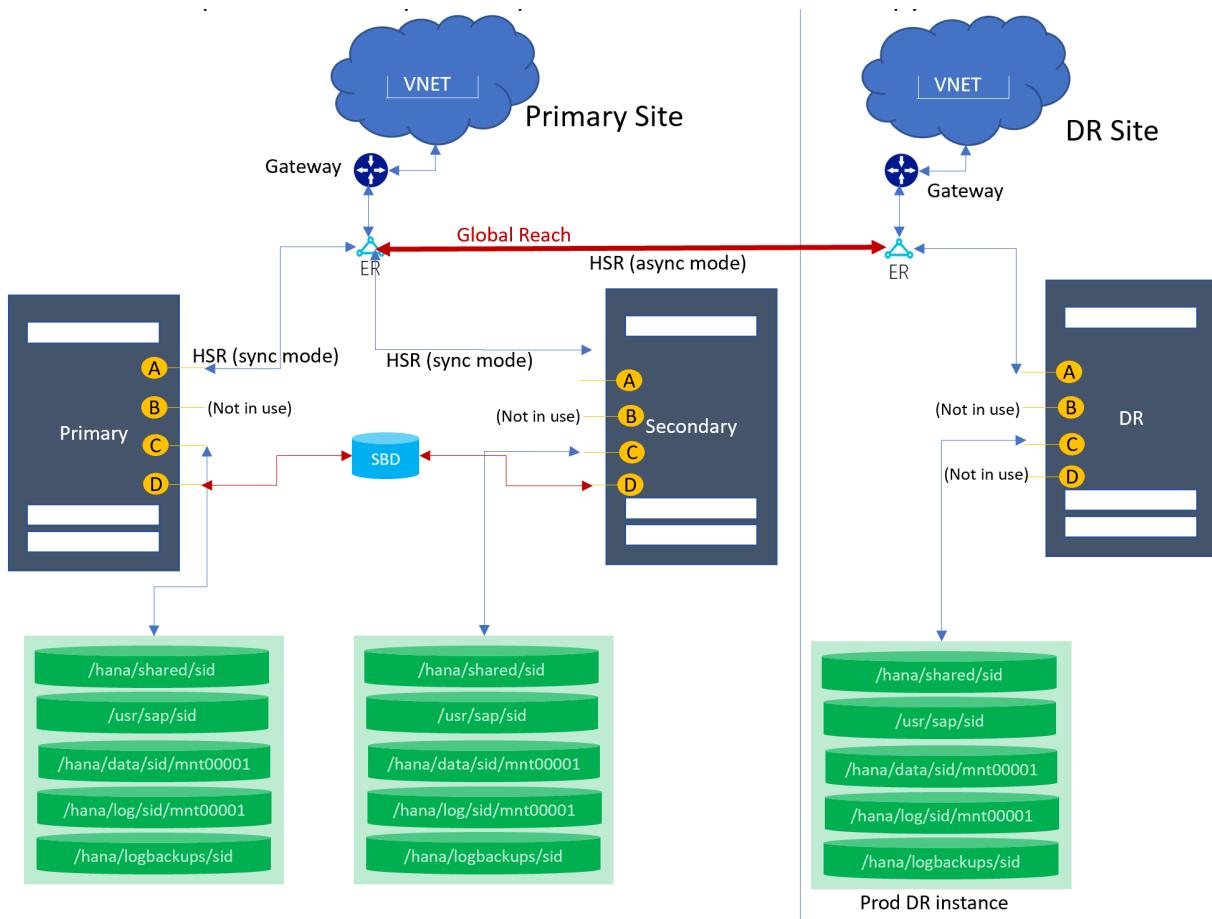
- /usr/sap/SID is a symbolic link to /hana/shared/SID.
- For MCOS: Volume size distribution is based on the database size in memory. To learn what database sizes in memory are supported in a multi-SID environment, see [Overview and architecture](#).
- At the DR site: The volumes and mount points are configured (marked as "PROD Instance at DR site") for the production HANA instance installation at the DR HLI unit.
- At the DR site: The data, log backups, log, and shared volumes for QA (marked as "QA instance installation") are configured for the QA instance installation.
- The primary node syncs with the DR node by using HANA system replication.

- **Global Reach** is used to link the ExpressRoute circuits together to make a private network between your regional networks.

## High availability and disaster recovery with HSR

This topology supports two nodes for the HANA system replication configuration for the local regions' high availability. For the DR, the third node at the DR region syncs with the primary site by using HSR (async mode).

### Architecture diagram



### Ethernet

The following network interfaces are preconfigured:

  Expand table

<b>NIC logical interface</b>	<b>SKU type</b>	<b>Name with SUSE OS</b>	<b>Name with RHEL OS</b>	<b>Use case</b>
A	TYPE I	eth0.tenant	eno1.tenant	Client-to-HLI/HSR
B	TYPE I	eth2.tenant	eno3.tenant	Configured but not in use
C	TYPE I	eth1.tenant	eno2.tenant	Node-to-storage
D	TYPE I	eth4.tenant	eno4.tenant	Configured but not in use
A	TYPE II	vlan<tenantNo>	team0.tenant	Client-to-HLI/HSR
B	TYPE II	vlan<tenantNo+2>	team0.tenant+2	Configured but not in use
C	TYPE II	vlan<tenantNo+1>	team0.tenant+1	Node-to-storage
D	TYPE II	vlan<tenantNo+3>	team0.tenant+3	Configured but not in use

## Storage

The following mount points are preconfigured:

[\[+\] Expand table](#)

<b>Mount point</b>	<b>Use case</b>
<b>At the primary site</b>	
/hana/shared/SID	HANA installation for production SID
/hana/data/SID/mnt00001	Data files installation for production SID
/hana/log/SID/mnt00001	Log files installation for production SID
/hana/logbackups/SID	Redo logs for production SID
<b>At the DR site</b>	
/hana/shared/SID	HANA installation for production SID
/hana/data/SID/mnt00001	Data files installation for production SID
/hana/log/SID/mnt00001	Log files installation for production SID
/hana/logbackups/SID	Redo logs for production SID

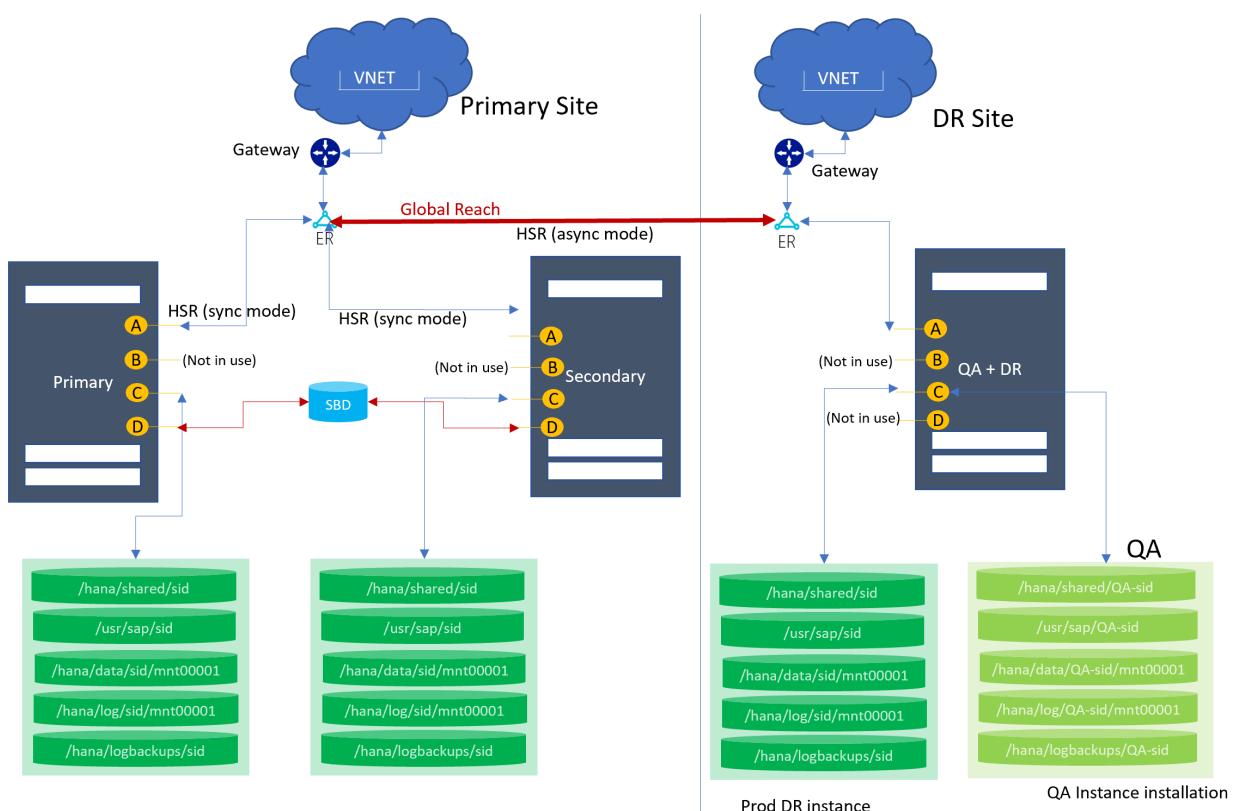
# Key considerations

- /usr/sap/SID is a symbolic link to /hana/shared/SID.
- At the DR site: The volumes and mount points are configured (marked as "PROD DR instance") for the production HANA instance installation at the DR HLI unit.
- The primary site node syncs with the DR node by using HANA system replication.
- **Global Reach** is used to link the ExpressRoute circuits together to make a private network between your regional networks.

## High availability and disaster recovery with HSR (cost optimized)

This topology supports two nodes for the HANA system replication configuration for the local regions' high availability. For the DR, the third node at the DR region syncs with the primary site by using HSR (async mode), while another instance (for example, QA) is already running out from the DR node.

## Architecture diagram



## Ethernet

The following network interfaces are preconfigured:

[\[\] Expand table](#)

<b>NIC logical interface</b>	<b>SKU type</b>	<b>Name with SUSE OS</b>	<b>Name with RHEL OS</b>	<b>Use case</b>
A	TYPE I	eth0.tenant	eno1.tenant	Client-to-HLI/HSR
B	TYPE I	eth2.tenant	eno3.tenant	Configured but not in use
C	TYPE I	eth1.tenant	eno2.tenant	Node-to-storage
D	TYPE I	eth4.tenant	eno4.tenant	Configured but not in use
A	TYPE II	vlan<tenantNo>	team0.tenant	Client-to-HLI/HSR
B	TYPE II	vlan<tenantNo+2>	team0.tenant+2	Configured but not in use
C	TYPE II	vlan<tenantNo+1>	team0.tenant+1	Node-to-storage
D	TYPE II	vlan<tenantNo+3>	team0.tenant+3	Configured but not in use

## Storage

The following mount points are preconfigured:

[\[\] Expand table](#)

<b>Mount point</b>	<b>Use case</b>
<b>At the primary site</b>	
/hana/shared/SID	HANA installation for production SID
/hana/data/SID/mnt00001	Data files installation for production SID
/hana/log/SID/mnt00001	Log files installation for production SID
/hana/logbackups/SID	Redo logs for production SID
<b>At the DR site</b>	
/hana/shared/SID	HANA installation for production SID
/hana/data/SID/mnt00001	Data files installation for production SID
/hana/log/SID/mnt00001	Log files installation for production SID

Mount point	Use case
/hana/logbackups/SID	Redo logs for production SID
/hana/shared/QA-SID	HANA installation for QA SID
/hana/data/QA-SID/mnt00001	Data files installation for QA SID
/hana/log/QA-SID/mnt00001	Log files installation for QA SID
/hana/logbackups/QA-SID	Redo logs for QA SID

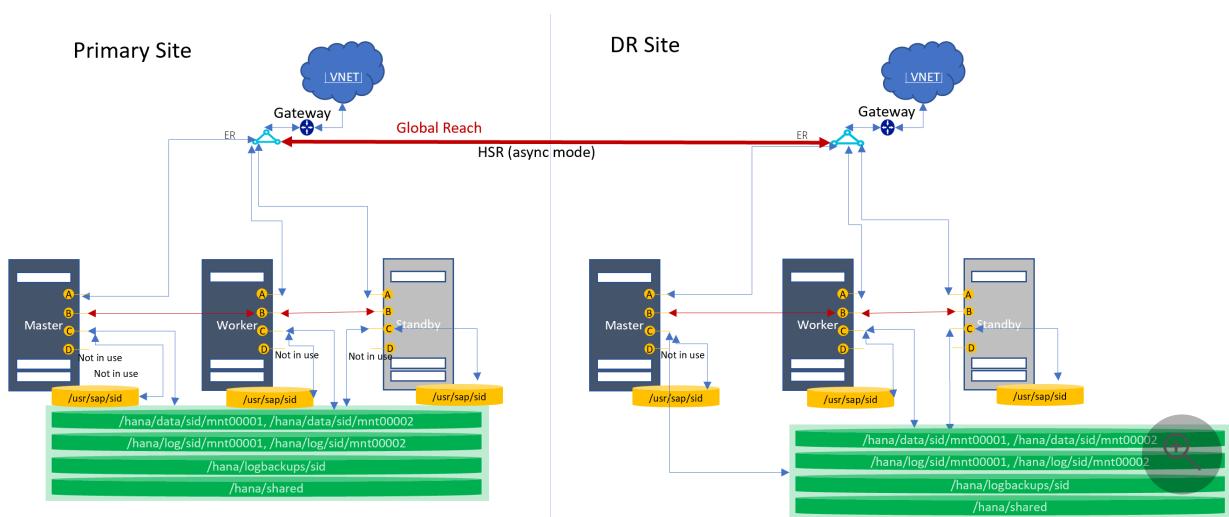
## Key considerations

- /usr/sap/SID is a symbolic link to /hana/shared/SID.
- At the DR site: The volumes and mount points are configured (marked as "PROD DR instance") for the production HANA instance installation at the DR HLI unit.
- At the DR site: The data, log backups, log, and shared volumes for QA (marked as "QA instance installation") are configured for the QA instance installation.
- The primary site node syncs with the DR node by using HANA system replication.
- **Global Reach** is used to link the ExpressRoute circuits together to make a private network between your regional networks.

## Scale-out with DR using HSR

This topology supports multiple nodes in a scale-out with a DR. You can request this topology with or without the standby node. The primary site node syncs with the DR site node by using HANA system replication (async mode).

## Architecture diagram



# Ethernet

The following network interfaces are preconfigured:

[\[+\] Expand table](#)

<b>NIC logical interface</b>	<b>SKU type</b>	<b>Name with SUSE OS</b>	<b>Name with RHEL OS</b>	<b>Use case</b>
A	TYPE I	eth0.tenant	eno1.tenant	Client-to-HLI/HSR
B	TYPE I	eth2.tenant	eno3.tenant	Node-to-node communication
C	TYPE I	eth1.tenant	eno2.tenant	Node-to-storage
D	TYPE I	eth4.tenant	eno4.tenant	Configured but not in use
A	TYPE II	vlan<tenantNo>	team0.tenant	Client-to-HLI/HSR
B	TYPE II	vlan<tenantNo+2>	team0.tenant+2	Node-to-node communication
C	TYPE II	vlan<tenantNo+1>	team0.tenant+1	Node-to-storage
D	TYPE II	vlan<tenantNo+3>	team0.tenant+3	Configured but not in use

# Storage

The following mount points are preconfigured:

[\[+\] Expand table](#)

<b>Mount point</b>	<b>Use case</b>
<b>On the primary node</b>	
/hana/shared	HANA installation for production SID
/hana/data/SID/mnt00001	Data files installation for production SID
/hana/log/SID/mnt00001	Log files installation for production SID
/hana/logbackups/SID	Redo logs for production SID
<b>On the DR node</b>	

Mount point	Use case
/hana/shared	HANA installation for production SID
/hana/data/SID/mnt00001	Data files installation for production SID
/hana/log/SID/mnt00001	Log files installation for production SID
/hana/logbackups/SID	Redo logs for production SID

## Key considerations

- /usr/sap/SID is a symbolic link to /hana/shared/SID.
- At the DR site: The volumes and mount points are configured for the production HANA instance installation at the DR HLI unit.
- The primary site node syncs with the DR node by using HANA system replication.
- **Global Reach** is used to link the ExpressRoute circuits together to make a private network between your regional networks.

## Next steps

Learn about deploying HANA Large Instances.

[SAP HANA \(Large Instances\) deployment](#)

# Understand Azure NetApp Files application volume group for SAP HANA

Article • 02/24/2023

This article helps you understand the use cases and key features of Azure NetApp Files application volume group for SAP HANA.

Application volume group for SAP HANA enables you to deploy all volumes required to install and operate an SAP HANA database according to best practices. Instead of individually creating the required SAP HANA volumes (including data, log, shared, log-backup, and data-backup volumes), application volume group for SAP HANA creates these volumes in a single "atomic" call. The atomic call ensures that either all volumes or no volumes at all are created.

Application volume group for SAP HANA provides technical improvements to simplify and standardize the process to help you streamline volume deployments for SAP HANA. As a result, you can focus on your application demands instead of managing technical settings such as individual QoS or sizes for volumes.

## Key features

Application volume group for SAP HANA is supported for all regions. It provides the following key features:

- Supporting SAP HANA configurations for both single and multiple host setups, including:
  - Volumes for a single or primary SAP HANA database
  - Volumes for an SAP HANA System Replication (HSR) secondary system
  - Volumes for a disaster recovery (DR) scenario using [cross-region replication](#)
- Creating the following volumes:
  - SAP HANA data volumes (one for each database host)
  - SAP HANA log volumes (one for each database host)
  - SAP HANA shared volumes (for the first SAP HANA host only)
  - Log-backup volumes (optional)
  - File-based data-backup volumes (optional)
- Creating volumes in a [manual QoS capacity pool](#). The volume size and the required performance (in MiB/s) are proposed based on user input for the memory size of

the database.

- The application volume group GUI and Azure Resource Manager (ARM) template provide best practices to simplify sizing management and volume creation. For example:
  - Proposing volume naming convention based on SAP System ID (SID) and volume type
  - Calculating the size and performance based on memory size

Application volume group for SAP HANA helps you simplify the deployment process and increase the storage performance for SAP HANA workloads. Some of the new features are as follows:

- Use of proximity placement group (PPG) instead of manual pinning.
  - You will anchor the SAP HANA VMs using a PPG to guarantee lowest possible latency. This PPG will be used to enforce that the data, log, and shared volumes are created in the close proximity to the SAP HANA VMs. See [Best practices about Proximity Placement Groups](#) for detail.
- Creation of separate storage endpoints (with different IP addresses) for data and log volumes.
  - This deployment method provides better performance and throughput for the SAP HANA database.

## Next steps

- Requirements and considerations for application volume group for SAP HANA
- Deploy the first SAP HANA host using application volume group for SAP HANA
- Add hosts to a multiple-host SAP HANA system using application volume group for SAP HANA
- Add volumes for an SAP HANA system as a secondary database in HSR
- Add volumes for an SAP HANA system as a DR system using cross-region replication
- Manage volumes in an application volume group
- Delete an application volume group
- Application volume group FAQs
- Troubleshoot application volume group errors

# Development and test environments for SAP workloads on Azure

Azure ExpressRoute

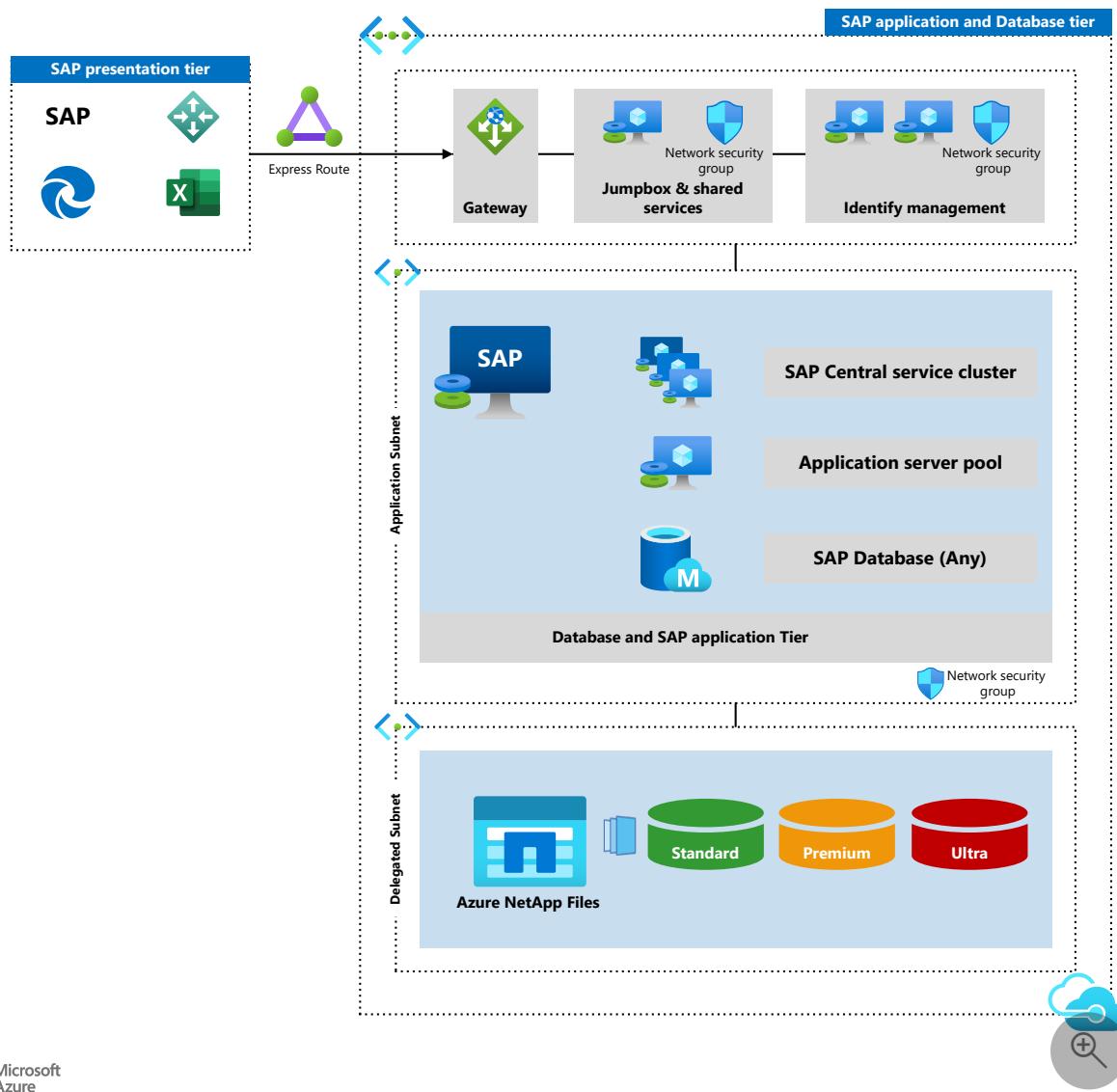
Azure Virtual Machines

Azure Virtual Network

Azure Resource Manager

This example shows how to establish a development and test environment for SAP NetWeaver in a Windows or Linux environment on Azure. The database used is AnyDB. (AnyDB is the SAP term for any supported DBMS that isn't SAP HANA.)

## Architecture



Download a [Visio file](#) of this architecture.

# Dataflow

This scenario demonstrates provisioning a single SAP system database and SAP application server on a single virtual machine. The data flows through the scenario as follows:

1. Customers use the SAP user interface or other client tools (Excel, a web browser, or other web application) to access the Azure-based SAP system.
2. Connectivity is provided by using an established ExpressRoute. The ExpressRoute connection is ended in Azure at the ExpressRoute gateway. Network traffic routes through the ExpressRoute gateway to the gateway subnet, and from the gateway subnet to the application-tier spoke subnet (see the [hub-spoke network topology](#)) and via a Network Security Gateway to the SAP application virtual machine.
3. The identity management servers provide authentication services.
4. The jump box provides local management capabilities.

## Components

- [Virtual networks](#) are the basis of network communication within Azure.
- [Azure Virtual Machines](#) provide on-demand, high-scale, secure, virtualized infrastructure using Windows or Linux servers.
- [Azure ExpressRoute](#) extends your on-premises networks into the Microsoft cloud over a private connection, which is facilitated by a connectivity provider.
- [Network security groups](#) limit network traffic to specific resources in a virtual network. A network security group contains a list of security rules that allow or deny inbound or outbound network traffic. The security rules are based on source or destination IP address, port, and protocol.
- [Resource groups](#) act as logical containers for Azure resources.
- [Azure Files](#) or [Azure NetApp Files](#) are recommended solutions to provide the storage for the SAP executables and HANA data and logs.

## Scenario details

Because this architecture is designed for non-production environments, it's deployed with only one virtual machine (VM). The VM size can be changed to accommodate your organization's needs.

For production use cases, review the SAP reference architectures available below:

- [SAP NetWeaver for AnyDB](#)
- [SAP S/4HANA](#)

- SAP on Azure large instances

## Potential use cases

Other relevant use cases include:

- Noncritical SAP nonproduction workloads (such sandbox, development, test, and quality assurance).
- Noncritical SAP business workloads.

## Considerations

These considerations implement the pillars of the Azure Well-Architected Framework, which is a set of guiding tenets that can be used to improve the quality of a workload. For more information, see [Microsoft Azure Well-Architected Framework](#).

Keep the following points in mind when establishing a development and test environment for SAP NetWeaver.

## Availability

Microsoft offers a service level agreement (SLA) for single VM instances. For more information on Microsoft Azure Service Level Agreement for Virtual Machines [SLA For Virtual Machines](#) ↗

## Scalability

For general guidance on designing scalable solutions, see the [performance efficiency checklist](#) in the Azure Architecture Center.

## Security

Security provides assurances against deliberate attacks and the abuse of your valuable data and systems. For more information, see [Overview of the security pillar](#).

For general guidance on designing secure solutions, see the [Azure Security Documentation](#).

## Data protection and cloning

For general guidance on protecting your application data, see [Azure Application Consistent Snapshot tool](#), which provides application consistent snapshots when used in combination with Azure NetApp Files.

## Resiliency

For general guidance on designing resilient solutions, see [Designing resilient applications for Azure](#).

## Cost optimization

Cost optimization is about looking at ways to reduce unnecessary expenses and improve operational efficiencies. For more information, see [Overview of the cost optimization pillar](#).

To help you explore the cost of running this scenario, all services are preconfigured in the cost calculator examples below. Change the appropriate variables to match the expected traffic for your use case.

We've provided four sample cost profiles based on amount of traffic you expect to receive:

[+] [Expand table](#)

Size	SAPs	VM Type	Storage	Azure Pricing Calculator
Small	8000	D8s_v3	2xP20, 1xP10	<a href="#">Small ↗</a>
Medium	16000	D16s_v3	3xP20, 1xP10	<a href="#">Medium ↗</a>
Large	32000	E32s_v3	3xP20, 1xP10	<a href="#">Large ↗</a>
Extra Large	64000	M64s	4xP20, 1xP10	<a href="#">Extra Large ↗</a>

### Note

This pricing is a guide that only indicates the VMs and storage costs. It excludes networking, backup storage, and data ingress/egress charges.

- [Small](#) : A small system consists of VM type D8s\_v3 with 8x vCPUs, 32-GB RAM, and 200 GB of temporary storage. It also contains premium storage: two 512-GB disks and one 128-GB disk.
- [Medium](#) : A medium system consists of VM type D16s\_v3 with 16x vCPUs, 64-GB RAM, and 400 GB of temporary storage. It also contains premium storage: three 512-GB disks and one 128-GB disk.
- [Large](#) : A large system consists of VM type E32s\_v3 with 32x vCPUs, 256-GB RAM, and 512 GB of temporary storage. It also contains premium storage: three 512-GB disks and one 128-GB disk.
- [Extra Large](#) : An extra-large system consists of a VM type M64s with 64x vCPUs, 1024-GB RAM, and 2000 GB of temporary storage. It also contains premium storage: four 512-GB disks and one 128-GB disk.

## Deploy this scenario

Select the link below to deploy the solution.



[Deploy to Azure](#)

### Note

SAP and Oracle are not installed during this deployment. You will need to deploy these components separately.

## Contributors

*This article is maintained by Microsoft. It was originally written by the following contributors.*

Principal author:

- [Andrew Dibbins](#) | Senior Engineer

*To see non-public LinkedIn profiles, sign in to LinkedIn.*

## Next steps

Learn more about the component technologies:

- [What is Azure Virtual Network?](#)

- [Linux virtual machines in Azure](#)
- [Windows virtual machines in Azure](#)
- [What is Azure ExpressRoute?](#)
- [Network security groups](#)
- [Use Azure to host and run SAP workload scenarios](#)
- [Installation of SAP HANA on Azure virtual machines](#)
- [Manage Azure Resource Manager resource groups by using Azure CLI](#)
- [High-availability architecture and scenarios for SAP NetWeaver](#)
- [What is Azure Files](#)
- [What is Azure NetApp Files](#)

## Related resources

Explore related architectures:

- [Run a Linux VM on Azure](#)
- [Run SAP NetWeaver in Windows on Azure](#)
- [SAP on Azure Architecture Guide](#)

# SAP BusinessObjects BI platform planning and implementation guide on Azure

Article • 06/16/2023

The purpose of this guide is to provide guidelines for planning, deploying, and configuring SAP BusinessObjects BI Platform, also known as SAP BOBI Platform on Azure. This guide is intended to cover common Azure services and features that are relevant for SAP BOBI Platform. This guide isn't an exhaustive list of all possible configuration options. It covers solutions common to typical deployment scenarios.

This guide isn't intended to replace the standard SAP BOBI Platform installation and administration guides, operating system, or any database documentation.

## Plan and implement SAP BusinessObjects BI platform on Azure

Microsoft Azure offers a wide range of services including compute, storage, networking, and many others for businesses to build their applications without lengthy procurement cycles. Azure virtual machines (VM) help companies to deploy on-demand and scalable computing resources for different SAP applications like SAP NetWeaver based applications, SAP Hybris, SAP BusinessObjects BI Platform, based on their business need. Azure also supports the cross-premises connectivity, which enables companies to integrate Azure virtual machines into their on-premises domains, their private clouds and their SAP system landscape.

This document provides guidance on planning and implementation consideration for SAP BusinessObjects BI Platform on Azure. It complements the SAP installation documentation and SAP Notes, which represent the primary resources for installations and deployments of SAP BOBI.

## Architecture overview

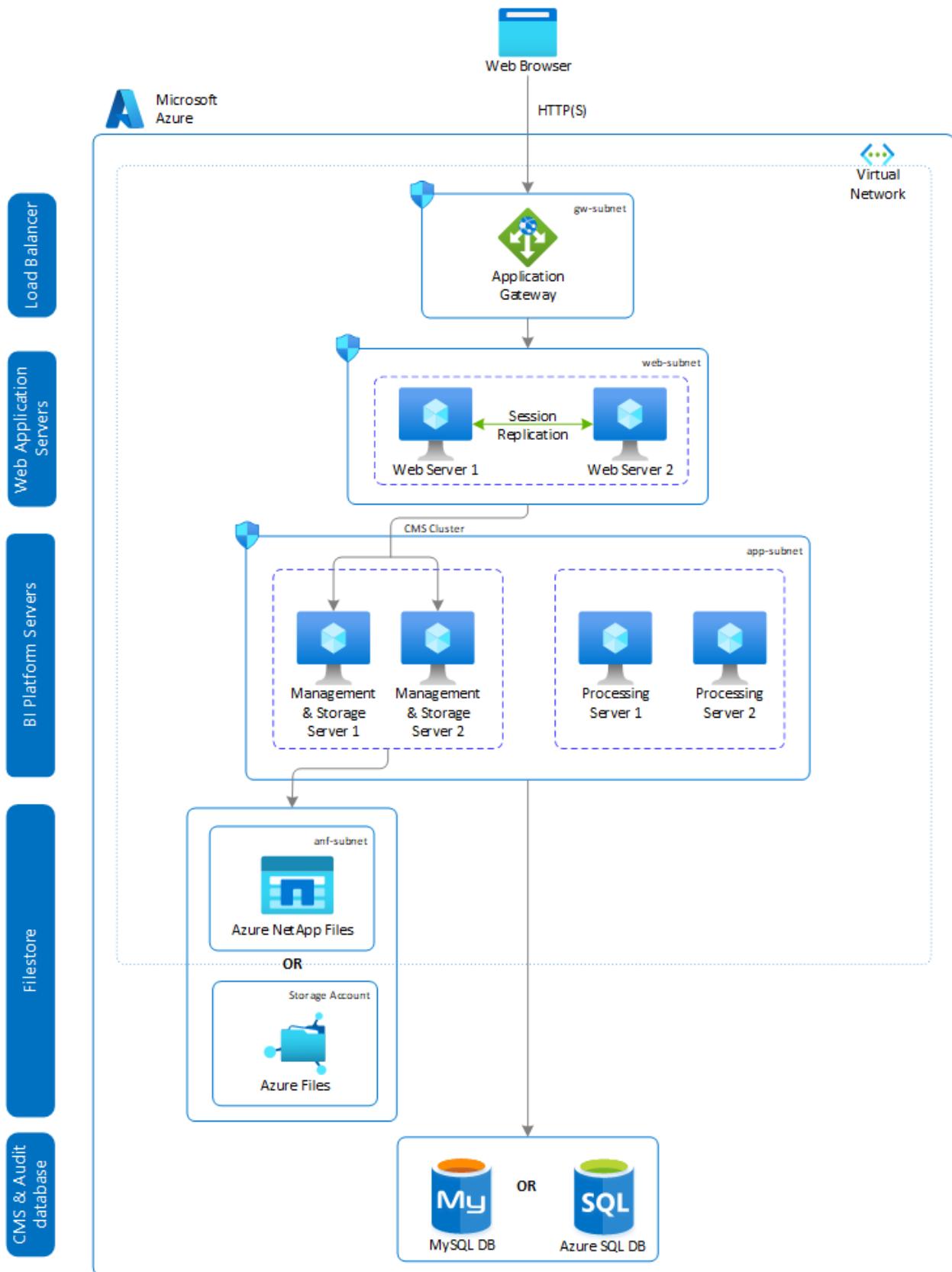
SAP BusinessObjects BI Platform is a self-contained system that can exist on a single Azure virtual machine or can be scaled into a cluster of many Azure Virtual Machines that run different components. SAP BOBI Platform consists of six conceptual tiers: Client Tier, Web Tier, Management Tier, Storage Tier, Processing Tier, and Data Tier. (For more

details on each tier, refer Administrator Guide in [SAP BusinessObjects Business Intelligence Platform](#) help portal). Following is the high-level details on each tier:

- **Client Tier:** It contains all desktop client applications that interact with the BI platform to provide different kind of reporting, analytic, and administrative capabilities.
- **Web Tier:** It contains web applications deployed to Java web application servers. Web applications provide BI Platform functionality to end users through a web browser.
- **Management Tier:** It coordinates and controls all the components that makes the BI Platform. It includes Central Management Server (CMS) and the Event Server and associated services
- **Storage Tier:** It's responsible for handling files, such as documents and reports. It also handles report caching to save system resources when user access reports.
- **Processing Tier:** It analyzes data, and produces reports and other output types. It's the only tier that accesses the databases that contain report data.
- **Data Tier:** It consists of the database servers hosting the CMS system databases and Auditing Data Store.

The SAP BI Platform consists of a collection of servers running on one or more hosts. It's essential that you choose the correct deployment strategy based on the sizing, business need and type of environment. For small installation like development or test, you can use a single Azure Virtual Machine for web application server, database server, and all BI Platform servers. In case you're using Database-as-a-Service (DBaaS) offering from Azure, database server runs separately from other components. For medium and large installation, you can have servers running on multiple Azure virtual machines.

The diagram below illustrates the architecture of a large-scale deployment of the SAP BOBI Platform on Azure virtual machines, with each component distributed. To ensure infrastructure resilience against service disruption, VMs can be deployed using either [flexible scale set](#), [availability sets](#) or [availability zones](#).



## Architecture details

- Load balancer

In SAP BOBI multi-instance deployment, Web application servers (or web tier) are running on two or more hosts. To distribute user load evenly across web servers, you can use a load balancer between end users and web servers. In Azure, you can

either use [Azure Load Balancer](#) or [Azure Application Gateway](#) to manage traffic to your web servers.

- Web application servers

The web server hosts the web applications of SAP BOBI Platform like CMC and BI Launch Pad. To achieve high availability for web server, you must deploy at least two web application servers to manage redundancy and load balancing. In Azure, these web application servers can be placed either in flexible scale set, availability zones or availability sets for better availability.

Tomcat is the default web application for SAP BI Platform. To achieve high availability for tomcat, enable session replication using [Static Membership Interceptor](#) in Azure. It ensures that user can access SAP BI web application even when tomcat service is disrupted.

 **Important**

By default Tomcat uses multicast IP and Port for clustering which is not supported on Azure (SAP Note [2764907](#)).

- BI platform servers

BI Platform servers include all the services that are part of SAP BOBI application (management tier, processing tier, and storage tier). When a web server receives a request, it detects each BI platform server (specifically, all CMS servers in a cluster) and automatically load balance their requests. In case if one of the BI Platform hosts fails, web server automatically send requests to other host.

To achieve high availability or redundancy for BI Platform, you must deploy the application in at least two Azure virtual machines. Based on the sizing, you can scale your BI Platform to run on more Azure virtual machines.

- File repository server (FRS)

File Repository Server contains all reports and other BI documents that have been created. In multi-instance deployment, BI Platform servers are running on multiple virtual machines and each VM should have access to these reports and other BI documents. So, a filesystem needs to be shared across all BI platform servers.

In Azure, you can either use [Azure Premium Files](#) or [Azure NetApp Files](#) for File Repository Server. Both of these Azure services have built-in redundancy.

- CMS & audit database

SAP BOBI Platform requires a database to store its system data, which is referred as CMS database. It's used to store BI platform information such as user, server, folder, document, configuration, and authentication details.

Azure offers [MySQL Database](#) and [Azure SQL database](#) Database-as-a-Service (DBaaS) offering that can be used for CMS database and Audit database. As this being a PaaS offering, customers don't have to worry about operation, availability, and maintenance of the databases. Customer can also choose their own database for CMS and Audit repository based on their business need.

## Support matrix

This section describes supportability of different SAP BOBI component like SAP BusinessObjects BI Platform version, Operating System and, Databases in Azure.

### SAP BusinessObjects BI platform

Azure Infrastructure as a Service (IaaS) enables you to deploy and configure SAP BusinessObjects BI Platform on Azure Compute. It supports following version of SAP BOBI Platform -

- SAP BusinessObjects BI Platform 4.3
- SAP BusinessObjects BI Platform 4.2 SP04+
- SAP BusinessObjects BI Platform 4.1 SP05+

The SAP BI Platform runs on different operating system and databases. Supportability of SAP BOBI platform between Operating System and Database version can be found in [Product Availability Matrix](#) for SAP BOBI.

### Operating system

Azure supports following operating systems for SAP BusinessObjects BI Platform deployment.

- Microsoft Windows Server
- SUSE Linux Enterprise Server (SLES)
- Red Hat Enterprise Linux (RHEL)
- Oracle Linux (OL)

The operating system version that is listed in [Product Availability Matrix \(PAM\)](#) for SAP BusinessObjects BI Platform are supported as long as they're compatible to run on Azure Infrastructure.

# Databases

The BI Platform needs database for CMS and Auditing Data store, which can be installed on any supported databases that are listed in [SAP Product Availability Matrix](#) that includes the following -

- Microsoft SQL Server
- [Azure SQL Database](#) (Supported database only for SAP BOBI Platform on Windows)

It's a fully managed SQL Server database engine, based on the latest stable Enterprise Edition of SQL Server. Azure SQL database handles most of the database management functions such as upgrading, patching, and monitoring without user involvement. With Azure SQL Database, you can create a highly available and high-performance data storage layer for the applications and solutions in Azure. For more details, check [Azure SQL Database](#) documentation.

- [Azure Database for MySQL](#) (Follow same compatibility guidelines as mentioned for MySQL AB in SAP PAM)

It's a relational database service powered by the MySQL community edition. Being a fully managed Database-as-a-Service (DBaaS) offering, it can handle mission-critical workloads with predictable performance and dynamic scalability. It has built-in high availability, automatic backups, software patching, automatic failure detection, and point-in-time restore for up to 35 days, which substantially reduce operation tasks. For more details, check [Azure Database for MySQL](#) documentation.

- SAP HANA
- SAP ASE
- IBM DB2
- Oracle (For version and restriction, check SAP Note [2039619](#))
- MaxDB

This document illustrates the guidelines to deploy **SAP BOBI Platform on Windows with Azure SQL Database** and **SAP BOBI Platform on Linux with Azure Database for MySQL**. It's also our recommended approach for running SAP BusinessObjects BI Platform on Azure.

# Sizing

Sizing is a process of determining the hardware requirement to run the application efficiently. For SAP BOBI Platform, sizing needs to be done using SAP sizing tool called [Quick Sizer](#). The tool provides the SAPS based on the input, which then needs to be mapped to certified Azure virtual machines types for SAP. SAP Note [1928533](#) provides the list of supported SAP products and Azure VM types along with SAPS. For more information on sizing, check [SAP BI Sizing Guide](#).

For storage need for SAP BOBI Platform, Azure offers different types of [Managed Disks](#). For SAP BOBI Installation directory, it's recommended to use premium managed disk and for the database that runs on virtual machines, follow the guidance that is provided in [DBMS deployment for SAP workload](#).

Azure supports two DBaaS offering for SAP BOBI Platform data tier - [Azure SQL Database](#) (BI Application running on Windows) and [Azure Database for MySQL](#) (BI Application running on Linux and Windows). So based on the sizing result, you can choose purchasing model that best fits your need.

## Tip

For quick sizing reference, consider 800 SAPS = 1 vCPU while mapping the SAPS result of SAP BOBI Platform database tier to Azure Database-as-a-Service (Azure SQL Database or Azure Database for MySQL).

## Sizing models for Azure SQL database

Azure SQL Database offers the following three purchasing models:

- vCore-based

It lets you choose the number of vCores, amount of memory, and the amount and speed of storage. The vCore-based purchasing model also allows you to use [Azure Hybrid Benefit for SQL Server](#) to gain cost savings. This model is suited for customer who value flexibility, control, and transparency.

There are three [Service Tier Options](#) being offered in vCore model that includes - General Purpose, Business Critical, and Hyperscale. The service tier defines the storage architecture, space, I/O limits, and business continuity options related to availability and disaster recovery. Following is high-level details on each service tier option -

1. **General Purpose** service tier is best suited for Business workloads. It offers budget-oriented, balanced, and scalable compute and storage options. For more information, refer [Resource options and limits](#).
2. **Business Critical** service tier offers business applications the highest resilience to failures by using several isolated replicas, and provides the highest I/O performance per database replica. For more information, refer [Resource options and limits](#).
3. **Hyperscale** service tier is best for business workloads with highly scalable storage and read-scale requirements. It offers higher resilience to failures by allowing configuration of more than one isolated database replica. For more information, refer [Resource options and limits](#).

- DTU-based

The DTU-based purchasing model offers a blend of compute, memory, and I/O resources in three service tiers, to support light and heavy database workloads. Compute sizes within each tier provide a different mix of these resources, to which you can add additional storage resources. It's best suited for customers who want simple, preconfigure resource options.

[Service Tiers](#) in the DTU-based purchasing model is differentiated by a range of compute sizes with a fixed amount of included storage, fixed retention period of backups, and fixed price.

- Serverless

The serverless model automatically scales compute based on workload demand, and bills for the amount of compute used per second. The serverless compute tier automatically pauses databases during inactive periods when only storage is billed, and automatically resumes databases when activity returns. For more information, refer [Resource options and limits](#).

It's more suitable for intermittent, unpredictable usage with low average compute utilization over time. So this model can be used for nonproduction SAP BOBI deployment.

 **Note**

For SAP BOBI, it's convenient to use vCore based model and choose either General Purpose or Business Critical service tier based on the business need.

## Sizing models for Azure database for MySQL

Azure Database for MySQL comes with three different pricing tiers. They're differentiated by the amount of compute in vCores, memory per vCore, and the storage technology used to store the date. Following is the high-level details on the options and for more details on different attributes, refer [Pricing Tier](#) for Azure Database for MySQL.

- Basic

It's used for the target workloads that require light compute and I/O performance.

- General Purpose

It's suited for most business workloads that require balanced compute and memory with scalable I/O throughput.

- Memory Optimized

For high-performance database workloads that require in-memory performance for faster transaction processing and higher concurrency.

 Note

For SAP BOBI, it is convenient to use General Purpose or Memory Optimized pricing tier based on the business workload.

## Azure resources

### Choosing regions

Azure region is one or a collection of data-centers that contains the infrastructure to run and hosts different Azure Services. This infrastructure includes large number of nodes that function as compute nodes or storage nodes, or run network functionality. Not all region offers the same services.

SAP BI Platform contains different components that might require specific VM types, Storage like Azure Files or Azure NetApp Files or Database as a Service (DBaaS) for its data tier that might not be available in certain regions. You can find out the exact information on VM types, Azure Storage types or, other Azure Services in [Products available by region](#) site. If you're already running your SAP systems on Azure, probably you have your region identified. In that case, you need to first investigate that the necessary services are available in those regions to decide the architecture of SAP BI Platform.

## Virtual machine scale sets with flexible orchestration

[Virtual machine scale sets](#) with flexible orchestration provide a logical grouping of platform-managed virtual machines. You have an option to create scale set within region or span it across availability zones. On creating, the flexible scale set within a region with `platformFaultDomainCount>1` ( $FD > 1$ ), the VMs deployed in the scale set would be distributed across specified number of fault domains in the same region. On the other hand, creating the flexible scale set across availability zones with `platformFaultDomainCount=1` ( $FD=1$ ) would distribute VMs across specified zone and the scale set would also distribute VMs across different fault domains within the zone on a best effort basis.

**For SAP workload only flexible scale set with FD=1 is supported.** The advantage of using flexible scale sets with  $FD=1$  for cross zonal deployment, instead of traditional availability zone deployment is that the VMs deployed with the scale set would be distributed across different fault domains within the zone in a best-effort manner. To learn more about SAP workload deployment with scale set, see [flexible virtual machine scale deployment guide](#).

## Availability zones

Availability Zones are physically separate locations within an Azure region. Each Availability Zone is made of one or more datacenters equipped with independent power, cooling, and networking.

To achieve high availability on each tier for SAP BI Platform, you can distribute VMs across Availability Zone by implementing high availability framework, which can provide the best SLA in Azure. For Virtual Machine SLA in Azure, check the latest version of [Virtual Machine SLAs ↗](#).

For data tier, Azure Database as a Service (DBaaS) service provides high availability framework by default. You just need to select the region and service inherent high availability, redundancy, and resiliency capabilities to mitigate database downtime from planned and unplanned outages, without requiring you to configure any additional components. For more details on the SLA for supported DBaaS offering on Azure, check [High availability in Azure Database for MySQL](#) and [High availability for Azure SQL Database](#).

## Availability sets

Availability set is a logical grouping capability for isolating Virtual Machine (VM) resources from each other on being deployed. Azure makes sure of the VMs you place

within an Availability Set run across multiple physical servers, compute racks, storage units, and network switches. If a hardware or software failure happens, only a subset of your VMs is affected and your overall solution stays operational. So when virtual machines are placed in availability sets, Azure Fabric Controller distributes the VMs over different [Fault ↗](#) and [Upgrade ↗](#) domains to prevent all VMs from being inaccessible because of infrastructure maintenance or failure within one Fault domain.

SAP BI Platform contains many different components and while designing the architecture you have to make sure that each of this component is resilient of any disruption. It can be achieved by placing Azure virtual machines of each component within availability sets. Keep in mind, when you mix VMs of different VM families within one availability set, you may come across problems that prevent you to include a certain VM type into such availability set. So have separate availability set for Web Application, BI Application for SAP BI Platform as highlighted in Architecture Overview.

Also the number of update and fault domains that can be used by an Azure Availability Set within an Azure Scale unit is finite. So if you keep adding VMs to a single availability set, two or more VMs will eventually end in the same fault or update domain. For more information, see the [Azure Availability Sets ↗](#) section of the Azure virtual machines planning and implementation for SAP document.

To understand the concept of Azure availability sets and the way availability sets relate to Fault and Upgrade Domains, read [manage availability](#) article.

#### **Important**

- The concepts of Azure availability zones and Azure availability sets are mutually exclusive. You can deploy a pair or multiple VMs into either a specific availability zone or an availability set, but you can't do both.
- If you planning to deploy across availability zones, it is advised to use **flexible scale set with FD=1** over standard availability zone deployment.

## Virtual machines

Azure Virtual Machine is a service offering that enables you to deploy custom images to Azure as Infrastructure-as-a-Service (IaaS) instances. It simplifies maintaining and operating applications by providing on-demand compute and storage to host, scale, and manage web application and connected applications.

Azure offers varieties of virtual machines for all your application needs. But for SAP workload, Azure has narrowed the selection to different VM families that are suitable for

SAP workload and SAP HANA workload more specifically. For more insight, check [What SAP software is supported for Azure deployments](#).

Based on the SAP BI Platform sizing, you need to map your requirement to Azure Virtual Machine, which is supported in Azure for SAP product. SAP Note [1928533](#) is a good starting point that lists supported Azure VM types for SAP Products on Windows and Linux. Also a point to keep in mind that beyond the selection of purely supported VM types, you also need to check whether those VM types are available in specific regions. You can check the availability of VM type on [Products available by region](#) page. For choosing the pricing model, you can refer to [Azure virtual machines for SAP workload](#)

## Storage

Azure Storage is an Azure-managed cloud service that provides storage that is highly available, secure, durable, scalable, and redundant. Some of the storage types have limited use for SAP scenarios. But several Azure Storage types are well suited or optimized for specific SAP workload scenarios. For more information, refer [Azure Storage types for SAP Workload](#) guide, as it highlights different storage options that are suited for SAP.

Azure Storage has different Storage types available for customers and details for the same can be read in the article [What disk types are available in Azure?](#). SAP BOBI Platform uses following Azure Storage to build the application -

- Azure-managed disks

It's a block-level storage volume that is managed by Azure. You can use the disks for SAP BOBI Platform application servers and databases, when installed on Azure virtual machines. There are different types of [Azure Managed Disks](#) available, but it's recommended to use [Premium SSDs](#) for SAP BOBI Platform application and database.

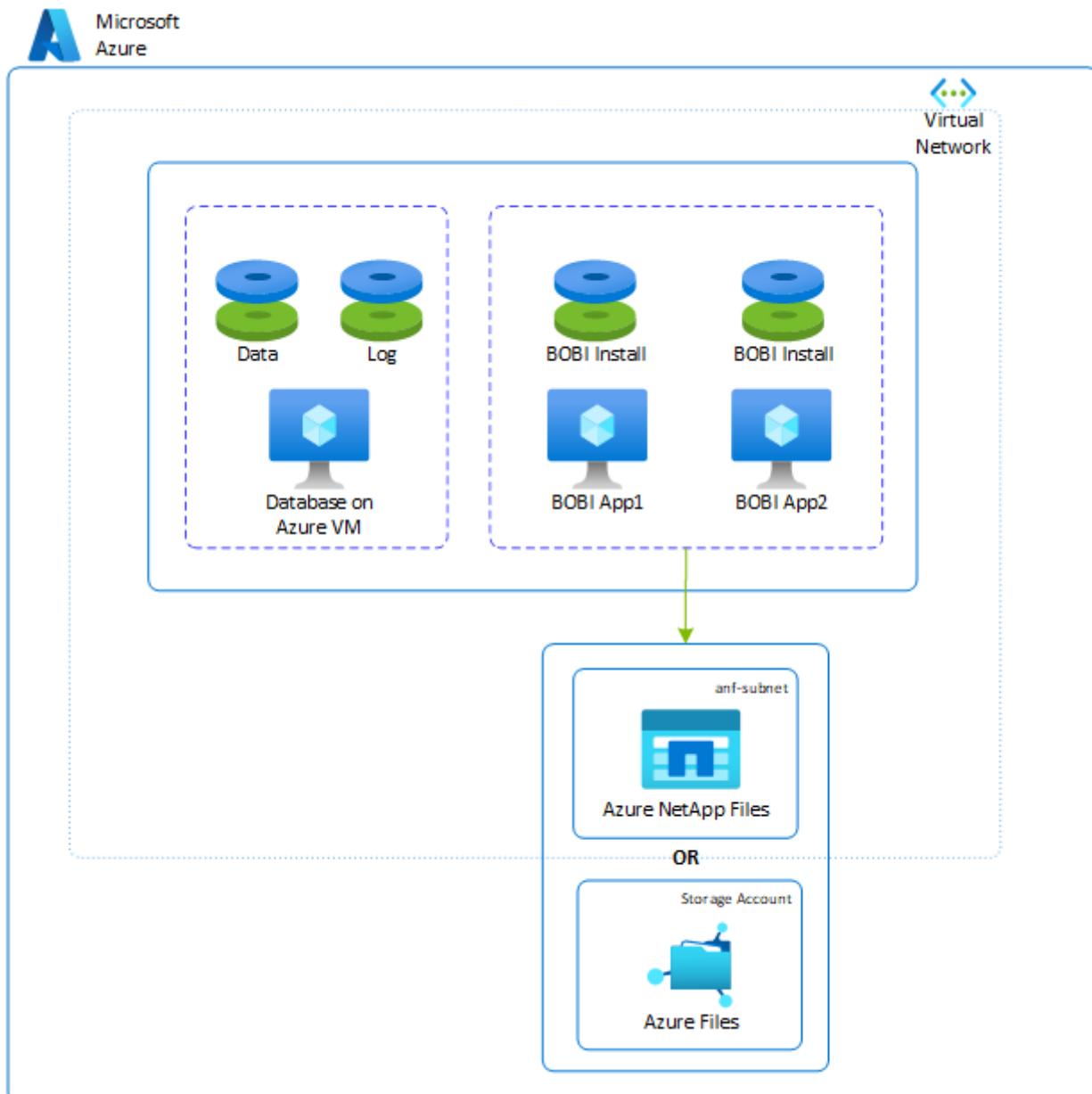
In below example, Premium SSDs are used for BOBI Platform installation directory. For database installed on virtual machine, you can use managed disks for data and log volume as per the guidelines. CMS and Audit databases are typically small and it doesn't have the same storage performance requirements as that of other SAP OLTP/OLAP databases.

- Azure Premium Files or Azure NetApp Files

In SAP BOBI Platform, File Repository Server (FRS) refers to the disk directories where contents like reports, universes, and connections are stored which are used by all application servers of that system. [Azure Premium Files](#) or [Azure NetApp](#)

[Files](#) storage can be used as a shared file system for SAP BOBI applications FRS. As this storage offering isn't available all regions, refer to [Products available by region](#) site to find out up-to-date information.

If the service is unavailable in your region, you can create NFS server from which you can share the file system to SAP BOBI application. But you'll also need to consider its high availability.



## Networking

SAP BOBI is a reporting and analytics BI platform that doesn't hold any business data. So the system is connected to other database servers from where it fetches all the data and provide insight to users. Azure provides a network infrastructure, which allows the mapping of all scenarios that can be realized with SAP BI Platform like connecting to on-premises system, systems in different virtual network and others. For more information check [Microsoft Azure Networking for SAP Workload](#).

For Database-as-a-Service offering, any newly created database (Azure SQL Database or Azure Database for MySQL) has a firewall that blocks all external connections. To allow access to the DBaaS service from BI Platform virtual machines, you need to specify one or more server-level firewall rules to enable access to your DBaaS server. For more information, see [Firewall rules](#) for Azure Database for MySQL and [Network Access Controls](#) section for Azure SQL database.

## Next steps

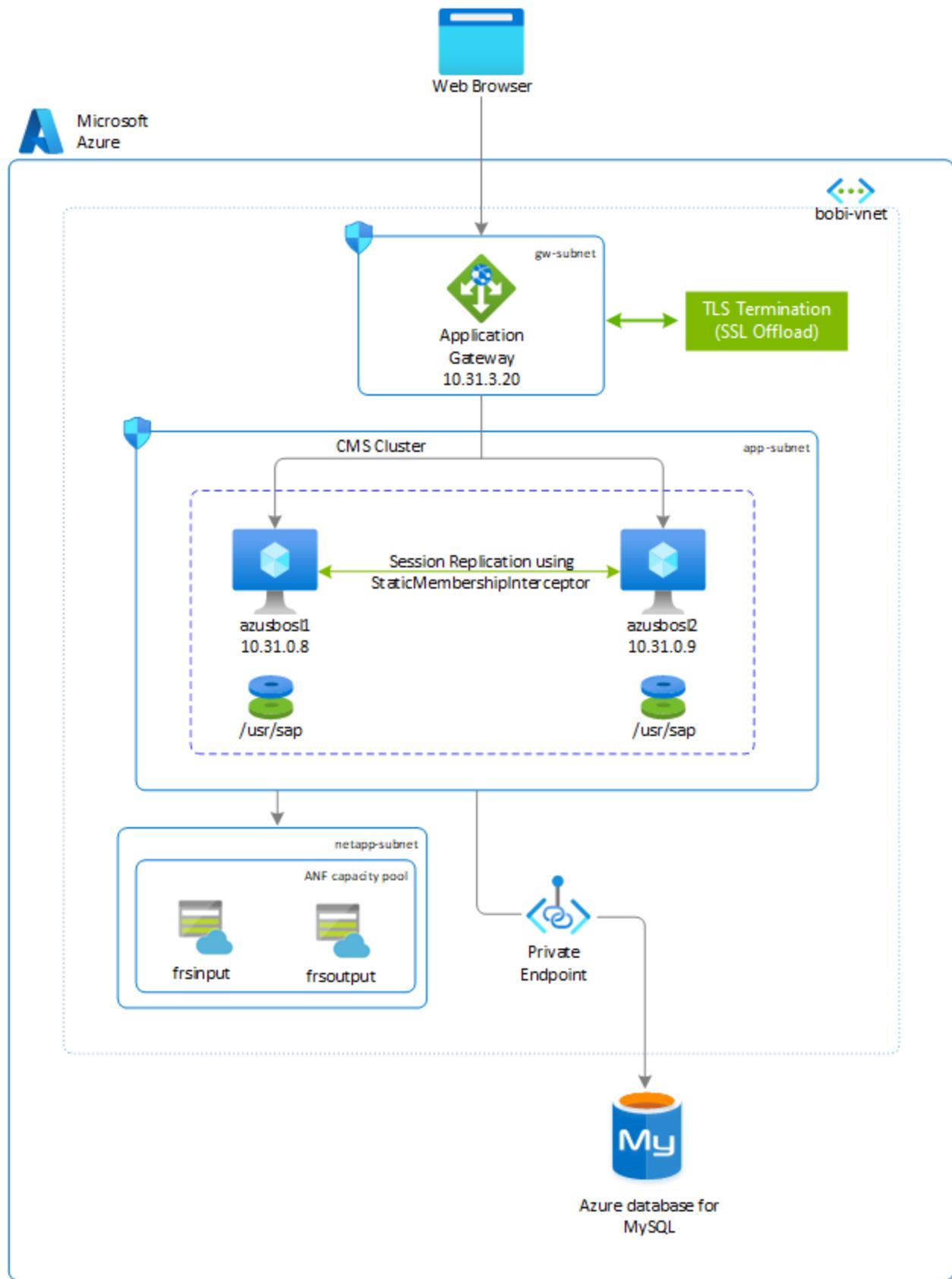
- [SAP BusinessObjects BI Platform Deployment on Linux](#)
- [Azure Virtual Machines planning and implementation for SAP](#)
- [Azure Virtual Machines deployment for SAP](#)
- [Azure Virtual Machines DBMS deployment for SAP](#)

# SAP BusinessObjects BI platform deployment guide for Linux on Azure

Article • 06/15/2023

This article describes the strategy to deploy SAP BusinessObjects BI (BOBI) platform on Azure for Linux. In this example, you configure two virtual machines with premium solid-state drive (SSD) managed disks as the install directory. You use Azure Database for MySQL for your CMS database, and you share Azure NetApp Files for your file repository server across both servers. On both virtual machines, you install the default Tomcat Java web application and BI platform application together. To load-balance user requests, you use Azure Application Gateway with native TLS/SSL offloading capabilities.

This type of architecture is effective for small deployments or non-production environments. For large deployments or production environments, you can have separate hosts for your web application. You can also have multiple BOBI application hosts, allowing the server to process more information.



Here's the product version and file system layout for this example:

- SAP BusinessObjects platform 4.3
- SUSE Linux Enterprise Server 12 SP5
- Azure Database for MySQL (Version: 8.0.15)
- MySQL C API Connector - libmysqlclient (Version: 6.1.11)

File system	Description	Size (GB)	Owner	Group	Storage
/usr/sap	The file system for installation of the SAP BOBI instance, the default Tomcat web application, and the database drivers (if necessary).	SAP sizing guidelines	bl1adm	sapsys	Managed premium disk - SSD
/usr/sap/frsinput	The mount directory is for the shared files across all BOBI hosts that will be used as the input file repository directory.	Business need	bl1adm	sapsys	Azure NetApp Files
/usr/sap/frsoutput	The mount directory is for the shared files across all BOBI hosts that will be used as the output file repository directory	Business need	bl1adm	sapsys	Azure NetApp Files

### ⓘ Important

While the setup of the SAP BusinessObjects platform is explained using Azure NetApp Files, you could use NFS on Azure Files as the input and output file repository.

## Deploy Linux virtual machine via Azure portal

In this section, you create two virtual machines with the Linux operating system image for the SAP BOBI platform. The high-level steps to create the virtual machines are as follows:

1. Create a [resource group](#).
2. Create a [virtual network](#).
  - Don't use a single subnet for all Azure services in the SAP BI platform deployment. Based on SAP BI platform architecture, you need to create multiple subnets. In this deployment, you create three subnets: one each for the application, the file repository store, and Application Gateway.
  - In Azure, Application Gateway and Azure NetApp Files must always be on a separate subnet. For more information, see [Azure Application Gateway](#) and [Guidelines for Azure NetApp Files network planning](#).
3. Select the suitable [availability options](#) depending on your preferred system configuration within an Azure region, whether it involves spanning across zones,

residing within a single zone, or operating in a zone-less region.

4. Create virtual machine 1, called (azusbosl1).
  - You can either use a custom image or choose an image from Azure Marketplace. For more information, see [Deploying a VM from the Azure Marketplace for SAP](#) or [Deploying a VM with a custom image for SAP](#).
5. Create virtual machine 2, called (azusbosl2).
6. Add one premium SSD disk. You'll use it as your SAP BOBI Installation directory.

## Provision Azure NetApp Files

Before you continue with the setup for Azure NetApp Files, familiarize yourself with the [Azure NetApp Files documentation](#).

Azure NetApp Files is available in several [Azure regions](#). Check to see whether your selected Azure region offers Azure NetApp Files.

Use [Azure NetApp Files availability by Azure Region](#) to check the availability of Azure NetApp Files by region.

## Deploy Azure NetApp Files resources

The following instructions assume that you've already deployed your [Azure virtual network](#). The Azure NetApp Files resources, and the VMs where the Azure NetApp Files resources will be mounted, must be deployed in the same Azure virtual network or in peered Azure virtual networks.

1. [Create an Azure NetApp Files account](#) in your selected Azure region.
2. [Set up an Azure NetApp Files capacity pool](#). The SAP BI platform architecture presented in this article uses a single Azure NetApp Files capacity pool at the Premium service level. For SAP BI File Repository Server on Azure, we recommend using an Azure NetApp Files Premium or Ultra [service Level](#).
3. [Delegate a subnet to Azure NetApp Files](#).
4. Deploy Azure NetApp Files volumes by following the instructions in [Create an NFS volume for Azure NetApp Files](#).

You can deploy the volumes as NFSv3 and NFSv4.1, because both protocols are supported for the SAP BOBI platform. Deploy the volumes in their respective Azure

NetApp Files subnets. The IP addresses of the Azure NetApp Files volumes are assigned automatically.

Keep in mind that the Azure NetApp Files resources and the Azure VMs must be in the same Azure virtual network or in peered Azure virtual networks. For example, *azusbobi-frsinput* and *azusbobi-frsoutput* are the volume names, and *nfs://10.31.2.4/azusbobi-frsinput* and *nfs://10.31.2.4/azusbobi-frsoutput* are the file paths for the Azure NetApp Files volumes.

- Volume azusbobi-frsinput (*nfs://10.31.2.4/azusbobi-frsinput*)
- Volume azusbobi-frsoutput (*nfs://10.31.2.4/azusbobi-frsoutput*)

## Important considerations

As you're creating your Azure NetApp Files for SAP BOBI platform file repository server, be aware of the following considerations:

- The minimum capacity pool is 4 tebibytes (TiB). The capacity pool size can be increased in 1 TiB increments.
- The minimum volume size is 100 gibibytes (GiB).
- Azure NetApp Files and all virtual machines where the Azure NetApp Files volumes will be mounted must be in the same Azure virtual network, or in [peered virtual networks](#) in the same region. Azure NetApp Files access over virtual network peering in the same region is supported. Azure NetApp Files access over global peering isn't currently supported.
- The selected virtual network must have a subnet that is delegated to Azure NetApp Files.
- The throughput and performance characteristics of an Azure NetApp Files volume is a function of the volume quota and service level, as documented in [Service level for Azure NetApp Files](#). While sizing the SAP Azure NetApp volumes, make sure that the resulting throughput meets the application requirements.
- With the Azure NetApp Files [export policy](#), you can control the allowed clients, the access type (for example, read-write or read only).
- The Azure NetApp Files feature isn't zone-aware yet. Currently, the feature isn't deployed in all availability zones in an Azure region. Be aware of the potential latency implications in some Azure regions.
- Azure NetApp Files volumes can be deployed as NFSv3 or NFSv4.1 volumes. Both protocols are supported for the SAP BI platform applications.

## Configure file systems on Linux servers

The steps in this section use the following prefix:

[A]: The step applies to all hosts.

## Format and mount the SAP file system

1. [A] List all attached disks.

Bash

```
sudo lsblk
NAME   MAJ:MIN RM  SIZE RO TYPE MOUNTPOINT
sda      8:0    0   30G  0 disk
└─sda1   8:1    0    2M  0 part
└─sda2   8:2    0  512M  0 part /boot/efi
└─sda3   8:3    0    1G  0 part /boot
└─sda4   8:4    0 28.5G  0 part /
sdb      8:16   0   32G  0 disk
└─sdb1   8:17   0   32G  0 part /mnt
sdc      8:32   0  128G  0 disk
sr0     11:0    1  628K  0 rom
# Premium SSD of 128 GB is attached to virtual machine, whose device
name is sdc
```

2. [A] Format the block device for /usr/sap.

Bash

```
sudo mkfs.xfs /dev/sdc
```

3. [A] Create the mount directory.

Bash

```
sudo mkdir -p /usr/sap
```

4. [A] Get the UUID of the block device.

Bash

```
sudo blkid
```

```
# It will display information about block device. Copy UUID of the
formatted block device
```

```
/dev/sdc: UUID="0eb5f6f8-fa77-42a6-b22d-7a9472b4dd1b" TYPE="xfs"
```

5. [A] Maintain the file system mount entry in /etc/fstab.

Bash

```
sudo echo "UUID=0eb5f6f8-fa77-42a6-b22d-7a9472b4dd1b /usr/sap xfs  
defaults,nofail 0 2" >> /etc/fstab
```

6. [A] Mount the file system.

Bash

```
sudo mount -a
```

```
sudo df -h
```

Filesystem	Size	Used	Avail	Use%	Mounted on
devtmpfs	7.9G	8.0K	7.9G	1%	/dev
tmpfs	7.9G	82M	7.8G	2%	/run
tmpfs	7.9G	0	7.9G	0%	/sys/fs/cgroup
/dev/sda4	29G	1.8G	27G	6%	/
tmpfs	1.6G	0	1.6G	0%	/run/user/1000
/dev/sda3	1014M	87M	928M	9%	/boot
/dev/sda2	512M	1.1M	511M	1%	/boot/efi
/dev/sdb1	32G	49M	30G	1%	/mnt
/dev/sdc	128G	29G	100G	23%	/usr/sap

## Mount the Azure NetApp Files volume

1. [A] Create mount directories.

Bash

```
sudo mkdir -p /usr/sap/frsinput  
sudo mkdir -p /usr/sap/frsoutput
```

2. [A] Configure the client operating system to support NFSv4.1 Mount (only applicable if using NFSv4.1).

If you're using Azure NetApp Files volumes with NFSv4.1 protocol, run the following configuration on all VMs where Azure NetApp Files NFSv4.1 volumes need to be mounted.

In this step, you need to verify NFS domain settings. Make sure that the domain is configured as the default Azure NetApp Files domain (`defaultv4iddomain.com`), and that the mapping is set to `nobody`.

Bash

```
sudo cat /etc/idmapd.conf
# Example
[General]
Domain = defaultv4iddomain.com
[Mapping]
Nobody-User = nobody
Nobody-Group = nobody
```

### ⓘ Important

Make sure to set the NFS domain in /etc/idmapd.conf on the VM to match the default domain configuration on Azure NetApp Files (defaultv4iddomain.com). If there's a mismatch, then the permissions for files on Azure NetApp Files volumes that are mounted on the VMs will be displayed as nobody.

Verify `nfs4_disable_idmapping`. It should be set to Y. To create the directory structure where `nfs4_disable_idmapping` is located, run the mount command. You won't be able to manually create the directory under /sys/modules, because access is reserved for the kernel / drivers.

Bash

```
# Check nfs4_disable_idmapping
cat /sys/module/nfs/parameters/nfs4_disable_idmapping

# If you need to set nfs4_disable_idmapping to Y
mkdir /mnt/tmp
mount -t nfs -o sec=sys,vers=4.1 10.31.2.4:/azusbobi-frsinput /mnt/tmp
umount /mnt/tmp

echo "Y" > /sys/module/nfs/parameters/nfs4_disable_idmapping

# Make the configuration permanent
echo "options nfs nfs4_disable_idmapping=Y" >> /etc/modprobe.d/nfs.conf
```

### 3. [A] Add mount entries.

If you're using NFSv3:

Bash

```
sudo echo "10.31.2.4:/azusbobi-frsinput /usr/sap/frsinput nfs
rw,hard,rsize=65536,wsize=65536,vers=3" >> /etc/fstab
```

```
sudo echo "10.31.2.4:/azusbobi-frsoutput /usr/sap/frsoutput nfs  
rw,hard,rsize=65536,wsize=65536,vers=3" >> /etc/fstab
```

If you're using NFSv4.1:

Bash

```
sudo echo "10.31.2.4:/azusbobi-frsinput /usr/sap/frsinput nfs  
rw,hard,rsize=65536,wsize=65536,vers=4.1,sec=sys" >> /etc/fstab  
sudo echo "10.31.2.4:/azusbobi-frsoutput /usr/sap/frsoutput nfs  
rw,hard,rsize=65536,wsize=65536,vers=4.1,sec=sys" >> /etc/fstab
```

#### 4. [A] Mount NFS volumes.

Bash

```
sudo mount -a  
  
sudo df -h  
  
Filesystem      Size   Used  Avail Use% Mounted on  
devtmpfs        7.9G   8.0K  7.9G  1% /dev  
tmpfs           7.9G   82M  7.8G  2% /run  
tmpfs           7.9G    0  7.9G  0% /sys/fs/cgroup  
/dev/sda4        29G   1.8G  27G  6% /  
tmpfs           1.6G    0  1.6G  0% /run/user/1000  
/dev/sda3       1014M  87M  928M  9% /boot  
/dev/sda2       512M  1.1M  511M  1% /boot/efi  
/dev/sdb1        32G   49M  30G  1% /mnt  
/dev/sdc         128G   29G  100G  23% /usr/sap  
10.31.2.4:/azusbobi-frsinput  101T   18G  100T  1% /usr/sap/frsinput  
10.31.2.4:/azusbobi-frsoutput 100T  512K  100T  1% /usr/sap/frsoutput
```

## Configure Azure Database for MySQL

This section provides details on how to provision Azure Database for MySQL by using the Azure portal. It also provides instructions on how to create the CMS and audit databases for the SAP BOBI platform, and a user account to access the database.

The guidelines are applicable only if you're using Azure Database for MySQL. For other databases, refer to the SAP or database-specific documentation for instructions.

### Create a database

Sign in to the Azure portal, and follow the steps in [Quickstart: Create an Azure Database for MySQL server by using the Azure portal](#). Here are a few points to note while you're

provisioning Azure Database for MySQL:

- Select the same region for Azure Database for MySQL as where your SAP BI platform application servers are running.
- Choose a supported database version, based on the [Product Availability Matrix \(PAM\) for SAP BI](#) specific to your SAP BOBI version.
- In **compute+storage**, select **Configure server**, and select the appropriate pricing tier based on your sizing output.
- **Storage Autogrowth** is enabled by default. Keep in mind that **storage** can only be scaled-up, not down.
- By default, **Back up Retention Period** is seven days. You can [optionally configure](#) it up to 35 days.
- Backups of Azure Database for MySQL are locally redundant by default. If you want server backups in geo-redundant storage, select **Geographically Redundant** from **Backup Redundancy Options**.

**ⓘ Important**

Changing the **Backup Redundancy Options** after server creation isn't supported.

**ⓘ Note**

The private link feature is only available for Azure Database for MySQL servers in the General Purpose or Memory Optimized pricing tiers. Ensure that the database server is in one of these pricing tiers.

## Configure Azure Private Link

In this section, you create a private link that allows SAP BOBI virtual machines to connect to Azure Database for MySQL through a private endpoint. Azure Private Link brings Azure services inside your private virtual network.

1. Select the database created in the previous section.
2. Go to **Security > Private endpoint connections**.
3. In **Private endpoint connections**, select **Private endpoint**.
4. Select **Subscription > Resource group > Location**.
5. Enter the **Name** of the private endpoint.

6. In the **Resource** section, specify the following:

- Resource type: Microsoft.DBforMySQL/servers
- Resource: MySQL database created in the previous section
- Target sub-resource: mysqlServer

7. In the **Networking** section, select the **Virtual network** and **Subnet** on which the SAP BOBI application is deployed.

**!** **Note**

If you have a network security group (NSG) enabled for the subnet, it will be disabled for private endpoints on this subnet only. Other resources on the subnet will still have NSG enforcement.

8. For **Integrate with private DNS zone**, accept the **default (yes)**.
9. Select your **private DNS zone** from the dropdown list.
10. Select **Review+Create**, and create a private endpoint.

For more information, see [Private Link for Azure Database for MySQL](#).

## Create the CMS and audit databases

1. Download and install MySQL Workbench from [MySQL website](#). Make sure you install MySQL Workbench on the server that can access Azure Database for MySQL.
2. Connect to the server by using MySQL Workbench. Follow the instructions in [Get connection information](#). If the connection test is successful, you get following message:

MySQL Workbench

**i** Successfully made the MySQL connection

Information related to this connection:

Host: azusbomysql.mysql.database.azure.com  
Port: 3306  
User: testuser@azusbomysql  
SSL: enabled with ECDHE-RSA-AES128-GCM-SHA256

A successful MySQL connection was made with the parameters defined for this connection.

OK

3. In the SQL query tab, run the following query to create a schema for the CMS and audit databases.

```
SQL

# Here cmsbl1 is the database name of CMS database. You can provide the
# name you want for CMS database.
CREATE SCHEMA `cmsbl1` DEFAULT CHARACTER SET utf8;

# auditbl1 is the database name of Audit database. You can provide the
# name you want for CMS database.
CREATE SCHEMA `auditbl1` DEFAULT CHARACTER SET utf8;
```

4. Create a user account to connect to the schema.

```
SQL

# Create a user that can connect from any host, use the '%' wildcard as
# a host part
CREATE USER 'cmsadmin'@'%' IDENTIFIED BY 'password';
CREATE USER 'auditadmin'@'%' IDENTIFIED BY 'password';

# Grant all privileges to a user account over a specific database:
GRANT ALL PRIVILEGES ON cmsbl1.* TO 'cmsadmin'@'%' WITH GRANT OPTION;
GRANT ALL PRIVILEGES ON auditbl1.* TO 'auditadmin'@'%' WITH GRANT
OPTION;

# Following any updates to the user privileges, be sure to save the
# changes by issuing the FLUSH PRIVILEGES
FLUSH PRIVILEGES;
```

5. To check the privileges and roles of the MySQL user account:

```
SQL

USE sys;
SHOW GRANTS FOR 'cmsadmin'@'%';
+-----+
| Grants for cmsadmin@%
| |
+-----+
| GRANT USAGE ON *.* TO `cmsadmin`@`%` 
| |
| GRANT ALL PRIVILEGES ON `cmsbl1`.* TO `cmsadmin`@`%` WITH GRANT
OPTION |
+-----+
| USE sys;
```

```
SHOW GRANTS FOR 'auditadmin'@'%';
+
+-----+
| Grants for auditadmin@%
|
+
+-----+
| GRANT USAGE ON *.* TO `auditadmin`@`%`
|
| GRANT ALL PRIVILEGES ON `auditbl1`.* TO `auditadmin`@`%` WITH GRANT
OPTION |
+
+-----+
```

## Install MySQL C API connector on a Linux server

For the SAP BOBI application server to access a database, it requires database client drivers. To access the CMS and audit databases, you must use the MySQL C API Connector for Linux. An ODBC connection to the CMS database isn't supported. This section provides instructions on how to set up MySQL C API Connector on Linux.

1. Refer to [MySQL drivers and management tools compatible with Azure Database for MySQL](#). Check for the **MySQL Connector/C (libmysqlclient)** driver in the article.
2. To download drivers, see [MySQL Product Archives](#).
3. Select the operating system and download the shared component rpm package of MySQL Connector. In this example, mysql-connector-c-shared-6.1.11 connector version is used.
4. Install the connector in all SAP BOBI application instances.

Bash

```
# Install rpm package
SLES: sudo zypper install <package>.rpm
RHEL: sudo yum install <package>.rpm
```

5. Check the path of libmysqlclient.so.

Bash

```
# Find the location of libmysqlclient.so file
whereis libmysqlclient

# sample output
libmysqlclient: /usr/lib64/libmysqlclient.so
```

6. Set `LD_LIBRARY_PATH` to point to the `/usr/lib64` directory for the user account that will be used for installation.

Bash

```
# This configuration is for bash shell. If you are using any other
# shell for sidadm, kindly set environment variable accordingly.
vi /home/bl1adm/.bashrc

export LD_LIBRARY_PATH=/usr/lib64
```

## Server preparation

The steps in this section use the following prefix:

[A]: The step applies to all hosts.

1. [A] Based on the flavor of Linux (SLES or RHEL), you need to set kernel parameters and install required libraries. Refer to the "System requirements" section in [Business Intelligence Platform Installation Guide for Unix](#).
2. [A] Ensure that the time zone on your machine is set correctly. In the Installation Guide, see [Additional Unix and Linux requirements](#).
3. [A] Create user account (`bl1adm`) and group (`sapsys`) under which the software's background processes can run. Use this account to run the installation and the software. The account doesn't require root privileges.
4. [A] Set the user account (`bl1adm`) environment to use a supported UTF-8 locale, and ensure that your console software supports UTF-8 character sets. To ensure that your operating system uses the correct locale, set the `LC_ALL` and `LANG` environment variables to your preferred locale in your (`bl1adm`) user environment.

Bash

```
# This configuration is for bash shell. If you are using any other
# shell for sidadm, kindly set environment variable accordingly.
vi /home/bl1adm/.bashrc

export LANG=en_US.utf8
export LC_ALL=en_US.utf8
```

5. [A] Configure user account (`bl1adm`).

Bash

```

# Set ulimit for bl1adm to unlimited
root@azusbosl1:~> ulimit -f unlimited bl1adm
root@azusbosl1:~> ulimit -u unlimited bl1adm

root@azusbosl1:~> su - bl1adm
bl1adm@azusbosl1:~> ulimit -a

core file size          (blocks, -c) unlimited
data seg size           (kbytes, -d) unlimited
scheduling priority     (-e) 0
file size               (blocks, -f) unlimited
pending signals          (-i) 63936
max locked memory       (kbytes, -l) 64
max memory size         (kbytes, -m) unlimited
open files              (-n) 1024
pipe size                (512 bytes, -p) 8
POSIX message queues    (bytes, -q) 819200
real-time priority      (-r) 0
stack size               (kbytes, -s) 8192
cpu time                 (seconds, -t) unlimited
max user processes       (-u) unlimited
virtual memory           (kbytes, -v) unlimited
file locks               (-x) unlimited

```

6. Download and extract media for SAP BusinessObjects BI platform from SAP Service Marketplace.

## Installation

Check the locale for user account **bl1adm** on the server:

Bash

```

bl1adm@azusbosl1:~> locale
LANG=en_US.utf8
LC_ALL=en_US.utf8

```

Go to the media of SAP BOBI platform, and run the following command with **bl1adm** user:

Bash

```

./setup.sh -InstallDir /usr/sap/BL1

```

Follow the [SAP BOBI platform](#) Installation Guide for Unix, specific to your version.

Here are a few points to note while you're installing the SAP BOBI platform:

- On **Configure Product Registration**, you can either use a temporary license key for SAP BusinessObjects Solutions from SAP Note [1288121](#), or you can generate a license key in SAP Service Marketplace.
- On **Select Install Type**, select **Full** installation on the first server (`azusbos11`). For the other server (`azusbos12`), select **Custom / Expand**, which will expand the existing BOBI setup.
- On **Select Default or Existing Database**, select **configure an existing database**, which will prompt you to select CMS and audit databases. Select **MySQL** for these database types.

You can also select **No auditing database**, if you don't want to configure auditing during installation.

- On **Select Java Web Application Server screen**, select appropriate options based on your SAP BOBI architecture. In this example, we have selected option 1, which installs a tomcat server on the same SAP BOBI platform.
- Enter CMS database information in **Configure CMS Repository Database - MySQL**. The following example shows input for CMS database information for a Linux installation. Azure Database for MySQL is used on default port 3306.

```
Configure CMS Repository Database - MySQL
Enter connection information about the existing database to use as the CMS repository

CMS Database Name
[cmsb11]                               ]

MySQL Server
[azusbomysql.mysql.database.azure.com]   ]

MySQL Port
[3306]                                  ]

User Name
[cmsadmin@azusbomysql]                  ]

Password
[*****]                                 ]

Reset existing database (1 = yes, 0 = no)
[1]
```

- (Optional) Enter audit database information in **Configure Audit Repository Database - MySQL**. The following example shows input for audit database information for a Linux installation.

```
Configure Auditing Database - MySQL
Enter connection information about the existing database to use for auditing

Auditing Database Name
[auditb11]                               ]

MySQL Server
[azusbomysql.mysql.database.azure.com]   ]

MySQL Port
[3306]                                  ]

User Name
[auditadmin@azusbomysql]                ]

Password
[*****]                                 ]
```

- Follow the instructions and enter required inputs to complete the installation.

For multi-instance deployment, run the installation setup on a second host (`azusbos12`).

For **Select Install Type**, select **Custom / Expand**, which will expand the existing BOBI setup.

In Azure Database for MySQL, a gateway redirects the connections to server instances. After the connection is established, the MySQL client displays the version of MySQL set in the gateway, not the actual version running on your MySQL server instance. To determine the version of your MySQL server instance, use the `SELECT VERSION();` command at the MySQL prompt. For more details, see [Supported Azure Database for MySQL server versions](#).

Properties	
<b>Data Source:</b>	<code>cmsbl1:azusbomysql.mysql.database.azure.com:3306</code>
<b>Database Name:</b>	<code>5.6.42.0</code>
<b>Database User Name:</b>	<code>cmsadmin@azusbomysql</code>
<b>Auditing:</b>	<code>Enabled</code>

```
SQL

# Run direct query to the database using MySQL Workbench

select version();

+-----+
| version() |
+-----+
| 8.0.15    |
+-----+
```

## Post-installation

After a multi-instance installation of the SAP BOBI platform, you need to perform additional, post-configuration steps, to support application high availability.

## Configure the cluster name

In a multi-instance deployment of the SAP BOBI platform, you want to run several CMS servers together in a cluster. A cluster consists of two or more CMS servers working together against a common CMS system database. If a node that is running on CMS fails, a node with another CMS will continue to service BI platform requests. By default in

the SAP BOBI platform, a cluster name reflects the hostname of the first CMS that you install.

To configure the cluster name on Linux, follow the instructions in the [SAP Business Intelligence Platform Administrator Guide](#). After configuring the cluster name, follow SAP Note [1660440](#) to set the default system entry on the CMC or BI launchpad sign-in page.

## Configure input and output filestore location to Azure NetApp Files

Filestore refers to the disk directories where the actual SAP BusinessObjects files are. The default location of file repository server for the SAP BOBI platform is located in the local installation directory. In a multi-instance deployment, it's important to set up the filestore on a shared storage, such as Azure NetApp Files. This allows access to the filestore from all storage tier servers.

1. If you haven't already created NFS volumes, create them in Azure NetApp Files.  
(Follow the instructions in the earlier section "Provision Azure NetApp Files.")
2. Mount the NFS volume. (Follow the instructions in the earlier section "Mount the Azure NetApp Files volume.")
3. Follow SAP Note [2512660](#) to change the path of file repository (both input and output).

## Session replication in Tomcat clustering

Tomcat supports clustering two or more application servers for session replication and failover. SAP BOBI platform sessions are serialized, so a user session can fail over seamlessly to another instance of Tomcat, even when an application server fails.

For example, suppose a user is connected to a web server that fails while the user is navigating a folder hierarchy in a SAP BI application. With a correctly configured cluster, the user can continue navigating the folder hierarchy without being redirected to the sign-in page.

See SAP Note [2808640](#) for steps to configure Tomcat clustering by using multicast. Note that Azure, however, doesn't support multicast. So to make the Tomcat cluster work in Azure, you must use [StaticMembershipInterceptor](#) (SAP Note [2764907](#)). For more information, see the blog post [Tomcat Clustering using Static Membership for SAP BusinessObjects BI Platform](#).

# Load-balancing web tier of SAP BI platform

In a SAP BOBI multi-instance deployment, Java Web Application servers (web tier) are running on two or more hosts. To distribute the user load evenly across web servers, you can use a load balancer between end users and web servers. In Azure, you can either use Azure Load Balancer or Azure Application Gateway to manage traffic to your web application servers. Details about each offering are explained in following section.

## Azure Load Balancer

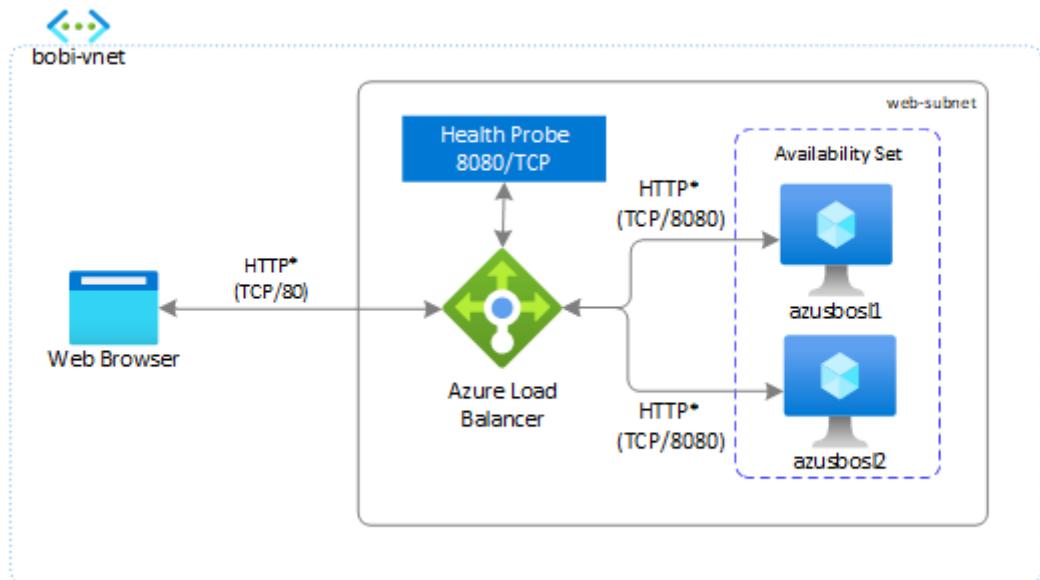
[Azure Load Balancer](#) is a high performance, low latency layer 4 (TCP, UDP) load balancer. It distributes traffic among healthy virtual machines (VMs). A load balancer health probe monitors a specified port on each VM, and only distributes traffic to an operational VM. You can either choose a public load balancer or an internal load balancer, depending on whether or not you want SAP BI platform accessible from the internet. It's zone redundant, ensuring high-availability across availability zones.

In the following diagram, refer to the Internal Load Balancer section. The web application server runs on port 8080, the default Tomcat HTTP port, which will be monitored by health probe. Any incoming request that comes from end users is redirected to the web application servers (`azusbos11` or `azusbos12`). Load Balancer doesn't support TLS/SSL termination (also known as TLS/SSL offloading). If you're using Load Balancer to distribute traffic across web servers, use Standard Load Balancer.

### Note

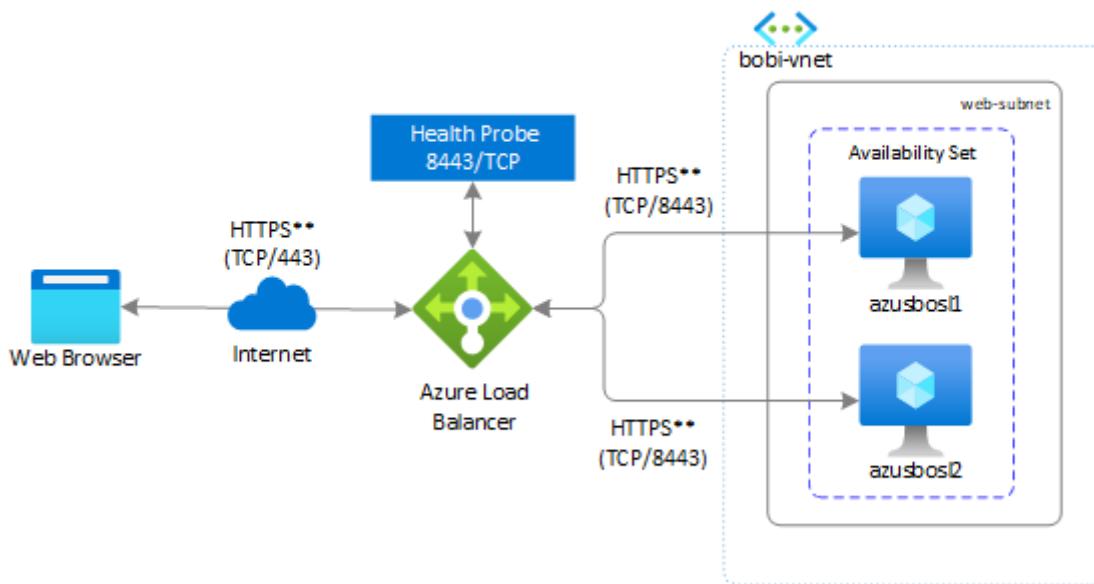
When VMs without public IP addresses are placed in the pool of internal (no public IP address) Standard Load Balancer, there will be no outbound internet connectivity, unless you perform additional configuration to allow routing to public end points. For more information, see [Public endpoint connectivity for Virtual Machines using Azure Standard Load Balancer in SAP high-availability scenarios](#).

## Internal Load Balancer



\* If you are using HTTPS port in internal load balancer, you have to activate HTTPS port for your web application.

## Public Load Balancer



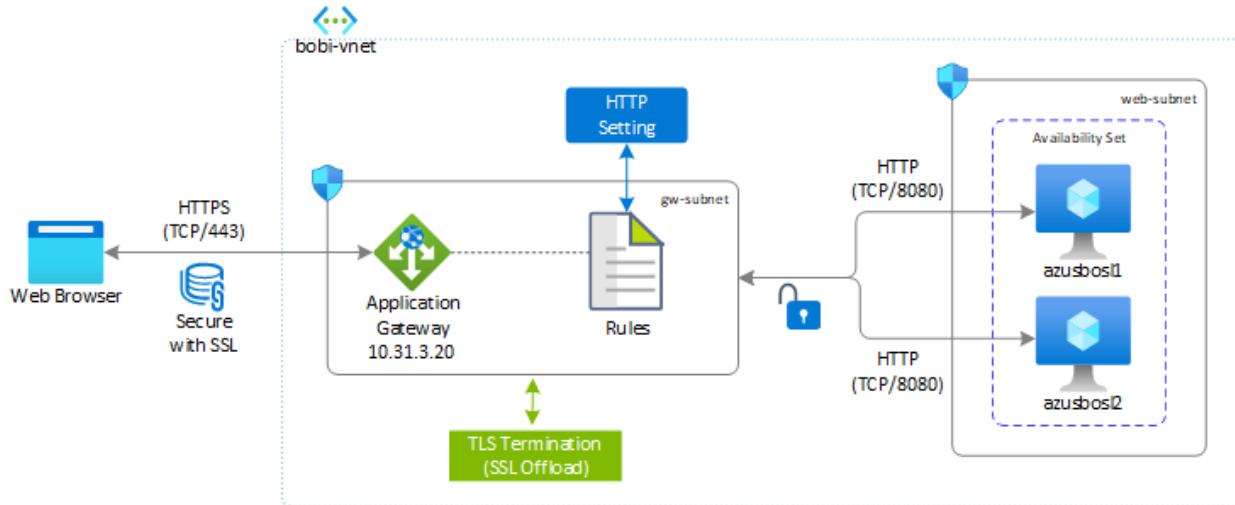
\*\* It is not advisable to use HTTP port when you are using Public Load Balancer

## Azure Application Gateway

Azure Application Gateway provides Application Delivery Controller (ADC) as a service. This service is used to help the application to direct user traffic to one or more web application servers. It offers various layer 7 load-balancing capabilities, such as TLS/SSL offloading, web application firewall (WAF), and cookie-based session affinity.

In SAP BI platform, Application Gateway directs application web traffic to the specified resources, either `azusbos1` or `azusbos2`. You assign a listener to a port, create rules, and add resources to a pool. In the following diagram, Application Gateway has a private IP

address (10.31.3.20) that acts as an entry point for users. It also handles incoming TLS/SSL (HTTPS - TCP/443) connections, decrypts the TLS/SSL, and passes on the unencrypted request (HTTP - TCP/8080) to the servers. It simplifies operations to maintain just one TLS/SSL certificate on Application Gateway.



To configure Application Gateway for a SAP BOBI web server, see the blog post [Load Balancing SAP BOBI Web Servers using Azure Application Gateway ↗](#).

#### ⓘ Note

Azure Application Gateway is preferable to load balance the traffic to a web server. It provides helpful features, such as SSL offloading, centralized SSL management to reduce encryption and decryption overhead on the server, a round-robin algorithm to distribute traffic, WAF capabilities, and high availability.

## SAP BOBI platform reliability on Azure

SAP BOBI platform includes different tiers, which are optimized for specific tasks and operations. When a component from any one tier becomes unavailable, a SAP BOBI application either becomes inaccessible or limited in its functionality. Make sure that each tier is designed to be reliable, to keep the application operational without any business disruption.

This guide explores how features native to Azure, in combination with the SAP BOBI platform configuration, improves the availability of SAP deployment. This section focuses on the following options:

- **Backup and restore:** It's a process of creating periodic copies of data and applications to a separate location. You can restore or recover to a previous state if the original data or applications are lost or damaged.

- **High availability:** A highly available platform has at least two of everything within an Azure region, to keep the application operational if one of the servers becomes unavailable.
- **Disaster recovery:** It's a process of restoring your application functionality if there's any catastrophic loss, such as an entire Azure region becoming unavailable because of some natural disaster.

Implementation of this solution varies based on the nature of the system setup in Azure. Tailor your backup/restore, high availability, and disaster recovery solutions according to your business requirements.

## Backup and restore

Backup and restore is an essential component of any business disaster recovery strategy. To develop a comprehensive strategy for SAP BOBI platform, identify the components that lead to system downtime or disruption in the application. In the SAP BOBI platform, backup of following components are vital to protect the application:

- SAP BOBI installation directory (Managed Premium Disks)
- File repository server (Azure NetApp Files or Azure Premium Files)
- CMS database (Azure Database for MySQL or a database on Azure Virtual Machines)

The following section describes how to implement a backup and restore strategy for each of these components.

### Backup and restore for SAP BOBI installation directory

In Azure, the simplest way to back up application servers and all the attached disks is by using [Azure Backup](#). It provides independent and isolated backups to guard against unintended destruction of the data on your VMs. Backups are stored in a recovery services vault, with built-in management of recovery points. Configuration and scaling are simple, backups are optimized, and you can easily restore when you need to.

As part of backup process, a snapshot is taken, and the data is transferred to the vault with no impact on production workloads. For more information, see [Snapshot consistency](#). You can also choose to back up a subset of the data disks in your VM, by using the selective disks backup and restore functionality. For more information, see [Azure VM Backup](#) and [FAQs - Backup Azure VMs](#).

### Backup and restore for file repository server

Based on your SAP BOBI deployment on Linux, you can use Azure NetApp Files as the filestore of your SAP BOBI platform. Choose from the following options for backup and restore based on the storage you use for filestore.

- **Azure NetApp Files:** You can create on-demand snapshots, and schedule automatic snapshots by using snapshot policies. Snapshot copies provide a point-in-time copy of your volume. For more information, see [Manage snapshots by using Azure NetApp Files](#).
- If you have created a separate NFS server, make sure you implement the backup and restore strategy for the same server.

## Backup and restore for CMS and audit databases

On Linux VMs, the CMS and audit databases can run on any of the supported databases. For more information, see the [support matrix](#). It's important that you adopt the backup and restore strategy based on the database used for the CMS and audit data store.

- Azure Database for MySQL automatically creates server backups, and stores them in user-configured, locally redundant or geo-redundant storage. Azure Database for MySQL takes backups of the data files and the transaction log. Depending on the supported maximum storage size, it either takes full and differential backups (4 TB max storage servers), or snapshot backups (up to 16 TB max storage servers). These backups allow you to restore a server at any point in time within your configured backup retention period. The default backup retention period is seven days, which you can [optionally configure](#) up to three days. All backups are encrypted by using AES 256-bit encryption. These backup files aren't user-exposed and can't be exported. These backups can only be used for restore operations in Azure Database for MySQL. You can use [mysqldump](#) to copy a database. For more information, see [Backup and restore in Azure Database for MySQL](#).
- For a database installed on an Azure virtual machine, you can use standard backup tools or [Azure Backup](#) for supported databases. You can also use supported third-party backup tools that provide an agent for backup and recovery of all SAP BOBI platform components.

## High availability

*High availability* refers to a set of technologies that can minimize IT disruptions by providing business continuity of applications and services. It does so through redundant, fault-tolerant, or failover-protected components inside the same datacenter. In our case,

the datacenters are within one Azure region. For more information, see [High-availability architecture and scenarios for SAP](#).

Based on the sizing result of the SAP BOBI platform, you need to design the landscape and determine the distribution of BI components across Azure Virtual Machines and subnets. The level of redundancy in the distributed architecture depends on the recovery time objective (RTO) and recovery point objective (RPO) that you need for your business. SAP BOBI platform includes different tiers, and components on each tier should be designed to achieve redundancy. For example:

- Redundant application servers, like BI application servers and web server.
- Unique components, like CMS database, file repository server, and Load Balancer.

The following sections describe how to achieve high availability on each component of the SAP BOBI platform.

## High availability for application servers

You can achieve high availability for application servers by employing redundancy. To do this, configure multiple instances of BI and web servers in various Azure VMs.

To reduce the impact of downtime due to [planned and unplanned events](#), it's a good idea to follow the [high availability architecture guidance](#).

For more information, see [Manage the availability of Linux virtual machines](#).

### Important

- The concepts of Azure availability zones and Azure availability sets are mutually exclusive. You can deploy a pair or multiple VMs into either a specific availability zone or an availability set, but you can't do both.
- If you planning to deploy across availability zones, it is advised to use [flexible scale set with FD=1](#) over standard availability zone deployment.

## High availability for a CMS database

If you're using Azure Database for MySQL for your CMS and audit databases, you have a locally redundant, high availability framework by default. You just need to select the region, and service inherent high availability, redundancy, and resiliency capabilities, without needing to configure any additional components. If the deployment strategy for the SAP BOBI platform is across availability zones, then you need to make sure you

achieve zone redundancy for your CMS and audit databases. For more information, see [High availability in Azure Database for MySQL](#) and [High availability for Azure SQL Database](#).

For other deployments for the CMS database, see the high availability information in the [DBMS deployment guides for SAP Workload](#).

## High availability for filestore

Filestore refers to the disk directories where contents like reports, universes, and connections are stored. It's shared across all application servers of that system. So you must make sure that it's highly available, along with other SAP BOBI platform components.

For SAP BOBI platform running on Linux, you can choose [Azure Premium Files](#) or [Azure NetApp Files](#) for file shares that are designed to be highly available and highly durable in nature. For more information, see [Redundancy for Azure Files](#).

Note that this file share service isn't available in all regions. See [Products available by region](#) to find up-to-date information. If the service isn't available in your region, you can create an NFS server from which you can share the file system to the SAP BOBI application. But you'll also need to consider its high availability.

## High availability for Load Balancer

To distribute traffic across a web server, you can either use Azure Load Balancer or Azure Application Gateway. The redundancy for either of these can be achieved based on the SKU you choose for deployment.

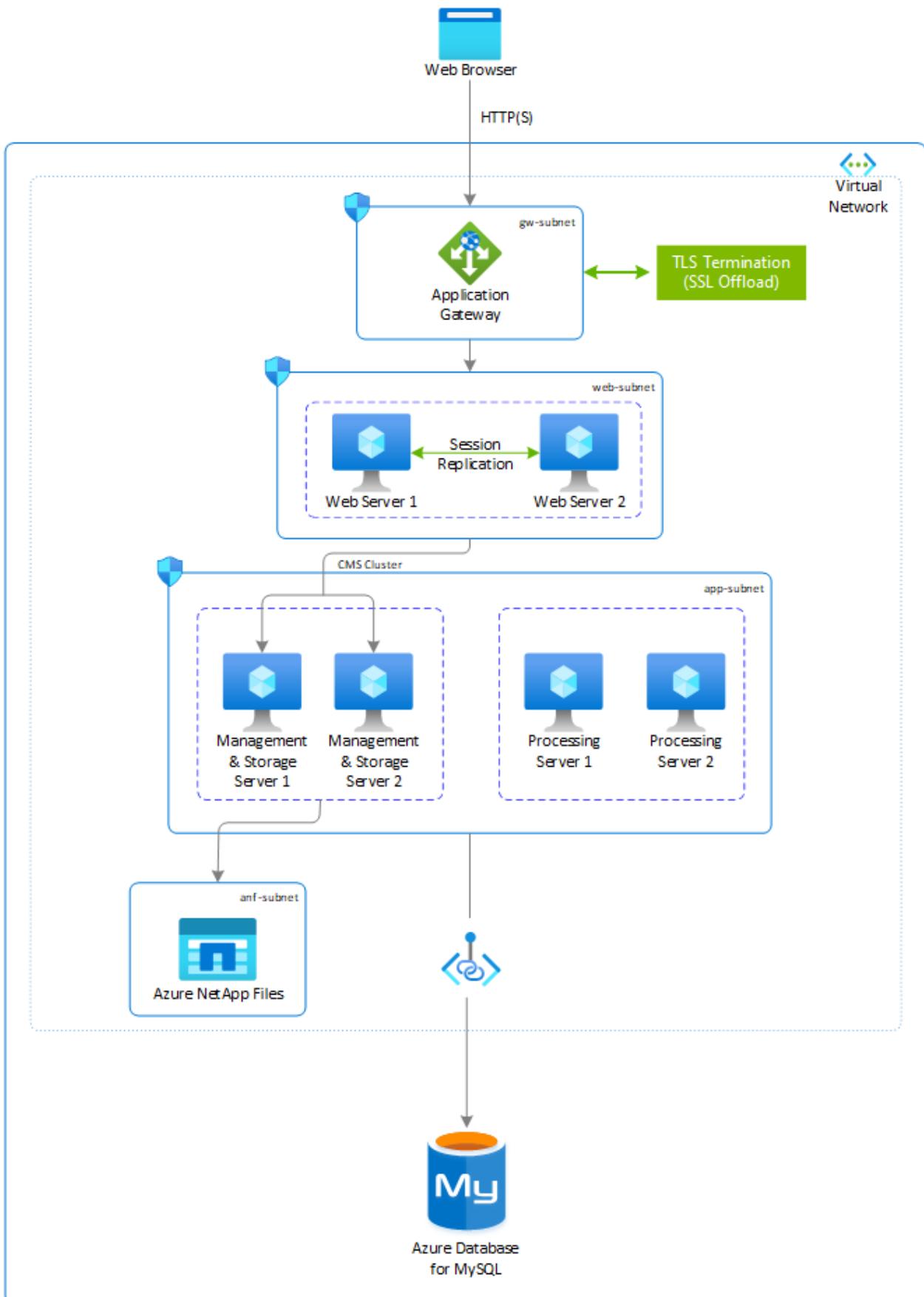
- For Azure Load Balancer, redundancy can be achieved by configuring Standard Load Balancer as zone-redundant. For more information, see [Standard Load Balancer and Availability Zones](#).
- For Application Gateway, high availability can be achieved based on the type of tier selected during deployment.
  - v1 SKU supports high-availability scenarios when you've deployed two or more instances. Azure distributes these instances across update and fault domains to ensure that instances don't all fail at the same time. You achieve redundancy within the zone.
  - v2 SKU automatically ensures that new instances are spread across fault domains and update domains. If you choose zone redundancy, the newest instances are also spread across availability zones to offer zonal failure

resiliency. For more details, see [Autoscaling and Zone-redundant Application Gateway v2](#).

## Reference high availability architecture for SAP BOBI platform

The following diagram shows the setup of SAP BOBI platform running on Linux server. The architecture showcases the use of different services, like Azure Application Gateway, Azure NetApp Files, and Azure Database for MySQL. These services offer built-in redundancy, which reduces the complexity of managing different high availability solutions.

Notice that the incoming traffic (HTTPS) is load-balanced by using Azure Application Gateway v1/v2 SKU, which is highly available when deployed on two or more instances. Multiple instances of the web server, management servers, and processing servers are deployed in separate VMs to achieve redundancy. Azure NetApp Files has built-in redundancy within the datacenter, so your Azure NetApp Files volumes for the file repository server will be highly available. The CMS database is provisioned on Azure Database for MySQL, which has inherent high availability. For more information, see [High availability in Azure Database for MySQL](#).



The preceding architecture provides insight on how a SAP BOBI deployment on Azure can be done. But it doesn't cover all possible configuration options. You can tailor your deployment based on your business requirements.

In several Azure regions, you can use availability zones. This means you can take advantage of an independent supply of power source, cooling, and network. It enables

you to deploy an application across two or three availability zones. If you want to achieve high availability across availability zones, you can deploy SAP BOBI platform across these zones, making sure that each component in the application is zone redundant.

## Disaster recovery

This section explains the strategy to provide disaster recovery protection for a SAP BOBI platform running on Linux. It complements the [Disaster Recovery for SAP](#) document, which represents the primary resources for the overall SAP disaster recovery approach. For SAP BOBI, refer to SAP Note [2056228](#), which describes the following methods to implement a disaster recovery environment safely.

- Fully or selectively using lifecycle management or federation to promote and distribute the content from the primary system.
- Periodically copying over the CMS and file repository server contents.

This guide focuses on the second option. It won't cover all possible configuration options for disaster recovery, but does cover a solution that features native Azure services in combination with a SAP BOBI platform configuration.

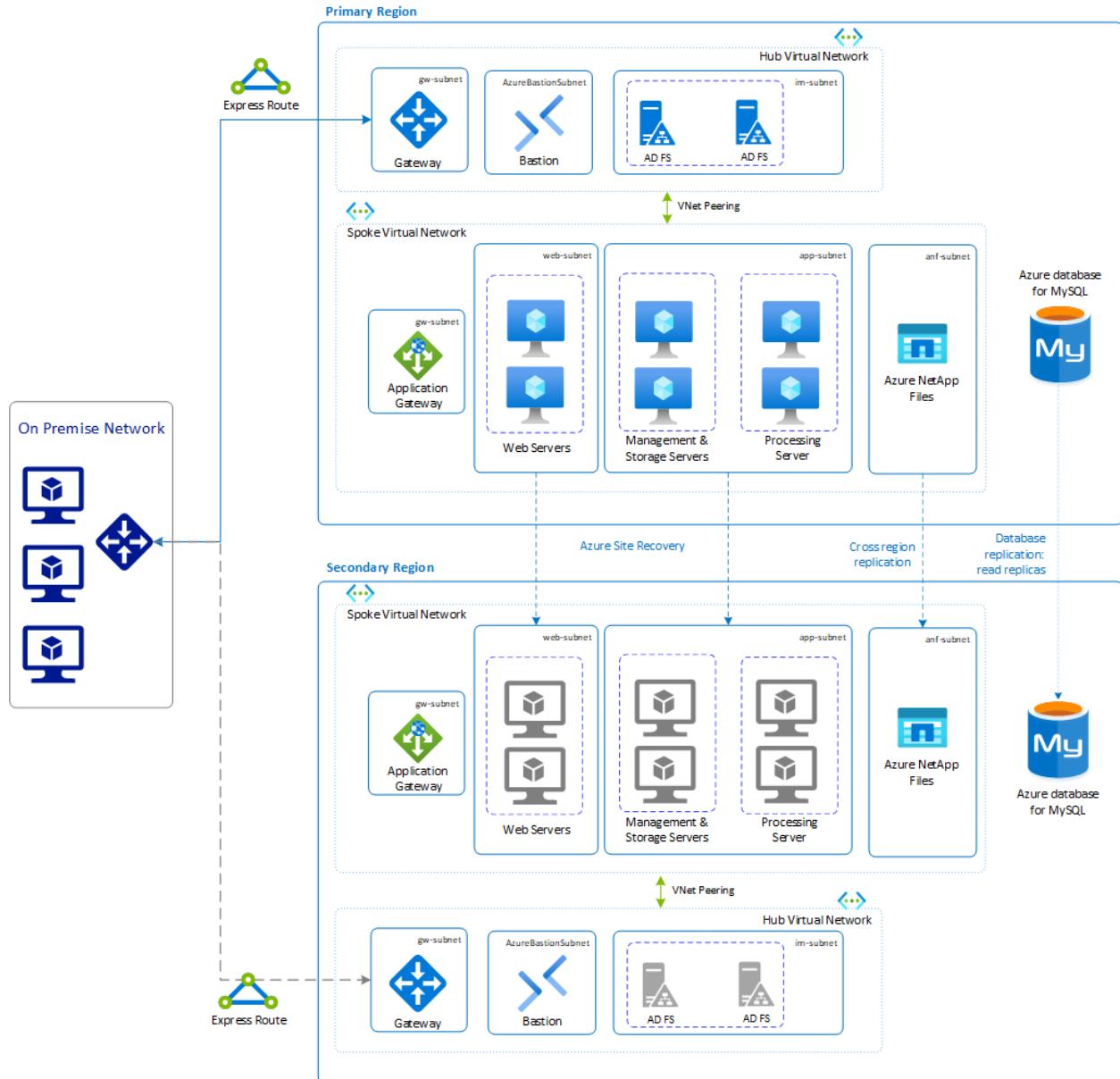
### Important

- The availability of each component in the SAP BOBI platform should be factored in the secondary region, and you must thoroughly test the entire disaster recovery strategy.
- In case where your SAP BI platform is configured with **flexible scale set** with FD=1, then you need to use **PowerShell** to set up Azure Site Recovery for disaster recovery. Currently, it's the only method available to configure disaster recovery for VMs deployed in scale set.

## Reference disaster recovery architecture for SAP BOBI platform

This reference architecture is running a multi-instance deployment of the SAP BOBI platform, with redundant application servers. For disaster recovery, you should fail over all the components of the SAP BOBI platform to a secondary region. In the following diagram, Azure NetApp Files is used as the filestore, Azure Database for MySQL as the CMS and audit repository, and Azure Application Gateway as the load balancer. The

strategy to achieve disaster recovery protection for each component is different, and these differences are described in the following sections.



## Load balancer

A load balancer is used to distribute traffic across web application servers of the SAP BOBI platform. On Azure, you can either use Azure Load Balancer or Azure Application Gateway for this purpose. To achieve disaster recovery for the load balancer services, you need to implement another Azure Load Balancer or Azure Application Gateway on the secondary region. To keep the same URL after a disaster recovery failover, you need to change the entry in the DNS, pointing to the load-balancing service running on the secondary region.

## VMs running web and BI application servers

Use [Azure Site Recovery](#) to replicate VMs running web and BI application servers on the secondary region. It replicates the servers and all its attached managed disk to the secondary region, so that when disasters and outages occur you can easily fail over to your replicated environment and continue working. To start replicating all the SAP application VMs to the Azure disaster recovery datacenter, follow the guidance in [Replicate a virtual machine to Azure](#).

## File repository servers

Filestore is a disk directory where the actual files, like reports and BI documents, are stored. It's important that all the files in the filestore are in sync to a disaster recovery region. Based on the type of file share service you use for SAP BOBI platform running on Linux, the appropriate disaster recovery strategy needs to be adopted to sync the content.

- **Azure NetApp Files** provides NFS and SMB volumes, so you can use any file-based copy tool to replicate data between Azure regions. For more information on how to copy a volume in another region, see [FAQs About Azure NetApp Files](#).

You can use Azure NetApp Files cross-region replication, currently in [preview ↗](#). Only changed blocks are sent over the network in a compressed, efficient format. This minimizes the amount of data required to replicate across the regions, saving data transfer costs. It also shortens the replication time, so you can achieve a smaller RPO. For more information, see [Requirements and considerations for using cross-region replication](#).

- **Azure premium files** only support locally redundant (LRS) and zone redundant storage (ZRS). For the disaster recovery strategy, you can use [AzCopy](#) or [Azure PowerShell](#) to copy your files to another storage account in a different region. For more information, see [Disaster recovery and storage account failover](#).

### Important

SMB Protocol for Azure Files is generally available, but NFS Protocol support for Azure Files is currently in preview. For more information, see [NFS 4.1 support for Azure Files is now in preview ↗](#).

## CMS database

The CMS and audit databases in the disaster recovery region must be a copy of the databases running in primary region. Based on the database type, it's important to copy

the database to the disaster recovery region based on the RTO and RPO that your business requires.

## Azure Database for MySQL

Azure Database for MySQL provides multiple options to recover a database if there's a disaster. Choose an appropriate option that works for your business.

- Enable cross-region read replicas to enhance your business continuity and disaster recovery planning. You can replicate from the source server to up to five replicas. Read replicas are updated asynchronously by using MySQL's binary log replication technology. Replicas are new servers that you manage similar to regular servers in Azure Database for MySQL. For more information, see [Read replicas in Azure Database for MySQL](#).
- Use the geo-restore feature to restore the server by using geo-redundant backups. These backups are accessible even when the region on which your server is hosted is offline. You can restore from these backups to any other region, and bring your server back online.

### ⓘ Note

Geo-restore is only possible if you provisioned the server with geo-redundant backup storage. Changing the **Backup Redundancy Options** after server creation isn't supported. For more information, see [Backup redundancy](#).

The following table shows the recommendation for disaster recovery of each tier used in this example.

SAP BOBI platform tiers	Recommendation
Azure Application Gateway	Parallel setup of Application Gateway on a secondary region.
Web application servers	Replicate by using Azure Site Recovery.
BI application servers	Replicate by using Site Recovery.
Azure NetApp Files	File-based copy tool to replicate data to a secondary region, or by using cross-region replication.

SAP BOBI platform tiers	Recommendation
Azure Database for MySQL	Cross-region read replicas, or restore backup from geo-redundant backups.

## Next steps

- Set up disaster recovery for a multi-tier SAP app deployment
- Azure Virtual Machines planning and implementation for SAP
- Azure Virtual Machines deployment for SAP
- Azure Virtual Machines DBMS deployment for SAP

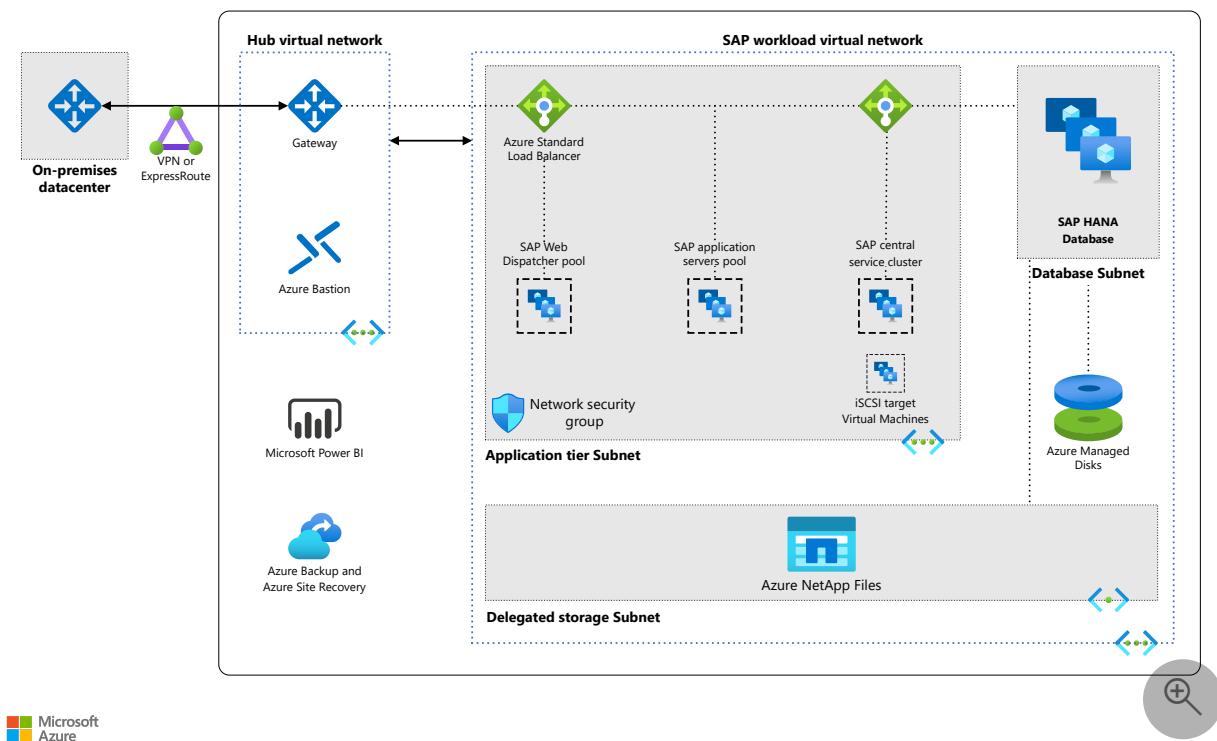
# Run SAP BW/4HANA with Linux virtual machines on Azure

Azure Bastion   Azure Managed Disks   Azure Virtual Machines   Azure Virtual Network

SAP HANA on Azure Large Instances

The following example focuses specifically on the SAP BW/4HANA application tier. It's suitable for a small-scale production environment of SAP BW/4HANA on Azure, where high availability is a priority.

## Architecture



Download a [Visio file](#) of this architecture.

## Components

This architecture makes use of the following technologies:

- [Azure Virtual Network](#) (VNet) securely connects Azure resources to each other and to an on-premises environment. In this architecture, multiple VNets are [peered together](#).

- [Linux virtual machines](#) are used for the application tier, including:
  - The SAP BusinessObjects (BOBJ) server pool.
  - The SAP Web Dispatcher pool.
  - The application servers pool.
  - The SAP Central Services cluster.
- [Load balancers](#) direct traffic to virtual machines in the application subnet. For high availability, this example uses [SAP Web Dispatcher](#) and [Azure Standard Load Balancer](#). These two services also support capacity extension by scaling out, or you can use Azure Application Gateway or other partner products, depending on the traffic type and required functionality you need, such as Secure Sockets Layer (SSL) termination and forwarding.
- [Network security groups](#) (NSGs) attach to a subnet or to the network interface cards (NICs) on a virtual machine. NSGs are used to restrict incoming, outgoing, and intra-subnet traffic in the virtual network.
- [Azure Bastion](#) provides secure access through the Azure portal to virtual machines that run in Azure, without using a jumpbox and its associated public IP address. This mechanism limits internet-facing exposure.
- [Azure Managed Disks](#). Premium or Ultra storage disks are recommended. These storage types provide data persistence for virtual machines with the SAP workload.
- [Azure NetApp Files](#) supports shared storage when using a cluster. It also supports shared storage when you need high-performance storage that can host SAP HANA data and log files. Azure NetApp Files is fully managed and scalable enough to meet the demands of most applications. It gives bare-metal performance, submillisecond latency, and integrated data management for your complex enterprise workloads on:
  - SAP HANA.
  - High-performance computing.
  - LOB applications.
  - High-performance file shares.
  - Virtual desktop infrastructure.
- [Power BI](#) enables users to access and visualize SAP BW/4HANA data from their Windows desktop. Installation requires the [SAP BW Connector](#) (implementation 2.0).

Microsoft Power BI Desktop imports data from various SAP sources, such as SAP BW/4HANA, for analysis and visualization. Power BI also complements SAP

BusinessObjects Universe by offering a business context or a semantics layer over the raw information.

- [Azure Backup](#) is an SAP Backint-certified data protection solution for SAP HANA in single-instance and scale-up deployments. Azure Backup also protects Azure Virtual Machines with general workloads.
- [Azure Site Recovery](#) is recommended as part of an automated disaster recovery solution for a multitier SAP NetWeaver application deployment. The [support matrix](#) details the capabilities and restrictions of this solution.

## Alternatives

- To help protect SAP global host files for SAP Central Services and the SAP transport directory, you can deploy [Network File System](#) (NFS) servers in a failover cluster configuration.
- [SIOS Protection Suite](#), available in Azure Marketplace, can be used to protect the global host files for Central Services instead of NFS or Azure NetApp Files.
- [Azure Application Gateway](#) is a web traffic load balancer. In one service, it provides SSL termination, a Web Application Firewall (WAF) service, and other handy high-availability and scalability features. Some SAP deployments have used it as a [gateway for the SAP Fiori front end](#) in their production landscape.

## Scenario details

SAP BW/4HANA is an enterprise data warehouse solution designed for the cloud and optimized for the SAP HANA platform. The following example focuses specifically on the SAP BW/4HANA application tier. It's suitable for a small-scale production environment of SAP BW/4HANA on Azure, where high availability is a priority.

This example workload also draws on the foundation of a pair of SAP on Azure reference architectures: [SAP NetWeaver \(Windows\) for AnyDB on virtual machines](#) and [SAP S/4HANA for Linux virtual machines on Azure](#). A similar deployment approach is used for SAP BW/4HANA workloads. The application layer is deployed using virtual machines that can be changed in size to accommodate your organization's needs.

The network layout has been simplified to demonstrate recommended architectural principles for an Azure enterprise deployment based on a [hub-spoke topology](#).

### Note

Many deployment considerations apply when deploying SAP workloads on Azure.

For more ideas and further information, see the [SAP on Azure planning and deployment checklist](#).

For details about the data persistence layer, see:

- [Run SAP HANA on Azure \(Large Instances\)](#)
- [Run SAP HANA on Linux virtual machines](#)

## Potential use cases

This scenario is relevant to the following use cases:

- Deployment of the SAP application layer separate from the DBMS layer
- Disaster recovery (DR) scenarios
- Deployments of the SAP application tier

## Recommendations

This architecture is designed for high availability, scalability, and resilience. For the best results on Azure, consider the recommendations in this section. Also, many of the recommendations for running SAP S/4HANA on Azure also apply to SAP BW/4HANA deployments. For details about SAP S/4HANA on Azure, see the [reference architecture](#).

## Virtual machines

For details about SAP support for Azure virtual machine types and throughput metrics (SAPS), see [SAP Note 1928533](#), "SAP Applications on Azure: Supported Products and Azure Virtual Machine Types." (To access this and other SAP notes, an SAP Service Marketplace account is required.)

For information about whether a virtual machine type has been certified for scale-out deployments of SAP HANA, see the "Scale-out" column in the [SAP HANA Hardware Directory](#).

## Application servers pool

In application servers pool, you can adjust the number of virtual machines based on your requirements. [Azure is certified](#) to run SAP BW/4HANA on Red Hat Enterprise Linux

and SUSE Linux Enterprise.

To manage logon groups for ABAP application servers, it's common to use the SMLG transaction to load-balance different groups, such as:

- Logon users.
- SM61 for batch server groups.
- RZ12 for RFC groups.

These transactions use the load-balancing capability within the message server of Central Services to distribute incoming sessions or workload among SAP application servers pool for SAP GUIs and RFC traffic.

## SAP Central Services cluster

This example shows a highly available cluster that uses Azure NetApp Files as a shared file storage solution. High availability for the Central Services cluster requires shared storage. Azure NetApp Files provides a simple highly available option so you don't have to deploy a Linux cluster infrastructure. An alternative is to set up a highly available [NFS service](#).

You can also deploy Central Services to a single virtual machine with Premium-managed disks and get a 99.9-percent availability [SLA ↗](#).

The virtual machines used for the application servers support multiple IP addresses per NIC. This feature supports the SAP recommended practice of using virtual host names for installations as outlined in [SAP Note 962955 ↗](#). Virtual host names decouple the SAP services from the physical host names and make it easier to migrate services from one physical host to another. This principle also applies to cloud virtual machines.

Application servers are connected to the highly available Central Services on Azure through the virtual host names of the Central Services or ERS services. These host names are assigned to the cluster front-end IP configuration of the load balancer. A load balancer supports many front-end IPs. Both the Central Services and ERS virtual IPs (VIPs) can be bound to one load balancer.

## Multi-SID installation

Azure also supports high availability in a [multi-SID installation](#) of the Linux and Windows clusters that host Central Services (ASCS/SCS). For details about deploying to a Pacemaker cluster, see the Azure multi-SID documentation for:

- [Windows](#).

- Red Hat Linux.
- SUSE Linux.

## Proximity placement groups

This example architecture also uses a [proximity placement group](#) to reduce network latency between virtual machines. This type of group places a location constraint on virtual machine deployments and minimizes the physical distance between them. The group's placement varies as follows:

- In a single SID installation, you should place all Central Services and application servers in the proximity placement group anchored by the SAP HANA database.
- In a multi-SID installation, you have the freedom to associate the Central Services and application servers with any single proximity placement group that's anchored by SAP HANA containers of different SIDs.

## Database

SAP BW/4HANA is designed for the SAP HANA database platform. Azure provides three scalability and deployment options:

- In a [scale-up SAP HANA deployment](#), the database tier uses two or more Linux virtual machines in a cluster to achieve high availability.
- A [scale-out deployment of SAP HANA](#) is supported for some virtual machine types.

## Storage

This example uses [Premium managed disks](#) for the non-shared storage of the application servers. It also uses [Azure NetApp Files](#) for cluster shared storage.

[Azure Premium SSD v2](#) is designed for performance-critical workloads like SAP. See [Deploy a Premium SSD v2](#) for information about the storage solution's benefits and current limitations.

[Ultra Disk Storage](#) significantly reduces disk latency. As a result, it benefits performance-critical applications like the SAP database servers. To compare block storage options in Azure, see [Azure managed disk types](#).

Standard managed disks aren't supported, as stated in [SAP Note 1928533](#). The use of standard storage isn't recommended for any SAP installations.

For the backup data store, we recommend using Azure [cool and archive access tiers](#). These storage tiers are cost-effective ways to store long-lived data that is infrequently accessed.

## Networking

Although not required, a [hub-spoke topology](#) is commonly deployed to provide logical isolation and security boundaries for an SAP landscape. For other networking details, see the [SAP S/4HANA reference architecture](#).

The hub VNet acts as a central point of connectivity to an on-premises network. The spokes are VNets that [peer](#) with the hub, and they can be used to isolate workloads. Traffic flows between the on-premises datacenter and the hub through a gateway connection.

Most customer implementations include one or more ExpressRoute circuits connecting on-premises networks to Azure. For less network bandwidth demand, VPN is a lower-cost alternative.

## Considerations

These considerations implement the pillars of the Azure Well-Architected Framework, which is a set of guiding tenets that can be used to improve the quality of a workload. For more information, see [Microsoft Azure Well-Architected Framework](#).

## Performance efficiency

Performance efficiency is the ability of your workload to scale to meet the demands placed on it by users in an efficient manner. For more information, see [Performance efficiency pillar overview](#).

SAP BW/4HANA is designed for real-time data warehousing tasks. SAP application servers carry on constant communications with the database servers, so minimizing latency from the application virtual machines to the database contributes to better application performance. Disk caching and server placement are two strategies that help reduce latency between these two components.

For performance-critical applications running on any database platforms, including SAP HANA, use [Premium managed disks](#) and enable [Write Accelerator](#) for the log volume. Write Accelerator is available for M-series virtual machines and improves write latency. However, when available, use [Ultra managed disks](#) in place of Premium disks without

Write Accelerator. Ultra disk capabilities continue to evolve. To see if these disks meet your requirements, review the latest information about the service scope of [ultra disks](#). Do this review especially if your implementation includes Azure resiliency features such as availability sets, Availability Zones, and cross-region replication.

To help performance by reducing the physical distance between the applications and database, use a proximity placement group, as mentioned earlier. [Scripts and utilities](#) are available on GitHub.

To optimize inter-server communications, use [Accelerated Networking](#), which is available for supported virtual machines, including D/DSv2, D/DSv3, E/ESv3, F/FS, FSv2, and Ms/Mms. In all SAP implementations, Accelerated Networking is required—especially when Azure NetApp Files is used.

To achieve high IO per second and disk bandwidth throughput, the common practices in storage volume [performance optimization](#) apply to Azure storage layout. For example, combining multiple disks together to create a striped disk volume improves IO performance. Enabling the read cache on storage content that changes infrequently enhances the speed of data retrieval.

## Scalability

This example architecture describes a small, production-level deployment with the flexibility to scale based on your requirements.

At the SAP application layer, Azure offers a wide range of virtual machine sizes for scaling up and scaling out. For an inclusive list, see [SAP Note 1928533](#). As we continue to certify more virtual machines types, you can scale up or down in the same cloud deployment.

## Availability

Resource redundancy is the general theme in highly available infrastructure solutions. If your organization has a less stringent SLA, use single-instance virtual machines with Premium disks, which offer an [uptime SLA](#).

To maximize application availability, you can deploy redundant resources in an availability set or across [Availability Zones](#). For more information, see the [SAP S/4HANA reference architecture](#).

This architecture places virtual machines that do the same role into an availability set. This configuration helps meet [SLAs](#) by guarding against downtime caused by Azure

infrastructure maintenance and unplanned outages. Two or more virtual machines per availability set are required to get a higher SLA.

## Azure Load Balancer

[Azure Load Balancer](#) is a network transmission layer service (layer 4). In cluster configurations, Azure Load Balancer directs traffic to the primary service instance or the healthy node if there's a fault. We recommend using [Azure Standard Load Balancer](#) for all SAP scenarios. It offers by-design security implementation and blocks outgoing traffic from the back-end pool unless you enable [outbound connectivity to public endpoints](#). In addition, you can also use an [Azure NAT Gateway](#) to get outbound connectivity.

Also, if you decide to deploy SAP workloads in [Azure Availability Zones](#), the Standard Load Balancer is zone-aware.

## Web Dispatcher

In this sample design, the SAP Web Dispatcher is used simply as an HTTP(s) load-balancing mechanism, for SAP traffic among the SAP application servers. To achieve [high availability](#) for the Web Dispatcher component, Azure Load Balancer implements either the failover cluster or the parallel Web Dispatcher setup. See [SAP Web Dispatcher](#) in the SAP documentation.

As a software load balancer, Web Dispatcher offers extra layer services that can do SSL termination and other offloading functions. These layer services are known as *layer 7* in the ISO networking model.

No other load balancer is needed for traffic from SAP GUI clients that connect an SAP server via DIAG protocol or Remote Function Calls (RFC). The Central Services message server balances the load through [logon groups](#) in the SAP application server.

The Web Dispatcher component is used as a load balancer for SAP traffic among the SAP application servers. To achieve [high availability of the SAP Web Dispatcher](#), Azure Load Balancer implements either the failover cluster or the parallel Web Dispatcher setup.

For internet-facing communications, a stand-alone solution in DMZ would be the recommended architecture to satisfy security concerns.

[Embedded Web Dispatcher](#) on ASCS is a special option, and proper sizing because of extra workload on ASCS should be taken into account.

## Central Services

To protect the [availability of SAP Central Services](#) (ASCS) on Azure Linux virtual machines, you must use the appropriate high availability extension (HAE) for your selected Linux distribution. HAE delivers Linux clustering software and OS-specific integration components for implementation.

To avoid a cluster split-brain problem, you can set up cluster node fencing using an iSCSI STONITH Block Device (SBD), as this example shows. Or you can instead use the [Azure Fence Agent](#). The improved Azure Fence Agent provides much faster service failover compared to the previous version of the agent for Red Hat and SUSE environments.

## Other application servers in the application servers tier

To achieve high availability for the SAP primary application servers and other application servers, load-balance traffic within the pool of application servers.

## Disaster recovery

Azure supports various [disaster recovery options](#) depending on your requirements. SAP application servers don't contain business data, so you can create SAP application servers in a secondary region before shutting them down. SAP application server software updates and configuration changes should be replicated to the disaster recovery side either manually or on a schedule. You can build a virtual machine in the disaster recovery region to run the Central Services role, which also doesn't persist business data. For details, see the [SAP S/4HANA reference architecture](#).

## Monitoring

To maximize the availability and performance of applications and services, use [Azure Monitor](#), which includes Azure Log Analytics and Azure Application Insights and provides sophisticated tools for collecting and analyzing telemetry. It can help you maximize the performance and availability of your cloud and on-premises resources and applications. You can use Azure Monitor to monitor infrastructure and application anomalies, send alerts to administrators, and automate reactions to predefined conditions.

For SAP applications that run on SAP HANA and other major database solutions, see [Azure Monitor for SAP solutions](#) to learn how Azure Monitor for SAP can help you manage the availability and performance of SAP services. Azure Monitor for SAP

provides a comprehensive initial set of metrics and telemetry for monitoring. The metric definitions are stored as SQL queries in JSON and can be modified to meet your requirements. The starting set of metrics is available on GitHub [here](#).

## Backup

For the SAP ASCS and application servers, we recommend using Azure Backup to protect the virtual machine contents. Azure Backup provides independent, isolated backups to help guard against accidental destruction of original data. Backups are stored in a [Recovery Services vault](#) that offers built-in management of recovery points. Configuration and scalability are simple, backups are optimized, and you can easily restore as needed.

Backup of the database tier varies depending on whether SAP HANA is deployed on [virtual machines](#) or [Azure Large Instances](#). See the [management and operations considerations](#) for SAP HANA on Linux virtual machines.

## Security

Security provides assurances against deliberate attacks and the abuse of your valuable data and systems. For more information, see [Overview of the security pillar](#).

SAP has its own User Management Engine (UME) to control role-based access and authorization within the SAP application and databases. For details, see the [Security Guide SAP BW/4HANA](#).

The [SAP S/4HANA reference architecture](#) provides other infrastructure security considerations that apply to SAP BW/4HANA.

## Contributors

*This article is maintained by Microsoft. It was originally written by the following contributors.*

Principal author:

- [Ben Trinh](#) | Principal Architect

*To see non-public LinkedIn profiles, sign in to LinkedIn.*

## Next steps

Learn more about the component technologies:

- [About SAP HANA database backup in Azure VMs](#)
- [Azure managed disks](#)
- [Create and deploy virtual machines in an availability set](#)
- [High availability for SAP NetWeaver on Azure VMs](#)
- [Installation of SAP HANA on Azure virtual machines](#)
- [Linux virtual machines in Azure](#)
- [Load Balancer documentation](#)
- [Network security groups](#)
- [SAP workload configurations with Azure Availability Zones](#)
- [Set up disaster recovery for a multi-tier SAP NetWeaver app deployment](#)
- [Use Azure to host and run SAP workload scenarios](#)
- [Use the SAP Business Warehouse connector in Power BI Desktop](#)
- [What is Azure Bastion?](#)
- [What is Azure Load Balancer?](#)
- [What is Azure Virtual Network?](#)
- [What is Power BI?](#)

## Related resources

Explore related architectures:

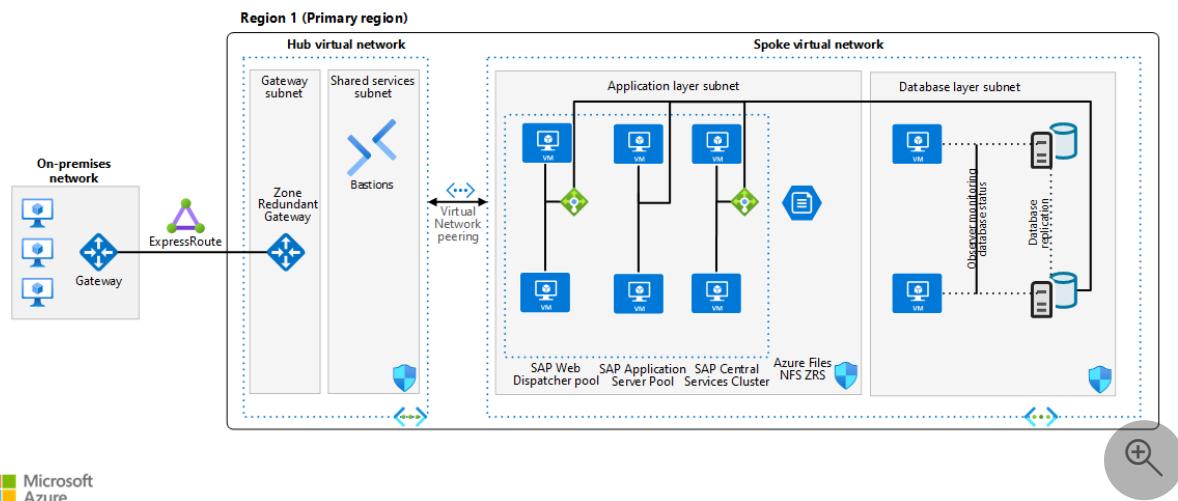
- [Run a Linux VM on Azure](#)
- [Run SAP HANA for Linux virtual machines in a scale-up architecture on Azure](#)
- [Run SAP HANA on Azure \(large instances\)](#)
- [SAP S/4HANA in Linux on Azure](#)
- [SAP S/4 HANA for large instances](#)
- [SAP on Azure Architecture Guide](#)

# SAP deployment on Azure using an Oracle database

Azure ExpressRoute    SAP HANA on Azure Large Instances    Azure Virtual Machines    Azure Virtual Network  
Azure NetApp Files

This reference architecture shows a set of proven practices for running a high-availability SAP NetWeaver with Oracle Database on Azure. The architecture principles are OS-agnostic, however, unless otherwise specified, the architecture is assumed to be based on Linux.

The first diagram shows a reference architecture for SAP on Oracle in Azure. The architecture uses availability sets.



Download a [Visio file](#) of this architecture and related architectures.

## ⓘ Note

To deploy this reference architecture, you need the appropriate licensing of SAP products and other non-Microsoft technologies.

## Components

This reference architecture describes a typical SAP production system running on Oracle Database in Azure, in a highly available deployment to maximize system availability. The architecture and its components can be customized based on business requirements.

(RTO, RPO, uptime expectations, system role) and potentially reduced to a single VM. The network layout is simplified to demonstrate the architectural principals of such SAP environment and not intended to describe a full enterprise network.

## Networking

**Virtual networks.** The [Azure Virtual Network](#) service connects Azure resources to each other with enhanced security. In this architecture, the virtual network connects to an on-premises environment via a virtual private network (VPN) gateway deployed in the hub of a [hub-spoke topology](#). SAP applications and database are contained in their own spoke virtual network. The virtual networks are subdivided into separate subnets for each tier: application (SAP NetWeaver), the database, and shared services (like Azure Bastion).

This architecture subdivides the virtual network address space into subnets. Place application servers on a separate subnet and database servers on another. Doing so allows you to secure them more easily by managing the subnet security policies rather than the individual servers and cleanly separates security rules applicable to databases from application servers.

**Virtual network peering.** This architecture uses a hub-and-spoke networking topology with multiple virtual networks that are [peered together](#). This topology provides network segmentation and isolation for services deployed on Azure. Peering enables transparent connectivity between peered virtual networks through the Microsoft backbone network. It doesn't incur a performance penalty if deployed within a single region.

**Zone-redundant gateway.** A gateway connects distinct networks, extending your on-premises network to the Azure virtual network. We recommend that you use [ExpressRoute](#) to create private connections that don't go over the public internet, but you can also use a [site-to-site](#) connection. Azure ExpressRoute or VPN gateways can be deployed across zones to guard against zone failures. See [Zone-redundant virtual network gateways](#) to understand the differences between a zonal deployment and a zone-redundant deployment. It's worth mentioning here that the IP addresses used need to be of Standard SKU for a zone deployment of the gateways.

**Network security groups.** To restrict incoming, outgoing, and intra-subnet traffic in the virtual network, create [network security groups](#) which are in turn assigned to specific subnets. DB and application subnets are secured with workload specific NSGs.

**Application security groups.** To define fine-grained network security policies inside your NSGs based on workloads that are centered on applications, use [application security](#)

[groups](#) instead of explicit IP addresses. They let you group VMs by name and help you secure applications by filtering traffic from trusted segments of your network.

**Network interface cards (NICs).** Network interface cards enable all communication among virtual machines on a virtual network. Traditional on-premises SAP deployments implement multiple NICs per machine to segregate administrative traffic from business traffic.

On Azure, the virtual network is a software-defined network that sends all traffic through the same network fabric. So it's not necessary to use multiple NICs for performance reasons. But if your organization needs to segregate traffic, you can deploy multiple NICs per VM and connect each NIC to a different subnet. You can then use network security groups to enforce different access control policies on each subnet.

Azure NICs support multiple IPs. This support conforms with the SAP recommended practice of using virtual host names for installations. For a complete outline, see [SAP note 962955](#). (To access SAP notes, you need an SAP Service Marketplace account.)

## Virtual machines

This architecture uses virtual machines (VM). For SAP application tier, VMs are deployed for all instance roles - web dispatcher and application servers - both central services SAP (A)SCS and ERS as well as application servers (PAS, AAS). Adjust the number of virtual machines based on your requirements. The [Azure Virtual Machines planning and implementation guide](#) includes details about running SAP NetWeaver on virtual machines.

Similarly for all Oracle purposes virtual machines are used, both for the Oracle Database as well as Oracle observer VMs. Observer VMs in this architecture are smaller compared to actual database servers.

- **Constrained vCPU VMs.** In order to potentially save cost on Oracle licensing, consider utilizing [vCPU constrained VMs](#)
- **Certified VM families for SAP.** For details about SAP support for Azure virtual machine types and throughput metrics (SAPS), see [SAP note 1928533](#). (To access SAP notes, you need an SAP Service Marketplace account.)

**Proximity Placement Groups (PPG).** To optimize network latency, you can use [proximity placement groups](#), which favor collocation, meaning that virtual machines are in the same datacenter to minimize application latency. They can greatly improve the user experience for most SAP applications. Due to potential restrictions with PPGs, adding the database AvSet to the SAP system's PPG should be done sparsely and only when

required for latency between SAP application and database traffic. For more details on the usage scenarios for PPGs see the linked documentation.

**Generation 2 (Gen2) virtual machines.** Azure offers the choice when deploying VMs if they should be generation 1 or 2. [Generation 2 VMs](#) support key features which are not available for generation 1 VMs. Particularly for very large Oracle databases this is of importance since some VM families such as [Mv2](#) or [MdsV2](#) are **only** supported as Gen2 VMs. Similarly, SAP on Azure certification for some newer VMs might require them to be only Gen2 for full support, even if Azure allows both on them. See details in [SAP Note 1928533 - SAP Applications on Microsoft Azure: Supported Products and Azure VM types ↗](#).

Since all other VMs supporting SAP allow the choice of either Gen2 only or Gen1+2 selectively, it is recommended to deploy all SAP VMs as Gen2, even if the memory requirements are very low. Even the smallest VMs once deployed as Gen2 can be scaled up to the largest available with a simple deallocate and resize. Gen1 VMs can only be resized to VM families allowed to run Gen1 VMs.

## Storage

This architecture uses [Azure managed disks](#) for virtual machines and [Azure Files NFS](#) or [Azure NetApp Files](#) for any NFS shared storage requirements such as sapmnt and SAP transport NFS volumes. Guidelines for storage deployment with SAP on Azure are in detail within the [Azure Storage types for SAP workload guide](#)

- **Certified storage for SAP.** Similar to certified VM types for SAP usage, see the details in [SAP note 2015553 ↗](#) and [SAP note 2039619 ↗](#).
- **Storage design for SAP on Oracle.** You can find a recommended storage design for SAP on Oracle in Azure in [Azure Virtual Machines Oracle DBMS deployment for SAP workload](#). This article provides specific guidance on file system layout, disk sizing recommendations, and other storage options.
- **Storing Oracle Database files.** On Linux ext4 or xfs filesystems need to be used for database, NTFS for Windows deployments. [Oracle ASM](#) is also supported for Oracle deployments for Oracle Database 12c Release 2 and higher.
- **Alternatives to managed disks.** You can alternatively use [Azure NetApp Files](#) for the Oracle database. For more information, see [SAP note 2039619 ↗](#) and the [Oracle on Azure](#) documentation. [Azure Files NFS](#) volumes are not intended for storing Oracle Database files, unlike Azure NetApp Files.
- **Azure Premium SSD v2** is designed for performance-critical workloads like SAP. See [Deploy a Premium SSD v2](#) for this storage solution's benefits and its current limitations.

# High availability

The preceding architecture depicts a highly available deployment, with each application layer contained on two or more virtual machines. The following components are used.

On Azure, SAP workload deployment can be either regional or zonal, depending on the availability and resiliency requirements of the SAP applications. Azure provides [different deployment options](#), like Virtual Machine Scale Sets with flexible orchestration (FD=1), availability zones, and availability sets to increase the availability of resources. To get a comprehensive understanding of the deployment options and their applicability across different Azure regions (including across zones, within a single zone, or in a region without zones), see [High-availability architecture and scenarios for SAP NetWeaver](#).

**Load balancers.** [Azure Load Balancer](#) is used to distribute traffic to virtual machines in the SAP subnets. When you incorporate Azure Load Balancer in a zonal deployment of SAP, be sure to select the Standard SKU load balancer. The Basic SKU balancer doesn't support zonal redundancy.

Consider the [decision factors](#) when deploying VMs between availability zones for SAP. Use of [proximity placement groups](#) with an availability zone deployment needs to be evaluated and used only for application tier VMs.

## ⓘ Note

Availability Zones support intra-region high availability, but they aren't effective for DR. The distances between zones are too short. Typical DR regions should be at least 100 miles away from the primary region.

**Oracle-specific components.** Oracle Database VMs are deployed in an availability set or in different availability zones. Each VM contains its own installation of the database software and VM-local database storage. Set up synchronous database replication through Oracle Data Guard between the databases to ensure consistency and allow low RTO & RPO service times in case of individual failures. Besides the database VMs, additional VMs with Oracle Data Guard Observer are needed for an Oracle Data Guard Fast-Start Failover setup. The Oracle observer VMs monitor the database and replication status and facilitate database failover in an automated way, without the need for any cluster manager. Database replication management can be performed then using Oracle Data Guard Broker for ease of use.

For details on Oracle Data Guard deployment see

- [SAP whitepaper - Setting up Oracle 12c Data Guard for SAP Customers ↗](#)

- Oracle Data Guard documentation on Azure

This architecture utilizes native Oracle tooling without any actual cluster setup or the need for a load balancer in the database tier. With Oracle Data Guard Fast-Start Failover and SAP configuration, the failover process is automated and SAP applications re-connect to the new primary database should a failover occur. Various 3rd party cluster solutions exist as an alternative, such as SIOS Protection Suite or Veritas InfoScale, details of which deployment can be found in respective 3rd party vendor's documentation respectively.

**Oracle RAC.** Oracle Real Application Cluster (RAC) is currently [not certified or supported by Oracle in Azure](#). However Oracle Data Guard technologies and architecture for high-availability can provide highly resilient SAP environments with protection against rack, data center, or regional interruptions of service.

**NFS tier.** For all highly available SAP deployments, a resilient NFS tier is required to be used, providing NFS volumes for SAP transport directory, sapmnt volume for SAP binaries as well as further volumes for (A)SCS and ERS instances. Options to provide an NFS tier are

- [Azure Files NFS](#) with zonal redundant storage (ZRS) - [SLES](#) and [RHEL](#) guides
- [Azure NetApp Files](#) deployment of NFS volumes - [SLES](#) and [RHEL](#) guides
- VM based NFS cluster - two additional VMs with local storage, replicated between VMs with DRBD (Distributed Replicated Block Device) - [SLES](#) and [RHEL](#) guides

**SAP central services cluster.** This reference architecture runs Central Services on discrete VMs. Central Services is a potential single point of failure (SPOF) when it's deployed to a single VM. To implement a highly available solution, cluster management software is needed which automates failover of (A)SCS and ERS instances to the respective VM. As this is tied strongly with the chosen NFS solution

Chosen cluster solution requires a mechanism to decide in case of software or infrastructure unavailability which VM should serve the respective service(s). With SAP on Azure, two options are available for Linux based implementation of STONITH - how to deal with unresponsive VM or application

- ***SUSE-Linux-only SBD (STONITH Block Device)*** - using one or three additional VMs which serve as iSCSI exports of a small block device, which is accessed regularly by the actual cluster member VMs, two (A)SCS/ERS VMs in this cluster pool. The VMs use these SBD mounts to cast votes and thus achieve quorum for cluster decisions. The architecture contained on this page does NOT contain the 1 or 3 additional SBD VMs. RedHat does not support any SBD implementations in Azure and thus this option is only available to SUSE SLES operating system.

- **Azure Fence Agent.** Without utilizing additional VMs, Azure management API is used for regular checks for VM availability.

Guides linked within the NFS tier section contain the necessary steps and design for respective cluster choice. Third party Azure certified cluster managers can be also utilized to provide high-availability of the SAP central services.

**SAP application servers pool.** Two or more application servers where high availability is achieved by load-balancing requests through SAP message server or web dispatchers. Each application server is independent and there is no network load balancing required for this pool of VMs.

**SAP web dispatcher pool.** The Web Dispatcher component is used as a load balancer for SAP traffic among the SAP application servers. To achieve [high availability of the SAP Web Dispatcher](#), Azure Load Balancer implements either the failover cluster or the parallel Web Dispatcher setup.

[Embedded Web Dispatcher](#) on (A)SCS is a special option. You should take into account proper sizing because of additional workload on (A)SCS.

For internet-facing communications, we recommend a stand-alone solution in the perimeter network (also known as *DMZ*) to satisfy security concerns.

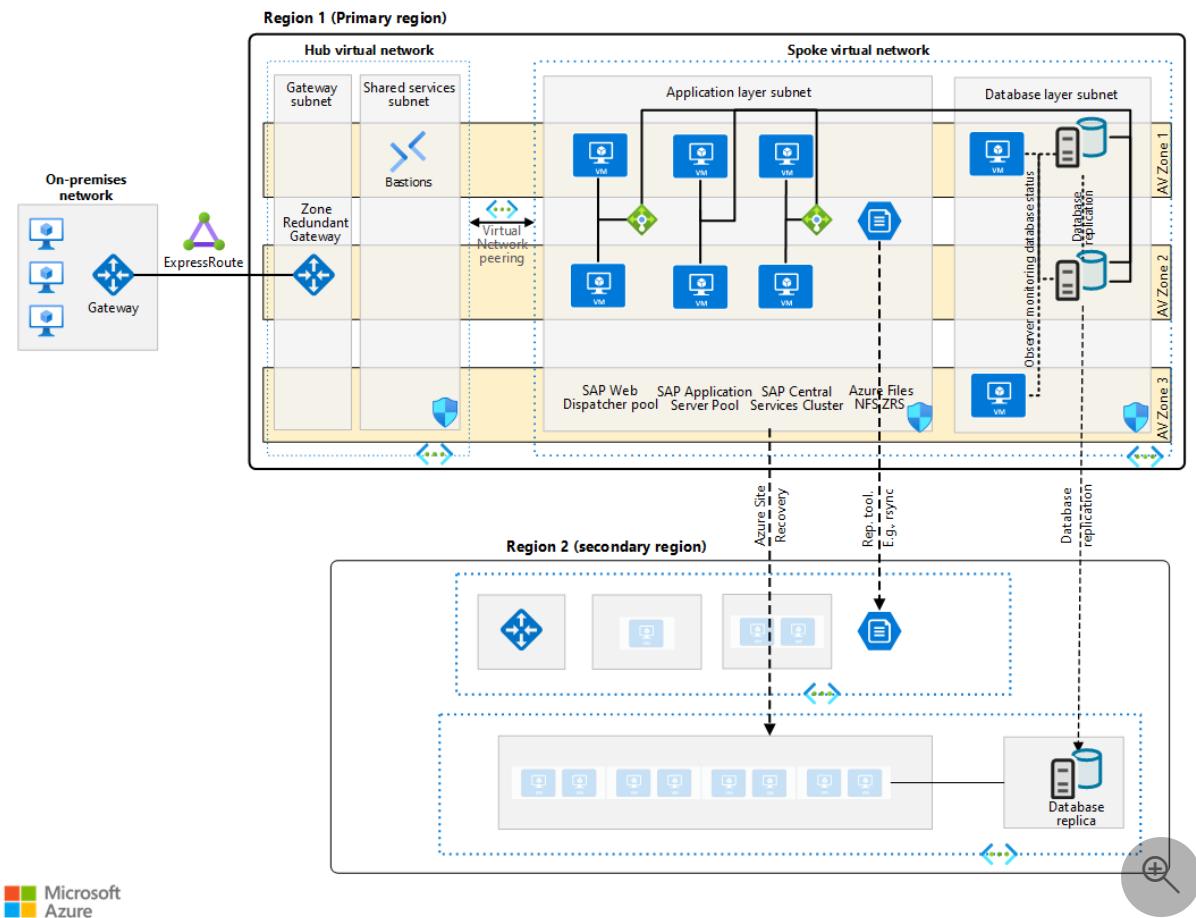
**Windows deployments.** This document, as prefaced in beginning, is focused primarily with Linux based deployments. For usage with Windows, same architectural principles apply and there are no architectural differences with Oracle between Linux and Windows.

For SAP application part, see the details in architecture guide [Run SAP NetWeaver in Windows on Azure](#).

## Considerations

### Disaster recovery

The following diagram shows the architecture of a production SAP system on Oracle in Azure. The architecture provides DR and uses availability zones.



Download a [Visio file](#) of this architecture and related architectures.

Every architectural layer in the SAP application stack uses a different approach to provide DR protection. For DR strategies and implementation details, see [Disaster recovery overview and infrastructure guidelines for SAP workload](#) and [Disaster recovery guidelines for SAP application](#).

## Backup

Backup for Oracle in Azure can be achieved through several means:

- **Azure Backup.** Azure provided and maintained scripts for Oracle Databases, to facilitate Oracle actions pre- and post backup execution.
- **Azure Storage.** Using file-based database backups, for example scheduled with SAP's BR tools, to be stored and versioned as files/directories on Azure Blob NFS, Azure Blob, or Azure Files storage services. See [documented details](#) how to achieve both Oracle data and log backups.
- **3rd party backup solutions.** See architecture of your backup storage provider, supporting Oracle in Azure.

For non-database VMs, [Azure Backup for VM](#) is recommended to protect SAP application VMs and surround infrastructure like SAP Web Dispatcher.

# Contributors

*This article is maintained by Microsoft. It was originally written by the following contributors.*

Principal author:

- [Robert Biro](#) ↗ | Senior Architect

## Next steps

- [High availability for SAP NetWeaver on Azure VMs](#)
- [Azure Virtual Machines planning and implementation for SAP NetWeaver](#)
- [Use Azure to host and run SAP workload scenarios](#)

## Communities

Communities can answer questions and help you set up a successful deployment.

Consider these resources:

- [Running SAP Applications on the Microsoft Platform blog](#) ↗
- [Azure Community Support](#) ↗
- [SAP Community](#) ↗
- [Stack Overflow for SAP](#) ↗

## Related resources

See these articles for more information and for examples of SAP workloads that use some of the same technologies:

- [SAP NetWeaver on SQL Server](#)
- [Run SAP NetWeaver in Windows on Azure](#)
- [Dev/test environments for SAP workloads on Azure](#)

# Run SAP HANA on Azure (Large Instances)

Azure ExpressRoute

Azure Virtual Machines

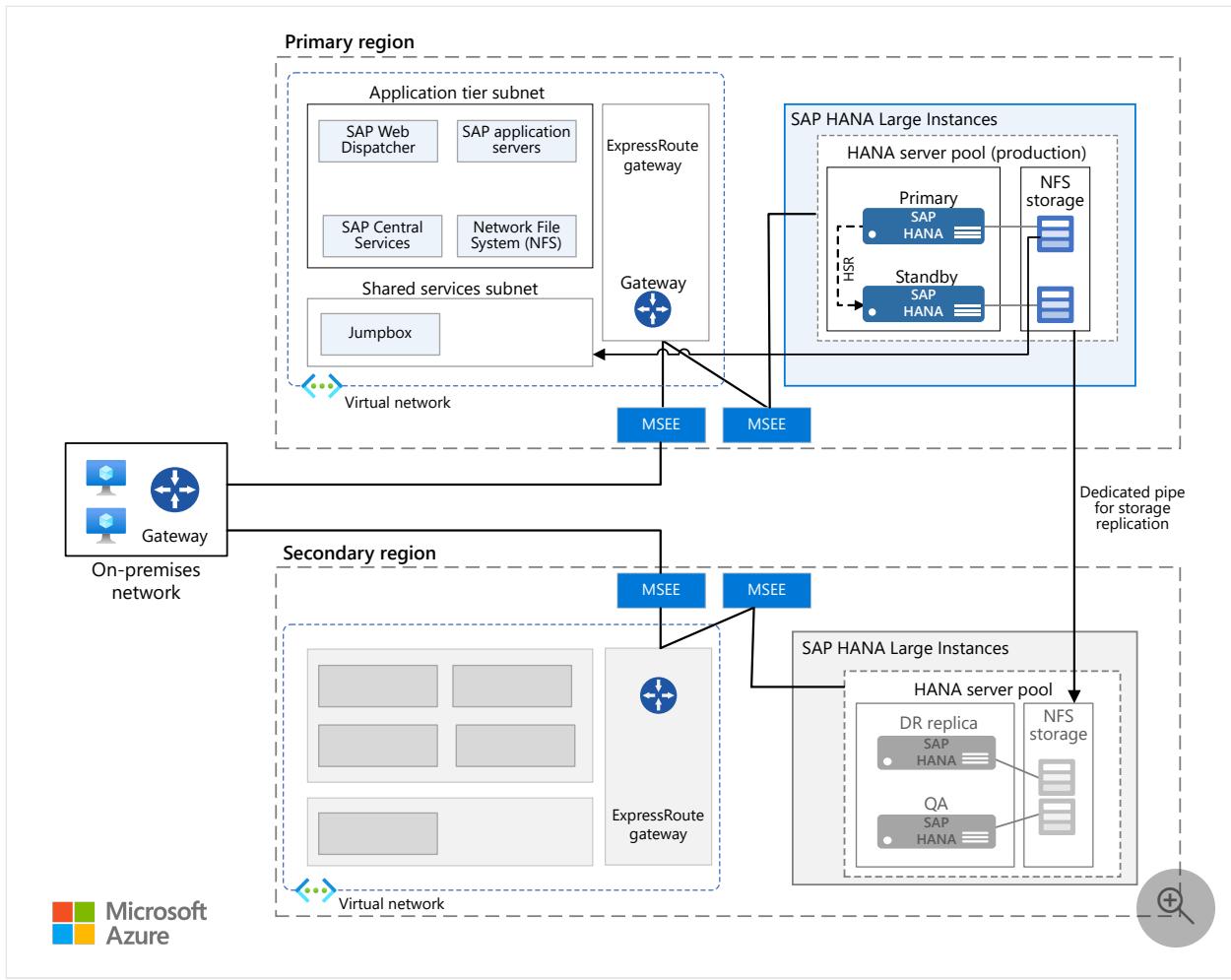
Azure Virtual Network

This reference architecture shows a set of proven practices for running SAP HANA on Azure (Large Instances) with high availability (HA) and disaster recovery (DR). Called HANA Large Instances (HLI), this offering is deployed on physical servers in Azure regions. This solution depicts a simple scale-up scenario to demonstrate core concepts in the deployment and operation of an SAP HANA system on Azure. For options, see other [installation scenarios for HANA Large Instances](#).

## ⓘ Note

Deploying this reference architecture requires appropriate licensing of SAP products and other non-Microsoft technologies.

## Architecture



[Download a Visio file](#) of this architecture.

## Workflow

This architecture consists of the following infrastructure components.

- **Virtual network.** The [Azure Virtual Network](#) (VNet) service securely connects Azure resources to each other and is subdivided into separate [subnets](#) for each layer. SAP application layers are deployed on Azure Virtual Machines (VMs) to connect to the HANA database layer residing on large instances.
- **HLI Revision 4.5 network.** As of July 2021, the updated revision of the HLI Rev 4 is available. This updated implementation [rev4.5] includes many improvements in the infrastructure, such as 100Gb/s networking for the NFS storage and better network redundancy of the DB server. In this design, the HLI servers are deployed in Azure datacenters in close physical proximity to the Azure VMs where the SAP application servers are running. When used in conjunction with an [ExpressRoute FastPath][fastpath] configuration, Rev 4.5 elevates application performance. These networking features also support the Rev 3 and Rev 4 deployments.

- **Virtual machines (VMs).** VMs are used in the SAP application layer and shared services layer. The latter includes a jump box used by administrators to set up HANA Large Instances and to provide access to other VMs. To colocate the SAP application servers in the same datacenter with the HANA Large Instance units, use [proximity placement groups](#).
- **HANA Large Instance.** This [physical server](#) is certified to meet SAP HANA Tailored Datacenter Integration (TDI) standards for running SAP HANA. This architecture uses two HANA Large Instances: a primary and a secondary compute unit. High availability at the data layer is provided through HANA System Replication (HSR).
- **High Availability Pair.** A group of HANA Large Instances blades are managed together to provide database redundancy and reliability.
- **Microsoft Enterprise Edge (MSEE).** MSEE is a connection point from a connectivity provider or your network edge through an ExpressRoute circuit.
- **Network interface cards (NICs).** To enable communication, the HANA Large Instance server provides four virtual NICs by default. This architecture requires one NIC for client communication, a second NIC for the node-to-node connectivity needed by HSR, a third NIC for HANA Large Instance storage, and a fourth for iSCSI used in high availability clustering.
- **Network File System (NFS) storage.** The [NFS](#) server supports the network file share that provides secure data persistence for HANA Large Instance.
- **ExpressRoute.** [ExpressRoute](#) is the recommended Azure networking service for creating private connections between an on-premises network and Azure VNets that do not go over the public internet. Azure VMs connect to HANA Large Instances using another ExpressRoute connection. The ExpressRoute connection between the Azure VNet and the HANA Large Instances is set up as part of the Microsoft offering.
- **Gateway.** The ExpressRoute Gateway is used to connect the Azure VNet used for the SAP application layer to the HANA Large Instance network. Use the [High Performance or Ultra Performance](#) SKU.
- **Disaster recovery (DR).** Options for DR include HANA System Replication (HSR), HANA file backup and restore, or storage replication. Upon request, the Microsoft Service Management team can configure the storage servers and volumes. You are responsible for scheduling the storage snapshot, testing the system, and getting familiar with the recovery process. Other considerations apply to the application tier for [SAP NetWeaver](#) and [SAP S/4HANA](#) on Azure.

# Recommendations

Requirements can vary, so use these recommendations as a starting point.

## HANA Large Instances compute

[HANA Large Instances](#) are physical servers based on the Intel Broadwell and Cascade Lake CPU architecture and configured in a large instance stamp—that is, a specific set of servers or blades. A compute unit equals one server or blade, and a stamp is made up of multiple servers or blades. Within a large instance stamp, servers are not shared and are dedicated to running one customer's deployment of SAP HANA.

A variety of [SKUs are available for HANA Large Instances](#), supporting up to 24 TB single instance (120 TB scale-out) of memory for BW/4HANA or other SAP HANA workloads.

Choose a SKU that fulfills the sizing requirements you determined in your architecture and design sessions. Always ensure that your sizing applies to the correct SKU.

Capabilities and deployment requirements [vary by type](#), and availability varies by region. You can also step up from one SKU to a larger SKU.

Microsoft helps establish the large instance setup, but it is your responsibility to verify the operating system's configuration settings. Make sure to review the most current SAP Notes for your exact Linux release.

## Storage

Storage layout is implemented according to the recommendation of the TDI for SAP HANA. HANA Large Instances come with a specific storage configuration for the standard TDI specifications. However, you can purchase additional storage in 1 TB increments.

To support the requirements of mission-critical environments including fast recovery, NFS is used and not direct attached storage. The NFS storage server for HANA Large Instances is hosted in a multi-tenant environment, where tenants are segregated and secured using compute, network, and storage isolation.

To support high availability at the primary site, use different storage layouts. For example, in a multi-host scale-out, the storage is shared. Another high availability option is application-based replication, such as HSR. For DR, however, a snapshot-based storage replication is used.

## Networking

This architecture uses both virtual and physical networks. The virtual network is part of Azure infrastructure as a service (IaaS) and connects to a discrete HANA Large Instances physical network through [ExpressRoute](#) circuits. A cross-premises gateway connects your workloads in the Azure VNet to your on-premises sites.

All HANA Large Instance deployments since July 2019 use Rev 4 stamps, which are deployed in close proximity to the Azure VM hosts used for the SAP application servers. As a result, Rev 4 deployment minimizes network latency between the application and database layers.

HANA Large Instances networks are isolated from each other for security. Instances residing in different regions do not communicate with each other, except for the dedicated storage replication. However, to use HSR, inter-region communications are required. [Azure Global Reach][globalreach], [IP routing tables](#), or proxies can be used to enable cross-regions HSR.

All Azure VNets that connect to HANA Large Instances in one region can be [cross-connected](#) via ExpressRoute to HANA Large Instances in a secondary region.

The ExpressRoute circuit for HANA Large Instances is included by default during provisioning. For setup, a specific network layout is needed, including required Classless Inter-Domain Routing (CIDR) address ranges and domain routing. For details, see [SAP HANA \(large instances\) infrastructure and connectivity on Azure](#).

To lower network latency and improve performance, consider enabling FastPath (also referred to as MSEE v2). This network configuration allows traffic from on-premises to the Azure VNet, and from the VNet to HANA Large Instances, to bypass the Azure gateway.

## Considerations

### Scalability

To scale up or down, you can choose from many sizes of servers that are available for HANA Large Instances. They are categorized as [Type I](#) and [Type II](#) and tailored for different workloads. Choose a size that can grow with your workload for the next three years. One-year commitments are also available.

A multi-host, scale-out deployment is generally used for BW/4HANA deployments as a kind of database partitioning strategy. As of this writing, BW/4HANA on HANA Large Instances can scale out to 120 TB. To scale out, plan the placement of HANA tables prior to installation. From an infrastructure standpoint, multiple hosts are connected to a

shared storage volume, enabling quick takeover by standby hosts in case one of the compute worker nodes in the HANA system fails.

S/4HANA and SAP Business Suite on HANA on a single blade can scale up to 24 TB with a single-instance node. HANA Large Instances and the Azure storage infrastructure also support S/4HANA and BW/4HANA scale-out deployments. For specific SKUs that are certified for scale-out, please consult the [SAP certified hardware directory][directory].

Memory requirements for HANA increase as data volume grows. Use your system's current memory consumption as the basis for predicting future consumption, and then map your demand into one of the HANA Large Instances sizes.

If you already have SAP deployments, SAP provides reports you can use to check the data used by existing systems and calculate memory requirements for a HANA instance. For example, see the following SAP Notes (access requires an SAP Service Marketplace account):

- SAP Note [1793345](#) - Sizing for SAP Suite on HANA
- SAP Note [1872170](#) - Suite on HANA and S/4 HANA sizing report
- SAP Note [2121330](#) - FAQ: SAP BW on HANA Sizing Report
- SAP Note [1736976](#) - Sizing Report for BW on HANA
- SAP Note [2296290](#) - New Sizing Report for BW on HANA

## Availability

Resource redundancy is the general theme in highly available infrastructure solutions. Work with SAP, your system integrator, or Microsoft to properly architect and implement a [high availability and disaster-recovery](#) strategy. This architecture follows the Azure [service level agreement](#) (SLA) for HANA on Azure (Large Instances). To assess your availability requirements, consider any single points of failure, the desired level of uptime for services, and these common metrics:

- Recovery Time Objective (RTO) means the duration of time in which the HANA Large Instances server is unavailable.
- Recovery Point Objective (RPO) means the maximum tolerable period in which customer data might be lost due to a failure.

For high availability, deploy more than one instance in a HA pair and use HSR in a synchronous mode to minimize data loss and downtime. In addition to a local, two-node high availability setup, HSR supports multi-tier replication, where a third node in a separate Azure region registers to the secondary replica of the clustered HSR pair as its replication target. This forms a replication daisy chain.

The failover to the DR node is a manual process without Linux clustering. For automatic fault detection and failover, you can configure Pacemaker to further lower downtime caused by software or hardware failure. Beginning with HANA 2.0 SPS 04, HSR also supports multi-target replication. Instead of a daisy chain, this form of replication has one primary and multiple secondary subscribers.

When you set up HANA Large Instances HSR with automatic failover, you can request the Microsoft Service Management team to set up a [STONITH device](#) for your HANA Large Instances servers.

## Disaster recovery

This architecture supports [disaster recovery](#) between HANA Large Instances in different Azure regions. There are two ways to support DR with HANA Large Instances:

- Storage replication. The primary storage contents are constantly replicated to the remote DR storage systems that are available on the designated DR HANA Large Instances server. In storage replication, the HANA database is not loaded into memory. This DR option is simpler from an administration perspective. To determine if this is a suitable strategy, consider the database load time against the availability SLA. Storage replication also enables you to perform point-in-time recovery. If multi-purpose (cost-optimized) DR is set up, you must purchase additional storage of the same size at the DR location. Microsoft provides self-services [storage snapshot and failover scripts](#) for HANA failover as part of the HANA Large Instances offering.
- Multi-tier or multi-target HSR with a third replica in the DR region (where the HANA database is loaded onto memory). This option supports a faster recovery time but does not support a point-in-time recovery. HSR requires a secondary system. HANA system replication traffic destined for the DR site can be routed through proxies such as nginx or IP tables. Alternatively, Global Reach can be used to link the ExpressRoute circuits together, enabling permitted users to connect to HANA Large Instances unit directly.

## Cost optimization

Use the [Azure pricing calculator](#) to estimate costs.

For more information, see the cost section in [Microsoft Azure Well-Architected Framework](#).

SKUs can affect the billing model. Here are some cost considerations.

## Virtual machines

In this reference architecture, virtual machines are used for hosting SAP applications, SAP services, and shared services such as management jump boxes. There are certain certified SKUs of HANA Large Instances. The configurations depend on the workload, CPU resources, desired memory, and budget.

[HANA Large Instances SKUs](#) are available as reserved VM instances. [Azure Reservations](#) can lower your cost if you can commit to one-year or three-year term. VM reservations can reduce costs up to 72 percent when compared to pay-as-you-go prices. You get a purpose-built SAP HANA infrastructure with compute, storage, and network. HANA Large Instances is coupled with NFS storage and networking and provides built-in support for backups through storage snapshots, high availability and disaster recovery and scale-out configurations. If your workload doesn't have a predictable time of completion or resource consumption, consider the pay-as-you-go option.

Use [Azure savings plan overview](#) and combine with Azure Reservations. Azure savings plan is a flexible cost-saving plan that generates significant savings off pay-as-you-go prices. You agree to a one-year or three-year contract and receive discounts on eligible compute services. Savings apply to these compute services regardless of the region, instance size, or operating system. For more information, see [Azure savings plan documentation](#).

Use [Azure Spot VMs](#) to run workloads that can be interrupted and do not require completion within a predetermined time frame or an SLA.

For more information, see the "SAP HANA on Azure Large Instances" section in [HLI for SAP HANA Virtual Machines Pricing](#).

## Azure ExpressRoute

For this architecture, Azure ExpressRoute is used as the networking service for creating private connections between an on-premises network and Azure virtual networks. Azure VMs connect to HANA Large Instances using another ExpressRoute connection and an ExpressRoute Gateway. [High Performance or Ultra Performance](#) is the recommended SKU.

All inbound data transfer is free. All outbound data transfer is charged based on a pre-determined rate. For more information, see [Azure ExpressRoute pricing](#).

### Note

You can optimize this reference architecture for cost by running one or multiple HANA containers in one HANA Large Instances blade. This setup is suitable for nonproduction HANA workloads.

## Backup

Based on your business requirements, choose from several options available.

 Expand table

Backup option	Pros	Cons
HANA backup	Native to SAP. Built-in consistency check.	Long backup and recovery times. Storage space consumption.
HANA snapshot	Native to SAP. Rapid backup and restore.	
Storage snapshot ↗	Included with HANA Large Instances. Optimized DR for HANA Large Instances. Boot volume backup support.	Maximum 254 snapshots per volume.
Log backup	Combined with full HANA data backup, offers point in time recovery.	
Other backup tools	Redundant backup location.	Additional licensing costs.

For details about a do-it-yourself approach to backup and more options provided with HANA Large Instances, see the [Backup and restore][backup-restore] article.

## Manageability

[Monitor HANA Large Instances resources](#)—such as CPU, memory, network bandwidth, and storage space—using SAP HANA Studio, SAP HANA Cockpit, SAP Solution Manager, and other native Linux tools. HANA Large Instances [Type I SKUs](#) don't come with built-in

monitoring tools. Type II SKUs offers prebuilt diagnostic tools for system activity logging and troubleshooting.

Microsoft offers basic tools and resources to help you [monitor HANA Large Instances](#) on Azure. The Microsoft support team can also assist you in troubleshooting technical issues.

## Security

- Since the end of 2018, [HANA Large Instances storage](#) is encrypted by default.
- Data in transit between HANA Large Instances and the VMs is not encrypted. To encrypt the data transfer, enable the application-specific encryption. See SAP Note [2159014](#) - FAQ: SAP HANA Security.
- Isolation provides security between the tenants in the multi-tenant HANA Large Instance environment. Tenants are isolated using their own VLAN.
- [Azure network security best practices](#) provide helpful guidance.
- As with any deployment, [operating system hardening](#) is recommended, including hardening the SUSE Linux image for SAP on Azure.
- For physical security, access to Azure datacenters is limited to authorized personnel only. No customers can access the physical servers.

For more information, see [SAP HANA Security—An Overview](#). (An SAP Service Marketplace account is required for access.)

## Communities

Communities can answer questions and help you set up a successful deployment. Consider the following:

- [Running SAP Applications on the Microsoft Platform Blog](#)
- [Stack Overflow SAP](#)

## Contributors

*This article is maintained by Microsoft. It was originally written by the following contributors.*

Principal author:

- [Ben Trinh](#) | Principal Architect

*To see non-public LinkedIn profiles, sign in to LinkedIn.*

## Related resources

You may wish to review the following [Azure example scenarios](#) that demonstrate specific solutions using some of the same technologies:

- [Running SAP production workloads using an Oracle Database on Azure](#)
- [Dev/test environments for SAP workloads on Azure](#)

# Deploy a SAP HANA scale-out system with standby node on Azure VMs by using Azure NetApp Files on Red Hat Enterprise Linux

Article • 07/11/2023

This article describes how to deploy a highly available SAP HANA system in a scale-out configuration with standby on Azure Red Hat Enterprise Linux virtual machines (VMs), by using [Azure NetApp Files](#) for the shared storage volumes.

In the example configurations, installation commands, and so on, the HANA instance is 03 and the HANA system ID is HN1. The examples are based on HANA 2.0 SP4 and Red Hat Enterprise Linux for SAP 7.6.

## ⓘ Note

This article contains references to terms that Microsoft no longer uses. When these terms are removed from the software, we'll remove them from this article.

Before you begin, refer to the following SAP notes and papers:

- [Azure NetApp Files documentation](#)
- SAP Note [1928533](#) ↗ includes:
  - A list of Azure VM sizes that are supported for the deployment of SAP software
  - Important capacity information for Azure VM sizes
  - Supported SAP software, and operating system (OS) and database combinations
  - The required SAP kernel version for Windows and Linux on Microsoft Azure
- SAP Note [2015553](#) ↗: Lists prerequisites for SAP-supported SAP software deployments in Azure
- SAP Note [2002167] has recommended OS settings for Red Hat Enterprise Linux
- SAP Note [2009879](#) ↗ has SAP HANA Guidelines for Red Hat Enterprise Linux
- SAP Note [3108302](#) ↗ has SAP HANA Guidelines for Red Hat Enterprise Linux 9.x
- SAP Note [2178632](#) ↗: Contains detailed information about all monitoring metrics reported for SAP in Azure
- SAP Note [2191498](#) ↗: Contains the required SAP Host Agent version for Linux in Azure
- SAP Note [2243692](#) ↗: Contains information about SAP licensing on Linux in Azure

- SAP Note [1999351](#): Contains additional troubleshooting information for the Azure Enhanced Monitoring Extension for SAP
- SAP Note [1900823](#): Contains information about SAP HANA storage requirements
- [SAP Community Wiki](#): Contains all required SAP notes for Linux
- [Azure Virtual Machines planning and implementation for SAP on Linux](#)
- [Azure Virtual Machines deployment for SAP on Linux](#)
- [Azure Virtual Machines DBMS deployment for SAP on Linux](#)
- General RHEL documentation
  - [High Availability Add-On Overview](#)
  - [High Availability Add-On Administration](#)
  - [High Availability Add-On Reference](#)
  - [Red Hat Enterprise Linux Networking Guide](#)
- Azure-specific RHEL documentation:
  - [Install SAP HANA on Red Hat Enterprise Linux for Use in Microsoft Azure](#)
- [NFS v4.1 volumes on Azure NetApp Files for SAP HANA](#)

## Overview

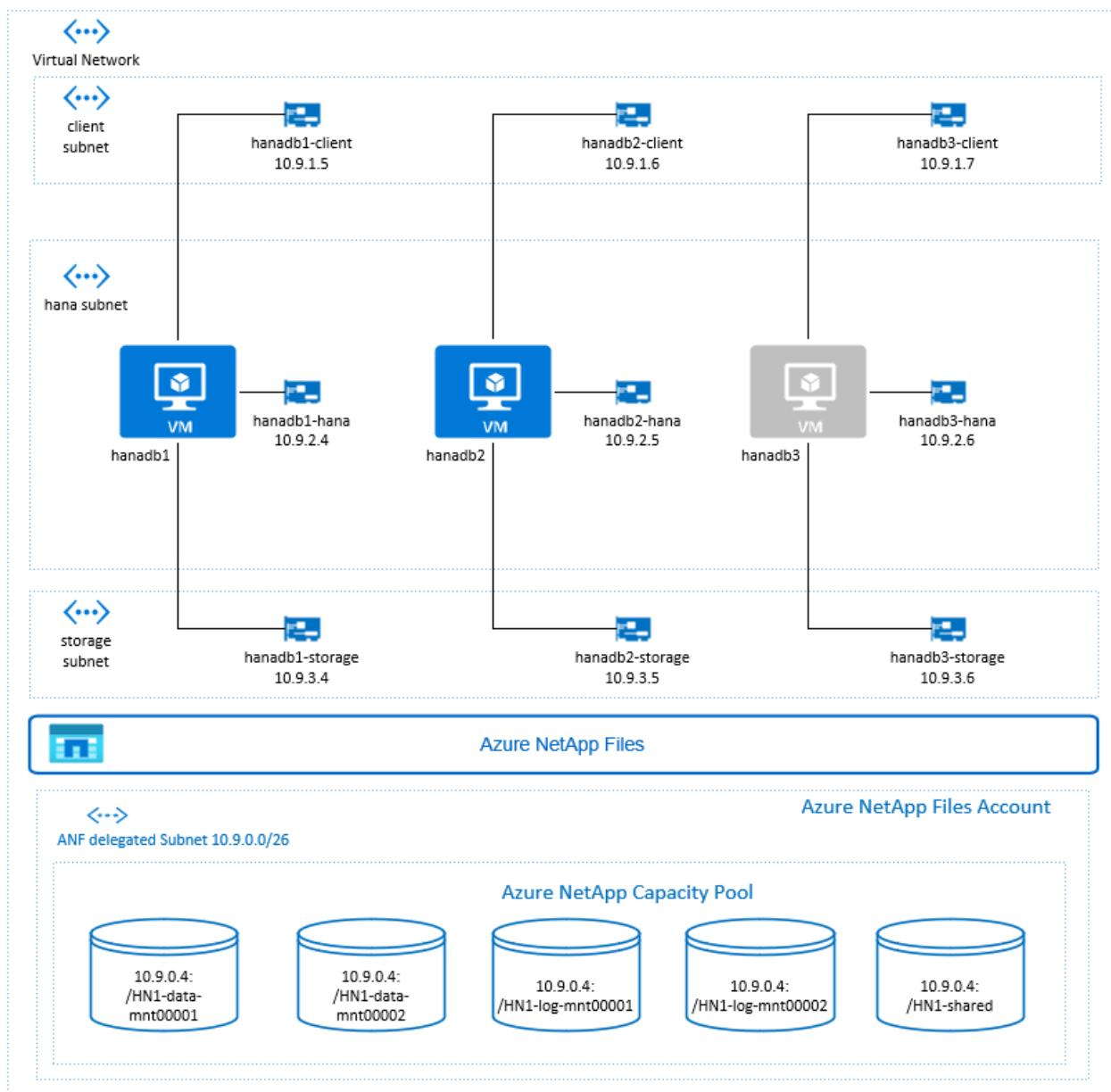
One method for achieving HANA high availability is by configuring host auto failover. To configure host auto failover, you add one or more virtual machines to the HANA system and configure them as standby nodes. When active node fails, a standby node automatically takes over. In the presented configuration with Azure virtual machines, you achieve auto failover by using [NFS on Azure NetApp Files](#).

### Note

The standby node needs access to all database volumes. The HANA volumes must be mounted as NFSv4 volumes. The improved file lease-based locking mechanism in the NFSv4 protocol is used for I/O fencing.

### Important

To build the supported configuration, you must deploy the HANA data and log volumes as NFSv4.1 volumes and mount them by using the NFSv4.1 protocol. The HANA host auto-failover configuration with standby node is not supported with NFSv3.



In the preceding diagram, which follows SAP HANA network recommendations, three subnets are represented within one Azure virtual network:

- For client communication
- For communication with the storage system
- For internal HANA inter-node communication

The Azure NetApp volumes are in separate subnet, [delegated to Azure NetApp Files](#).

For this example configuration, the subnets are:

- `client` 10.9.1.0/26
- `storage` 10.9.3.0/26
- `hana` 10.9.2.0/26
- `anf` 10.9.0.0/26 (delegated subnet to Azure NetApp Files)

# Set up the Azure NetApp Files infrastructure

Before you proceed with the setup for Azure NetApp Files infrastructure, familiarize yourself with the [Azure NetApp Files documentation](#).

Azure NetApp Files is available in several [Azure regions](#). Check to see whether your selected Azure region offers Azure NetApp Files.

For information about the availability of Azure NetApp Files by Azure region, see [Azure NetApp Files Availability by Azure Region](#).

## Important considerations

As you're creating your Azure NetApp Files volumes for SAP HANA scale-out with stand by nodes scenario, be aware of the important considerations documented in [NFS v4.1 volumes on Azure NetApp Files for SAP HANA](#).

## Sizing for HANA database on Azure NetApp Files

The throughput of an Azure NetApp Files volume is a function of the volume size and service level, as documented in [Service level for Azure NetApp Files](#).

While designing the infrastructure for SAP HANA on Azure with Azure NetApp Files, be aware of the recommendations in [NFS v4.1 volumes on Azure NetApp Files for SAP HANA](#).

The configuration in this article is presented with simple Azure NetApp Files Volumes.

### Important

For production systems, where performance is a key, we recommend to evaluate and consider using [Azure NetApp Files application volume group for SAP HANA](#).

## Deploy Azure NetApp Files resources

The following instructions assume that you've already deployed your [Azure virtual network](#). The Azure NetApp Files resources and VMs, where the Azure NetApp Files resources will be mounted, must be deployed in the same Azure virtual network or in peered Azure virtual networks.

1. Create a NetApp account in your selected Azure region by following the instructions in [Create a NetApp account](#).

2. Set up an Azure NetApp Files capacity pool by following the instructions in [Set up an Azure NetApp Files capacity pool](#).

The HANA architecture presented in this article uses a single Azure NetApp Files capacity pool at the *Ultra Service* level. For HANA workloads on Azure, we recommend using an Azure NetApp Files *Ultra* or *Premium service Level*.

3. Delegate a subnet to Azure NetApp Files, as described in the instructions in [Delegate a subnet to Azure NetApp Files](#).
4. Deploy Azure NetApp Files volumes by following the instructions in [Create an NFS volume for Azure NetApp Files](#).

As you're deploying the volumes, be sure to select the **NFSv4.1** version. Deploy the volumes in the designated Azure NetApp Files [subnet](#). The IP addresses of the Azure NetApp volumes are assigned automatically.

Keep in mind that the Azure NetApp Files resources and the Azure VMs must be in the same Azure virtual network or in peered Azure virtual networks. For example, **HN1-data-mnt00001**, **HN1-log-mnt00001**, and so on, are the volume names and `nfs://10.9.0.4/HN1-data-mnt00001`, `nfs://10.9.0.4/HN1-log-mnt00001`, and so on, are the file paths for the Azure NetApp Files volumes.

- volume **HN1-data-mnt00001** (`nfs://10.9.0.4/HN1-data-mnt00001`)
- volume **HN1-data-mnt00002** (`nfs://10.9.0.4/HN1-data-mnt00002`)
- volume **HN1-log-mnt00001** (`nfs://10.9.0.4/HN1-log-mnt00001`)
- volume **HN1-log-mnt00002** (`nfs://10.9.0.4/HN1-log-mnt00002`)
- volume **HN1-shared** (`nfs://10.9.0.4/HN1-shared`)

In this example, we used a separate Azure NetApp Files volume for each HANA data and log volume. For a more cost-optimized configuration on smaller or non-productive systems, it's possible to place all data mounts on a single volume and all logs mounts on a different single volume.

## Deploy Linux virtual machines via the Azure portal

First you need to create the Azure NetApp Files volumes. Then do the following steps:

1. Create the [Azure virtual network subnets](#) in your [Azure virtual network](#).
2. Deploy the VMs.

3. Create the additional network interfaces, and attach the network interfaces to the corresponding VMs.

Each virtual machine has three network interfaces, which correspond to the three Azure virtual network subnets (`client`, `storage` and `hana`).

For more information, see [Create a Linux virtual machine in Azure with multiple network interface cards](#).

 **Important**

For SAP HANA workloads, low latency is critical. To achieve low latency, work with your Microsoft representative to ensure that the virtual machines and the Azure NetApp Files volumes are deployed in close proximity. When you're [onboarding new SAP HANA system](#) that's using SAP HANA Azure NetApp Files, submit the necessary information.

The next instructions assume that you've already created the resource group, the Azure virtual network, and the three Azure virtual network subnets: `client`, `storage` and `hana`. When you deploy the VMs, select the client subnet, so that the client network interface is the primary interface on the VMs. You will also need to configure an explicit route to the Azure NetApp Files delegated subnet via the storage subnet gateway.

 **Important**

Make sure that the OS you select is SAP-certified for SAP HANA on the specific VM types you're using. For a list of SAP HANA certified VM types and OS releases for those types, go to the [SAP HANA certified IaaS platforms](#) site. Click into the details of the listed VM type to get the complete list of SAP HANA-supported OS releases for that type.

1. Create an availability set for SAP HANA. Make sure to set the max update domain.
2. Create three virtual machines (`hanadb1`, `hanadb2`, `hanadb3`) by doing the following steps:
  - a. Use a Red Hat Enterprise Linux image in the Azure gallery that's supported for SAP HANA. We used a RHEL-SAP-HA 7.6 image in this example.
  - b. Select the availability set that you created earlier for SAP HANA.
  - c. Select the client Azure virtual network subnet. Select [Accelerated Network](#).

When you deploy the virtual machines, the network interface name is automatically generated. In these instructions for simplicity we'll refer to the automatically generated network interfaces, which are attached to the client Azure virtual network subnet, as **hanadb1-client**, **hanadb2-client**, and **hanadb3-client**.

3. Create three network interfaces, one for each virtual machine, for the **storage** virtual network subnet (in this example, **hanadb1-storage**, **hanadb2-storage**, and **hanadb3-storage**).
4. Create three network interfaces, one for each virtual machine, for the **hana** virtual network subnet (in this example, **hanadb1-hana**, **hanadb2-hana**, and **hanadb3-hana**).
5. Attach the newly created virtual network interfaces to the corresponding virtual machines by doing the following steps:
  - a. Go to the virtual machine in the [Azure portal](#).
  - b. In the left pane, select **Virtual Machines**. Filter on the virtual machine name (for example, **hanadb1**), and then select the virtual machine.
  - c. In the **Overview** pane, select **Stop** to deallocate the virtual machine.
  - d. Select **Networking**, and then attach the network interface. In the **Attach network interface** drop-down list, select the already created network interfaces for the **storage** and **hana** subnets.
  - e. Select **Save**.
  - f. Repeat steps b through e for the remaining virtual machines (in our example, **hanadb2** and **hanadb3**).
  - g. Leave the virtual machines in stopped state for now. Next, we'll enable [accelerated networking](#) for all newly attached network interfaces.
6. Enable accelerated networking for the additional network interfaces for the **storage** and **hana** subnets by doing the following steps:
  - a. Open [Azure Cloud Shell](#) in the [Azure portal](#).
  - b. Execute the following commands to enable accelerated networking for the additional network interfaces, which are attached to the **storage** and **hana** subnets.

```

az network nic update --id /subscriptions/your
subscription/resourceGroups/your resource
group/providers/Microsoft.Network/networkInterfaces/hanadb1-storage --
accelerated-networking true
az network nic update --id /subscriptions/your
subscription/resourceGroups/your resource
group/providers/Microsoft.Network/networkInterfaces/hanadb2-storage --
accelerated-networking true
az network nic update --id /subscriptions/your
subscription/resourceGroups/your resource
group/providers/Microsoft.Network/networkInterfaces/hanadb3-storage --
accelerated-networking true

az network nic update --id /subscriptions/your
subscription/resourceGroups/your resource
group/providers/Microsoft.Network/networkInterfaces/hanadb1-hana --
accelerated-networking true
az network nic update --id /subscriptions/your
subscription/resourceGroups/your resource
group/providers/Microsoft.Network/networkInterfaces/hanadb2-hana --
accelerated-networking true
az network nic update --id /subscriptions/your
subscription/resourceGroups/your resource
group/providers/Microsoft.Network/networkInterfaces/hanadb3-hana --
accelerated-networking true

```

7. Start the virtual machines by doing the following steps:

- In the left pane, select **Virtual Machines**. Filter on the virtual machine name (for example, **hanadb1**), and then select it.
- In the **Overview** pane, select **Start**.

## Operating system configuration and preparation

The instructions in the next sections are prefixed with one of the following:

- [A]: Applicable to all nodes
- [1]: Applicable only to node 1
- [2]: Applicable only to node 2
- [3]: Applicable only to node 3

Configure and prepare your OS by doing the following steps:

1. [A] Maintain the host files on the virtual machines. Include entries for all subnets.

The following entries were added to `/etc/hosts` for this example.

```
# Storage
10.9.3.4 hanadb1-storage
10.9.3.5 hanadb2-storage
10.9.3.6 hanadb3-storage
# Client
10.9.1.5 hanadb1
10.9.1.6 hanadb2
10.9.1.7 hanadb3
# Hana
10.9.2.4 hanadb1-hana
10.9.2.5 hanadb2-hana
10.9.2.6 hanadb3-hana
```

2. [A] Add a network route, so that the communication to the Azure NetApp Files goes via the storage network interface.

In this example will use `Networkmanager` to configure the additional network route.

The following instructions assume that the storage network interface is `eth1`.

First, determine the connection name for device `eth1`. In this example the connection name for device `eth1` is `Wired connection 1`.

```
# Execute as root
nmcli connection
# Result
#NAME UUID TYPE
DEVICE
#System eth0 5fb06bd0-0bb0-7ffb-45f1-d6edd65f3e03 ethernet
eth0
#Wired connection 1 4b0789d1-6146-32eb-83a1-94d61f8d60a7 ethernet
eth1
```

Then configure additional route to the Azure NetApp Files delegated network via `eth1`.

```
# Add the following route
# ANFDelegatedSubnet/cidr via StorageSubnetGW dev
StorageNetworkInterfaceDevice
nmcli connection modify "Wired connection 1" +ipv4.routes "10.9.0.0/26
10.9.3.1"
```

Reboot the VM to activate the changes.

### 3. [A] Install the NFS client package.

```
yum install nfs-utils
```

### 4. [A] Prepare the OS for running SAP HANA on Azure NetApp with NFS, as described in SAP note [3024346 - Linux Kernel Settings for NetApp NFS](#). Create configuration file */etc/sysctl.d/91-NetApp-HANA.conf* for the NetApp configuration settings.

```
vi /etc/sysctl.d/91-NetApp-HANA.conf
# Add the following entries in the configuration file
net.core.rmem_max = 16777216
net.core.wmem_max = 16777216
net.ipv4.tcp_rmem = 4096 131072 16777216
net.ipv4.tcp_wmem = 4096 16384 16777216
net.core.netdev_max_backlog = 300000
net.ipv4.tcp_slow_start_after_idle=0
net.ipv4.tcp_no_metrics_save = 1
net.ipv4.tcp_moderate_rcvbuf = 1
net.ipv4.tcp_window_scaling = 1
net.ipv4.tcp_timestamps = 1
net.ipv4.tcp_sack = 1
```

### 5. [A] Create configuration file */etc/sysctl.d/ms-az.conf* with additional optimization settings.

```
vi /etc/sysctl.d/ms-az.conf
```

```
# Add the following entries in the configuration file
net.ipv6.conf.all.disable_ipv6 = 1
net.ipv4.tcp_max_syn_backlog = 16348
net.ipv4.conf.all.rp_filter = 0
sunrpc.tcp_slot_table_entries = 128
vm.swappiness=10
```

### 💡 Tip

Avoid setting net.ipv4.ip\_local\_port\_range and net.ipv4.ip\_local\_reserved\_ports explicitly in the sysctl configuration files to allow SAP Host Agent to manage the port ranges. For more details see SAP note [2382421](#).

5. [A] Adjust the sunrpc settings, as recommended in SAP note [3024346 - Linux Kernel Settings for NetApp NFS](#).

```
vi /etc/modprobe.d/sunrpc.conf
# Insert the following line
options sunrpc tcp_max_slot_table_entries=128
```

6. [A] Red Hat for HANA configuration.

Configure RHEL as described in SAP Note [2292690](#), [2455582](#), [2593824](#), and Red Hat note [2447641](#).

### ⚠ Note

If installing HANA 2.0 SP04 you will be required to install package `compat-sap-c++-7` as described in SAP note [2593824](#), before you can install SAP HANA.

## Mount the Azure NetApp Files volumes

1. [A] Create mount points for the HANA database volumes.

```
mkdir -p /hana/data/HN1/mnt00001
```

```
mkdir -p /hana/data/HN1/mnt0002
mkdir -p /hana/log/HN1/mnt0001
mkdir -p /hana/log/HN1/mnt0002
mkdir -p /hana/shared
mkdir -p /usr/sap/HN1
```

2. [1] Create node-specific directories for /usr/sap on **HN1**-shared.

```
# Create a temporary directory to mount HN1-shared
mkdir /mnt/tmp
# if using NFSv3 for this volume, mount with the following command
mount 10.9.0.4:/HN1-shared /mnt/tmp
# if using NFSv4.1 for this volume, mount with the following command
mount -t nfs -o sec=sys,nfsvers=4.1 10.9.0.4:/HN1-shared /mnt/tmp
cd /mnt/tmp
mkdir shared usr-sap-hanadb1 usr-sap-hanadb2 usr-sap-hanadb3
# unmount /hana/shared
cd
umount /mnt/tmp
```

3. [A] Verify the NFS domain setting. Make sure that the domain is configured as the default Azure NetApp Files domain, i.e. **defaultv4iddomain.com** and the mapping is set to **nobody**.

#### **ⓘ Important**

Make sure to set the NFS domain in **/etc/idmapd.conf** on the VM to match the default domain configuration on Azure NetApp Files:

**defaultv4iddomain.com**. If there's a mismatch between the domain configuration on the NFS client (i.e. the VM) and the NFS server, i.e. the Azure NetApp configuration, then the permissions for files on Azure NetApp volumes that are mounted on the VMs will be displayed as **nobody**.

```
sudo cat /etc/idmapd.conf
# Example
[General]
Domain = defaultv4iddomain.com
[Mapping]
Nobody-User = nobody
```

```
Nobody-Group = nobody
```

4. [A] Verify `nfs4_disable_idmapping`. It should be set to Y. To create the directory structure where `nfs4_disable_idmapping` is located, execute the mount command. You won't be able to manually create the directory under `/sys/modules`, because access is reserved for the kernel / drivers.

```
# Check nfs4_disable_idmapping
cat /sys/module/nfs/parameters/nfs4_disable_idmapping
# If you need to set nfs4_disable_idmapping to Y
mkdir /mnt/tmp
mount 10.9.0.4:/HN1-shared /mnt/tmp
umount /mnt/tmp
echo "Y" > /sys/module/nfs/parameters/nfs4_disable_idmapping
# Make the configuration permanent
echo "options nfs nfs4_disable_idmapping=Y" >>
/etc/modprobe.d/nfs.conf
```

For more details on how to change `nfs4_disable_idmapping` parameter see <https://access.redhat.com/solutions/1749883>.

5. [A] Mount the shared Azure NetApp Files volumes.

```
sudo vi /etc/fstab
# Add the following entries
10.9.0.4:/HN1-data-mnt0001 /hana/data/HN1/mnt0001 nfs
rw,nfsvers=4.1,hard,timeo=600,rsize=262144,wsize=262144,noatime,lock,_netdev,sec=sys 0 0
10.9.0.4:/HN1-data-mnt0002 /hana/data/HN1/mnt0002 nfs
rw,nfsvers=4.1,hard,timeo=600,rsize=262144,wsize=262144,noatime,lock,_netdev,sec=sys 0 0
10.9.0.4:/HN1-log-mnt0001 /hana/log/HN1/mnt0001 nfs
rw,nfsvers=4.1,hard,timeo=600,rsize=262144,wsize=262144,noatime,lock,_netdev,sec=sys 0 0
10.9.0.4:/HN1-log-mnt0002 /hana/log/HN1/mnt0002 nfs
rw,nfsvers=4.1,hard,timeo=600,rsize=262144,wsize=262144,noatime,lock,_netdev,sec=sys 0 0
10.9.0.4:/HN1-shared/shared /hana/shared nfs
rw,nfsvers=4.1,hard,timeo=600,rsize=262144,wsize=262144,noatime,lock,_netdev,sec=sys 0 0
# Mount all volumes
```

```
sudo mount -a
```

For workloads, that require higher throughput, consider using the `nconnect` mount option, as described in [NFS v4.1 volumes on Azure NetApp Files for SAP HANA](#). Check if `nconnect` is supported by Azure NetApp Files on your Linux release.

#### 6. [1] Mount the node-specific volumes on `hanadb1`.

```
sudo vi /etc/fstab
# Add the following entries
10.9.0.4:/HN1-shared/usr-sap-hanadb1 /usr/sap/HN1 nfs
rw,nfsvers=4.1,hard,timeo=600,rsize=262144,wsize=262144,noatime,lock,_n
etdev,sec=sys 0 0
# Mount the volume
sudo mount -a
```

#### 7. [2] Mount the node-specific volumes on `hanadb2`.

```
sudo vi /etc/fstab
# Add the following entries
10.9.0.4:/HN1-shared/usr-sap-hanadb2 /usr/sap/HN1 nfs
rw,nfsvers=4.1,hard,timeo=600,rsize=262144,wsize=262144,noatime,lock,_n
etdev,sec=sys 0 0
# Mount the volume
sudo mount -a
```

#### 8. [3] Mount the node-specific volumes on `hanadb3`.

```
sudo vi /etc/fstab
# Add the following entries
10.9.0.4:/HN1-shared/usr-sap-hanadb3 /usr/sap/HN1 nfs
rw,nfsvers=4.1,hard,timeo=600,rsize=262144,wsize=262144,noatime,lock,_n
etdev,sec=sys 0 0
# Mount the volume
sudo mount -a
```

## 9. [A] Verify that all HANA volumes are mounted with NFS protocol version **NFSv4**.

```
sudo nfsstat -m
# Verify that flag vers is set to 4.1
# Example from hanadb1
/hana/data/HN1/mnt00001 from 10.9.0.4:/HN1-data-mnt00001
Flags:
rw,noatime,vers=4.1,rsize=262144,wsize=262144,namlen=255,hard,proto=tcp
,timeo=600,retrans=2,sec=sys,clientaddr=10.9.3.4,local_lock=none,addr=1
0.9.0.4
/hana/log/HN1/mnt00002 from 10.9.0.4:/HN1-log-mnt00002
Flags:
rw,noatime,vers=4.1,rsize=262144,wsize=262144,namlen=255,hard,proto=tcp
,timeo=600,retrans=2,sec=sys,clientaddr=10.9.3.4,local_lock=none,addr=1
0.9.0.4
/hana/data/HN1/mnt00002 from 10.9.0.4:/HN1-data-mnt00002
Flags:
rw,noatime,vers=4.1,rsize=262144,wsize=262144,namlen=255,hard,proto=tcp
,timeo=600,retrans=2,sec=sys,clientaddr=10.9.3.4,local_lock=none,addr=1
0.9.0.4
/hana/log/HN1/mnt00001 from 10.9.0.4:/HN1-log-mnt00001
Flags:
rw,noatime,vers=4.1,rsize=262144,wsize=262144,namlen=255,hard,proto=tcp
,timeo=600,retrans=2,sec=sys,clientaddr=10.9.3.4,local_lock=none,addr=1
0.9.0.4
/usr/sap/HN1 from 10.9.0.4:/HN1-shared/usr-sap-hanadb1
Flags:
rw,noatime,vers=4.1,rsize=262144,wsize=262144,namlen=255,hard,proto=tcp
,timeo=600,retrans=2,sec=sys,clientaddr=10.9.3.4,local_lock=none,addr=1
0.9.0.4
/hana/shared from 10.9.0.4:/HN1-shared/shared
Flags:
rw,noatime,vers=4.1,rsize=262144,wsize=262144,namlen=255,hard,proto=tcp
,timeo=600,retrans=2,sec=sys,clientaddr=10.9.3.4,local_lock=none,addr=1
0.9.0.4
```

# Installation

In this example for deploying SAP HANA in scale-out configuration with standby node with Azure, we've used HANA 2.0 SP4.

## Prepare for HANA installation

1. [A] Before the HANA installation, set the root password. You can disable the root password after the installation has been completed. Execute as `root` command

```
passwd .
```

2. [1] Verify that you can log in via SSH to **hanadb2** and **hanadb3**, without being prompted for a password.

```
ssh root@hanadb2  
ssh root@hanadb3
```

3. [A] Install additional packages, which are required for HANA 2.0 SP4. For more information, see SAP Note [2593824](#).

```
yum install libgcc_s1 libstdc++6 compat-sap-c++-7 libatomic1
```

4. [2], [3] Change ownership of SAP HANA `data` and `log` directories to `hn1adm`.

```
# Execute as root  
sudo chown hn1adm:sapsys /hana/data/HN1  
sudo chown hn1adm:sapsys /hana/log/HN1
```

5. [A] Disable the firewall temporarily, so that it doesn't interfere with the HANA installation. You can re-enable it, after the HANA installation is done.

```
# Execute as root  
systemctl stop firewalld  
systemctl disable firewalld
```

## HANA installation

1. [1] Install SAP HANA by following the instructions in the [SAP HANA 2.0 Installation and Update guide](#). In this example, we install SAP HANA scale-out with master,

one worker, and one standby node.

- a. Start the **hdblcm** program from the HANA installation software directory. Use the `internal_network` parameter and pass the address space for subnet, which is used for the internal HANA inter-node communication.

```
./hdblcm --internal_network=10.9.2.0/26
```

- b. At the prompt, enter the following values:

- For **Choose an action**: enter 1 (for install)
- For **Additional components for installation**: enter 2, 3
- For installation path: press Enter (defaults to /hana/shared)
- For **Local Host Name**: press Enter to accept the default
- Under **Do you want to add hosts to the system?**: enter y
- For **comma-separated host names to add**: enter hanadb2, hanadb3
- For **Root User Name [root]**: press Enter to accept the default
- For roles for host hanadb2: enter 1 (for worker)
- For **Host Failover Group** for host hanadb2 [default]: press Enter to accept the default
- For **Storage Partition Number** for host hanadb2 [<<assign automatically>>]: press Enter to accept the default
- For **Worker Group** for host hanadb2 [default]: press Enter to accept the default
- For **Select roles** for host hanadb3: enter 2 (for standby)
- For **Host Failover Group** for host hanadb3 [default]: press Enter to accept the default
- For **Worker Group** for host hanadb3 [default]: press Enter to accept the default
- For **SAP HANA System ID**: enter HN1
- For **Instance number [00]**: enter 03
- For **Local Host Worker Group** [default]: press Enter to accept the default
- For **Select System Usage / Enter index [4]**: enter 4 (for custom)
- For **Location of Data Volumes [/hana/data/HN1]**: press Enter to accept the default
- For **Location of Log Volumes [/hana/log/HN1]**: press Enter to accept the default
- For **Restrict maximum memory allocation? [n]**: enter n

- For **Certificate Host Name For Host hanadb1** [hanadb1]: press Enter to accept the default
- For **Certificate Host Name For Host hanadb2** [hanadb2]: press Enter to accept the default
- For **Certificate Host Name For Host hanadb3** [hanadb3]: press Enter to accept the default
- For **System Administrator (hn1adm) Password**: enter the password
- For **System Database User (system) Password**: enter the system's password
- For **Confirm System Database User (system) Password**: enter system's password
- For **Restart system after machine reboot? [n]**: enter n
- For **Do you want to continue (y/n)**: validate the summary and if everything looks good, enter y

## 2. [1] Verify global.ini

Display global.ini, and ensure that the configuration for the internal SAP HANA inter-node communication is in place. Verify the **communication** section. It should have the address space for the `hana` subnet, and `listeninterface` should be set to `.internal`. Verify the **internal\_hostname\_resolution** section. It should have the IP addresses for the HANA virtual machines that belong to the `hana` subnet.

```
sudo cat /usr/sap/HN1/SYS/global/hdb/custom/config/global.ini
# Example
#global.ini last modified 2019-09-10 00:12:45.192808 by hdbnameserve
[communication]
internal_network = 10.9.2.0/26
listeninterface = .internal
[internal_hostname_resolution]
10.9.2.4 = hanadb1
10.9.2.5 = hanadb2
10.9.2.6 = hanadb3
```

## 3. [1] Add host mapping to ensure that the client IP addresses are used for client communication. Add section `public_host_resolution`, and add the corresponding IP addresses from the client subnet.

```
sudo vi /usr/sap/HN1/SYS/global/hdb/custom/config/global.ini
#Add the section
```

```
[public_hostname_resolution]
map_hanadb1 = 10.9.1.5
map_hanadb2 = 10.9.1.6
map_hanadb3 = 10.9.1.7
```

4. [1] Restart SAP HANA to activate the changes.

```
sudo -u hn1adm /usr/sap/hostctrl/exe/sapcontrol -nr 03 -function
StopSystem HDB
sudo -u hn1adm /usr/sap/hostctrl/exe/sapcontrol -nr 03 -function
StartSystem HDB
```

5. [1] Verify that the client interface will be using the IP addresses from the `client` subnet for communication.

```
# Execute as hn1adm
/usr/sap/HN1/HDB03/exe/hdbsql -u SYSTEM -p "password" -i 03 -d
SYSTEMDB 'select * from SYS.M_HOST_INFORMATION'|grep net_publicname
# Expected result
"hanadb3","net_publicname","10.9.1.7"
"hanadb2","net_publicname","10.9.1.6"
"hanadb1","net_publicname","10.9.1.5"
```

For information about how to verify the configuration, see SAP Note [2183363 - Configuration of SAP HANA internal network](#).

6. [A] Re-enable the firewall.

- Stop HANA

```
sudo -u hn1adm /usr/sap/hostctrl/exe/sapcontrol -nr 03 -
function StopSystem HDB
```

- Re-enable the firewall

```
# Execute as root
systemctl start firewalld
systemctl enable firewalld
```

- Open the necessary firewall ports

**ⓘ Important**

Create firewall rules to allow HANA inter node communication and client traffic. The required ports are listed on [TCP/IP Ports of All SAP Products](#). The following commands are just an example. In this scenario with used system number 03.

```
# Execute as root
sudo firewall-cmd --zone=public --add-port=
{30301,30303,30306,30307,30313,30315,30317,30340,30341,30342,1128,
1129,40302,40301,40307,40303,40340,50313,50314,30310,30302}/tcp --
permanent
sudo firewall-cmd --zone=public --add-port=
{30301,30303,30306,30307,30313,30315,30317,30340,30341,30342,1128,
1129,40302,40301,40307,40303,40340,50313,50314,30310,30302}/tcp
```

- Start HANA

```
sudo -u hn1adm /usr/sap/hostctrl/exe/sapcontrol -nr 03 -
function StartSystem HDB
```

7. To optimize SAP HANA for the underlying Azure NetApp Files storage, set the following SAP HANA parameters:

- `max_parallel_io_requests 128`
- `async_read_submit on`
- `async_write_submit_active on`
- `async_write_submit_blocks all`

For more information, see [I/O stack configuration for SAP HANA](#).

Starting with SAP HANA 2.0 systems, you can set the parameters in `global.ini`.

For more information, see SAP Note [1999930](#).

For SAP HANA 1.0 systems versions SPS12 and earlier, these parameters can be set during the installation, as described in SAP Note [2267798](#).

8. The storage that's used by Azure NetApp Files has a file size limitation of 16 terabytes (TB). SAP HANA is not implicitly aware of the storage limitation, and it won't automatically create a new data file when the file size limit of 16 TB is reached. As SAP HANA attempts to grow the file beyond 16 TB, that attempt will result in errors and, eventually, in an index server crash.

### Important

To prevent SAP HANA from trying to grow data files beyond the **16-TB limit** of the storage subsystem, set the following parameters in `global.ini`.

- `datavolume_striping = true`
- `datavolume_striping_size_gb = 15000` For more information, see SAP Note [2400005](#). Be aware of SAP Note [2631285](#).

## Test SAP HANA failover

1. Simulate a node crash on an SAP HANA worker node. Do the following:

- a. Before you simulate the node crash, run the following commands as `hn1adm` to capture the status of the environment:

```
# Check the landscape status
python
/usr/sap/HN1/HDB03/exe/python_support/landscapeHostConfiguration.py
| Host      | Host      | Host      | Failover | Remove | Storage      | Storage
| Failover  | Failover  | NameServer | NameServer | IndexServer | IndexServer
IndexServer | Host      | Host      | Worker    | Worker    |
|           | Active    | Status    | Status    | Status    | Config      | Actual
| Config   | Actual    | Config    | Actual    | Config    | Actual
| Config   | Actual    | Config    | Actual    |
|           |           |           |           |           | Partition  |
Partition | Group     | Group     | Role      | Role      | Role       |
| Role     | Roles     | Roles     | Groups    | Groups    | Groups    |
```

```

| ----- | ----- | ----- | ----- | ----- | ----- | -----
- | ----- | ----- | ----- | ----- | ----- | ----- |
----- | ----- | ----- | ----- | ----- |
| hanadb1 | yes | ok | | | | 1 |
1 | default | default | master 1 | master | worker | |
master | worker | worker | default | default | |
| hanadb2 | yes | ok | | | | 2 |
2 | default | default | master 2 | slave | worker | slave |
| worker | worker | default | default | |
| hanadb3 | yes | ignore | | | | 0 |
0 | default | default | master 3 | slave | standby | |
standby | standby | standby | default | - | |

# Check the instance status
sapcontrol -nr 03 -function GetSystemInstanceList
GetSystemInstanceList
OK
hostname, instanceNr, httpPort, httpsPort, startPriority, features,
dispstatus
hanadb2, 3, 50313, 50314, 0.3, HDB|HDB_WORKER, GREEN
hanadb1, 3, 50313, 50314, 0.3, HDB|HDB_WORKER, GREEN
hanadb3, 3, 50313, 50314, 0.3, HDB|HDB_STANDBY, GREEN

```

- b. To simulate a node crash, run the following command as root on the worker node, which is **hanadb2** in this case:

```
echo b > /proc/sysrq-trigger
```

- c. Monitor the system for failover completion. When the failover has been completed, capture the status, which should look like the following:

```

# Check the instance status
sapcontrol -nr 03 -function GetSystemInstanceList
GetSystemInstanceList
OK
hostname, instanceNr, httpPort, httpsPort, startPriority, features,
dispstatus
hanadb1, 3, 50313, 50314, 0.3, HDB|HDB_WORKER, GREEN
hanadb2, 3, 50313, 50314, 0.3, HDB|HDB_WORKER, GRAY
hanadb3, 3, 50313, 50314, 0.3, HDB|HDB_STANDBY, GREEN
# Check the landscape status
python
/usr/sap/HN1/HDB03/exe/python_support/landscapeHostConfiguration.py
| Host | Host | Host | Failover | Remove | Storage | Storage
| Failover | Failover | NameServer | NameServer | IndexServer |
```

IndexServer	Host	Host	Worker	Worker		
Config	Active	Status	Status	Status	Config	Actual
Config	Actual	Config	Actual	Config	Config	Actual
Config	Actual	Config	Actual			
Partition	Group	Group	Role	Role	Role	
Role	Roles	Roles	Groups	Groups	Groups	
-----	-----	-----	-----	-----	-----	-----
-	-----	-----	-----	-----	-----	-----
-----	-----	-----	-----	-----	-----	-----
1   hanadb1   yes   ok       1						
1   default   default   master 1   master   worker						
master   worker   worker   default   default						
0   hanadb2   no   info       2						
0   default   default   master 2   slave   worker						
standby   worker   standby   default   -						
2   hanadb3   yes   info       0						
2   default   default   master 3   slave   standby						
standby   worker   default   default						

### ⓘ Important

When a node experiences kernel panic, avoid delays with SAP HANA failover by setting `kernel.panic` to 20 seconds on *all* HANA virtual machines. The configuration is done in `/etc/sysctl`. Reboot the virtual machines to activate the change. If this change isn't performed, failover can take 10 or more minutes when a node is experiencing kernel panic.

## 2. Kill the name server by doing the following:

- Prior to the test, check the status of the environment by running the following commands as `hn1adm`:

```
#Landscape status
python
/usr/sap/HN1/HDB03/exe/python_support/landscapeHostConfiguration.py
| Host | Host | Host | Failover | Remove | Storage | Storage
| Failover | Failover | NameServer | NameServer | IndexServer |
IndexServer | Host | Host | Worker | Worker | 
| | Active | Status | Status | Status | Config | Actual
| Config | Actual | Config | Actual | Config | Actual
| Config | Actual | Config | Actual | 
| | | | | | Partition |
Partition | Group | Group | Role | Role | Role
| Role | Roles | Roles | Groups | Groups | 
| ----- | ----- | ----- | ----- | ----- | -----
```

```

- | ----- | ----- | ----- | ----- | ----- | -----
----- | ----- | ----- | ----- | ----- | 
| hanadb1 | yes   | ok    |       |       |       1 |
1 | default | default | master 1 | master   | worker   |
master   | worker  | worker  | default | default |
| hanadb2 | yes   | ok    |       |       |       2 |
2 | default | default | master 2 | slave    | worker   | slave
| worker  | worker  | default | default |
| hanadb3 | yes   | ignore |       |       |       0 |
0 | default | default | master 3 | slave    | standby  |
standby  | standby | standby | default | -      |
# Check the instance status
sapcontrol -nr 03 -function GetSystemInstanceList
GetSystemInstanceList
OK
hostname, instanceNr, httpPort, httpsPort, startPriority, features,
dispstatus
hanadb2, 3, 50313, 50314, 0.3, HDB|HDB_WORKER, GREEN
hanadb3, 3, 50313, 50314, 0.3, HDB|HDB_STANDBY, GREEN
hanadb1, 3, 50313, 50314, 0.3, HDB|HDB_WORKER, GREEN

```

b. Run the following commands as **hn1adm** on the active master node, which is **hanadb1** in this case:

```
hn1adm@hanadb1:/usr/sap/HN1/HDB03> HDB kill
```

The standby node **hanadb3** will take over as master node. Here is the resource state after the failover test is completed:

```

# Check the instance status
sapcontrol -nr 03 -function GetSystemInstanceList
GetSystemInstanceList
OK
hostname, instanceNr, httpPort, httpsPort, startPriority, features,
dispstatus
hanadb2, 3, 50313, 50314, 0.3, HDB|HDB_WORKER, GREEN
hanadb3, 3, 50313, 50314, 0.3, HDB|HDB_STANDBY, GREEN
hanadb1, 3, 50313, 50314, 0.3, HDB|HDB_WORKER, GRAY
# Check the landscape status
python
/usr/sap/HN1/HDB03/exe/python_support/landscapeHostConfiguration.py
| Host    | Host    | Host    | Failover | Remove | Storage   | Storage
| Failover | Failover | NameServer | NameServer | IndexServer |
IndexServer | Host    | Host    | Worker   | Worker   |

```

		Active	Status	Status	Status	Config	Actual
Config	Actual	Config	Config	Actual	Config	Config	Actual
Config	Actual	Config	Config	Actual			
Partition	Group	Group	Role	Role	Role	Role	
Role	Roles	Roles	Groups	Groups	Groups	Groups	
-----	-----	-----	-----	-----	-----	-----	-----
--	-----	-----	-----	-----	-----	-----	-----
-----	-----	-----	-----	-----	-----	-----	-----
0   hanadb1   no   info       1							
0   default   default   master 1   slave   worker							
standby   worker   standby   default   -							
1   hanadb2   yes   ok       2							
2   default   default   master 2   slave   worker   slave							
worker   worker   default   default							
1   hanadb3   yes   info       0							
1   default   default   master 3   master   standby							
master   standby   worker   default   default							

- c. Restart the HANA instance on **hanadb1** (that is, on the same virtual machine, where the name server was killed). The **hanadb1** node will rejoin the environment and will keep its standby role.

```
hn1adm@hanadb1:/usr/sap/HN1/HDB03> HDB start
```

After SAP HANA has started on **hanadb1**, expect the following status:

```
# Check the instance status
sapcontrol -nr 03 -function GetSystemInstanceList
GetSystemInstanceList
OK
hostname, instanceNr, httpPort, httpsPort, startPriority, features,
dispstatus
hanadb2, 3, 50313, 50314, 0.3, HDB|HDB_WORKER, GREEN
hanadb3, 3, 50313, 50314, 0.3, HDB|HDB_STANDBY, GREEN
hanadb1, 3, 50313, 50314, 0.3, HDB|HDB_WORKER, GREEN
# Check the landscape status
python
/usr/sap/HN1/HDB03/exe/python_support/landscapeHostConfiguration.py
| Host      | Host      | Host      | Failover | Remove | Storage      | Storage
| Failover  | Failover  | NameServer | NameServer | IndexServer | IndexServer
IndexServer | Host      | Host      | Worker    | Worker    | |
|           | Active    | Status    | Status    | Status    | Config     | Actual
| Config   | Actual   | Config    | Actual    | Config   | Config    | Actual
```

	Config	Actual	Config	Actual				Partition
Partition	Group	Group	Role	Role	Role	Role		
	Role	Roles	Roles	Groups	Groups			
	-----	-----	-----	-----	-----	-----	-----	-----
-	-----	-----	-----	-----	-----	-----	-----	-----
-----	-----	-----	-----	-----	-----	-----	-----	-----
hanadb1	no	info				1		
0	default	default	master 1	slave	worker			
standby	worker	standby	default	-				
hanadb2	yes	ok				2		
2	default	default	master 2	slave	worker		slave	
worker	worker	default	default					
hanadb3	yes	info				0		
1	default	default	master 3	master	standby			
master	standby	worker	default	default				

d. Again, kill the name server on the currently active master node (that is, on node **hanadb3**).

```
hn1adm@hanadb3:/usr/sap/HN1/HDB03> HDB kill
```

Node **hanadb1** will resume the role of master node. After the failover test has been completed, the status will look like this:

```
# Check the instance status
sapcontrol -nr 03 -function GetSystemInstanceList
GetSystemInstanceList
OK
hostname, instanceNr, httpPort, httpsPort, startPriority, features,
dispsstatus
hanadb2, 3, 50313, 50314, 0.3, HDB|HDB_WORKER, GREEN
hanadb3, 3, 50313, 50314, 0.3, HDB|HDB_STANDBY, GRAY
hanadb1, 3, 50313, 50314, 0.3, HDB|HDB_WORKER, GREEN
# Check the landscape status
python
/usr/sap/HN1/HDB03/exe/python_support/landscapeHostConfiguration.py
| Host | Host | Host | Failover | Remove | Storage | Storage
| Failover | Failover | NameServer | NameServer | IndexServer |
IndexServer | Host | Host | Worker | Worker |
| | Active | Status | Status | Status | Config | Actual
| Config | Actual | Config | Actual | Config | Actual
| Config | Actual | Config | Actual |
| | | | | | Partition |
Partition | Group | Group | Role | Role | Role
```

Role	Roles	Roles	Groups	Groups	
hanadb1	yes	ok			1
default	default	master 1	master	worker	
master	worker	worker	default	default	
hanadb2	yes	ok			2
default	default	master 2	slave	worker	slave
worker	worker	default	default		
hanadb3	no	ignore			0
default	default	master 3	slave	standby	
standby	standby	standby	default	-	

e. Start SAP HANA on hanadb3, which will be ready to serve as a standby node.

```
hn1adm@hanadb3:/usr/sap/HN1/HDB03> HDB start
```

After SAP HANA has started on hanadb3, the status looks like the following:

```
# Check the instance status
sapcontrol -nr 03 -function GetSystemInstanceList & python
/usr/sap/HN1/HDB03/exe/python_support/landscapeHostConfiguration.py
GetSystemInstanceList
OK
hostname, instanceNr, httpPort, httpsPort, startPriority, features,
dispstatus
GetSystemInstanceList
OK
hostname, instanceNr, httpPort, httpsPort, startPriority, features,
dispstatus
hanadb2, 3, 50313, 50314, 0.3, HDB|HDB_WORKER, GREEN
hanadb3, 3, 50313, 50314, 0.3, HDB|HDB_STANDBY, GREEN
hanadb1, 3, 50313, 50314, 0.3, HDB|HDB_WORKER, GREEN
# Check the landscape status
python
/usr/sap/HN1/HDB03/exe/python_support/landscapeHostConfiguration.py
| Host      | Host      | Host      | Failover | Remove | Storage      | Storage
| Failover  | Failover  | NameServer | NameServer | IndexServer | IndexServer
| Host      | Host      | Worker     | Worker     |          |          |
|          | Active    | Status     | Status     | Status     | Config      | Actual
| Config   | Actual    | Config     | Actual     | Config     | Actual
| Config   | Actual    | Config     | Actual     |          |          |
|          |          |           |           |           | Partition   |
Partition | Group     | Group     | Role      | Role      | Role      | Role
| Role     | Roles     | Roles     | Groups    | Groups    | Groups    |
```

-	-	-	-	-	-	-	-	-	-
1	hanadb1	yes	ok				1		
1	default	default	master 1	master	worker				
master	worker	worker	default	default					
2	hanadb2	yes	ok				2		
2	default	default	master 2	slave	worker				
worker	worker	default	default						
0	hanadb3	no	ignore				0		
0	default	default	master 3	slave	standby				
standby	standby	standby	default	-					

## Next steps

- [Azure Virtual Machines planning and implementation for SAP](#)
- [Azure Virtual Machines deployment for SAP](#)
- [Azure Virtual Machines DBMS deployment for SAP](#)
- [NFS v4.1 volumes on Azure NetApp Files for SAP HANA](#)
- To learn how to establish high availability and plan for disaster recovery of SAP HANA on Azure VMs, see [High Availability of SAP HANA on Azure Virtual Machines \(VMs\)](#).

# Run SAP HANA for Linux virtual machines in a scale-up architecture on Azure

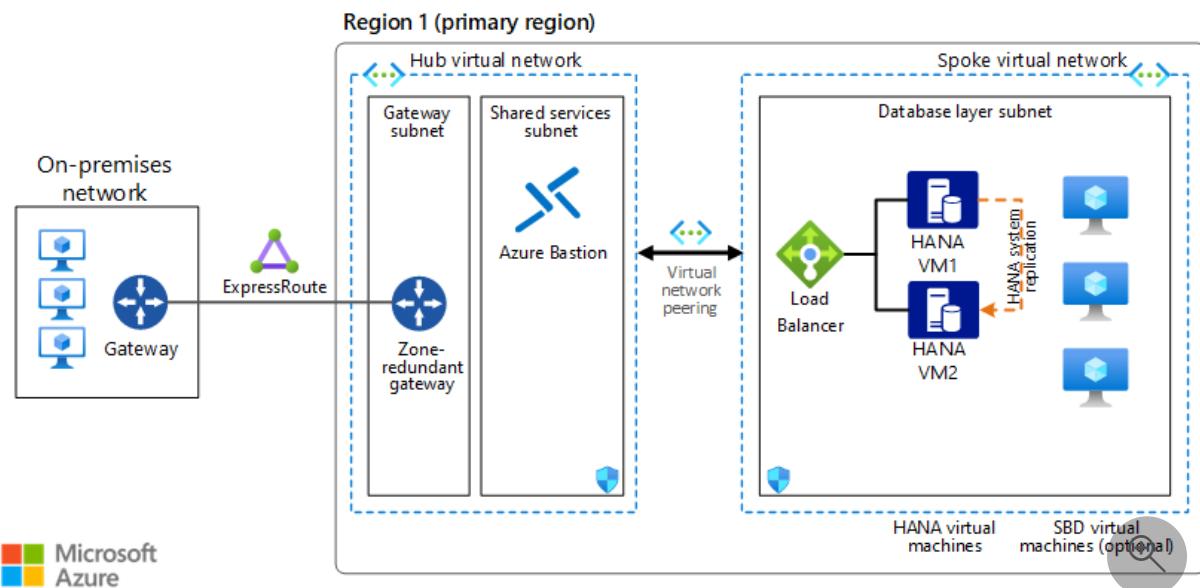
Azure    Azure Virtual Machines

This reference architecture shows a set of proven practices for running SAP HANA in a highly available, scale-up environment that supports disaster recovery on Azure. This implementation focuses on the database layer only.

## Architecture

This reference architecture describes a common production system. You can choose the virtual machine sizes to accommodate your organization's needs. This configuration can also be reduced to one virtual machine, depending on business requirements.

The following diagram shows a reference architecture for SAP HANA on Azure:



Download a [Visio file](#) that contains the diagrams in this article.

### Note

To deploy this reference architecture, you need the appropriate licensing of SAP products and other non-Microsoft technologies.

# Workflow

This reference architecture describes a typical SAP HANA database running in Azure, in a highly available deployment to maximize system availability. The architecture and its components can be customized based on business requirements (RTO, RPO, uptime expectations, system role) and potentially reduced to a single VM. The network layout is simplified to demonstrate the architectural principals of such SAP environment and not intended to describe a full enterprise network.

## Networking

**Virtual networks.** The [Azure Virtual Network](#) service connects Azure resources to each other with enhanced security. In this architecture, the virtual network connects to an on-premises environment via an ExpressRoute gateway deployed in the hub of a [hub-spoke topology](#). SAP HANA database is contained in own spoke virtual network. The spoke virtual networks contains one subnet for the database virtual machines (VMs).

If applications connecting to SAP HANA are running on VMs, the application VMs should be located in same virtual network but within a dedicated application subnet. Alternatively, if SAP HANA connection isn't the primary database, the application VMs can be located in other virtual networks. Separating into subnets by workload allows easier enablement of network security groups (NSG) to set security rules applicable to SAP HANA VMs only.

**Zone-redundant gateway.** A gateway connects distinct networks, extending your on-premises network to the Azure virtual network. We recommend that you use [ExpressRoute](#) to create private connections that don't go over the public internet. You can also use a [site-to-site](#) connection. Azure ExpressRoute or VPN gateways can be deployed across zones to guard against zone failures. See [Zone-redundant virtual network gateways](#) to understand the differences between a zonal deployment and a zone-redundant deployment. The IP addresses used need to be of Standard SKU for a zone deployment of the gateways.

**Network security groups (NSG).** To restrict incoming and outgoing network traffic of the virtual network, create [network security groups](#), which are in turn assigned to specific subnets. DB and application subnets are secured with workload specific NSGs.

**Application security groups (ASG).** To define fine-grained network security policies inside your NSGs based on workloads that are centered on applications, use [application security groups](#) instead of explicit IP addresses. They let you group network interfaces of VMs by name and help you secure applications by filtering traffic from trusted segments of your network.

**Network interface cards (NICs).** Network interface cards enable all communication among virtual machines on a virtual network. Traditional on-premises SAP deployments implement multiple NICs per machine to segregate administrative traffic from business traffic.

On Azure, it's not necessary to use multiple NICs for performance reasons. Multiple NICs share the same network throughput limit of a VM. But if your organization needs to segregate traffic, you can deploy multiple NICs per VM and connect each NIC to a different subnet. You can then use network security groups to enforce different access control policies on each subnet.

Azure NICs support multiple IPs. This support conforms with the SAP recommended practice of using virtual host names for installations. For a complete outline, see [SAP note 962955](#). (To access SAP notes, you need an SAP Service Marketplace account.)

#### Note

As specified in [SAP Note 2731110](#), do not place any network virtual appliance (NVA) in between the application and the database layers for any SAP application stack. Doing so introduces significant data packets processing time and unacceptably slows application performance.

## Virtual machines

This architecture uses virtual machines (VM). Azure offers single-node scale up to 23.5 Tebibytes (TiB) of memory on virtual machines. The [SAP Certified and Supported SAP HANA Hardware Directory](#) lists the virtual machines that are certified for the SAP HANA database. For details about SAP support for virtual machine types and throughput metrics (SAPS), see [SAP Note 1928533 - SAP Applications on Microsoft Azure: Supported Products and Azure VM types](#). (To access this and other SAP notes, an SAP Service Marketplace account is required.)

Microsoft and SAP jointly certify a range of virtual machine sizes for SAP HANA workloads. For example, smaller deployments can run on an [Edsv4](#) or [Edsv5](#) virtual machine with 160 GiB or more of RAM. To support the largest SAP HANA memory sizes on virtual machines, as much as 23 TiB, you can use [Mv2-series](#) virtual machines. M208 virtual machine types achieve approximately 260,000 SAPS, and M832ixs virtual machine types achieve approximately 795,900 SAPS.

**Generation 2 (Gen2) virtual machines.** When you deploy VMs, you can use either generation 1 or generation 2 VMs. [Generation 2 VMs](#) support key features that aren't

available for generation 1 VMs. For SAP HANA, this is particularly important because some VM families, like [Mv2](#) and [MdsV2](#), are supported only as Gen2 VMs. Similarly, SAP on Azure certification for some newer VMs might require them to be only Gen2 for full support, even if Azure allows both on them. See details in [SAP Note 1928533 - SAP Applications on Microsoft Azure: Supported Products and Azure VM types](#).

Because all other VMs supporting SAP HANA allow the choice of either Gen2 only or Gen1+2 selectively, we recommend that you deploy all SAP VMs as Gen2 only. This applies also to VMs with low memory requirements. Even the smallest 160 GiB SAP HANA VM can run as Gen2 VM and can be, when deallocated, be resized to the largest VM available in your region and subscription.

**Proximity Placement Groups (PPG).** To optimize network latency, you can use [proximity placement groups](#), which favor collocation, meaning that virtual machines are in the same datacenter to minimize latency between SAP HANA and connecting application VMs. For SAP HANA architecture itself, no PPGs are needed, they're only an option colocating SAP HANA with application tier VMs. Due to potential restrictions with PPGs, adding the database AvSet to the SAP system's PPG should be done sparsely and **only when required** for latency between SAP application and database traffic. For more information on the usage scenarios of PPGs, see the linked documentation. As PPGs restrict workloads to a single datacenter, a PPG can't span multiple availability zones.

## Components

- [Azure Virtual Network](#)
- [Azure ExpressRoute](#)
- [Azure Virtual Machines](#)
- [Azure NetApp Files](#)
- [Azure Load Balancer](#)
- [Azure Disk Storage](#)

## Considerations

This section describes key considerations for running SAP HANA on Azure.

## Scalability

This architecture runs SAP HANA on virtual machines that can scale up to 23 TiB in one instance.

If your workload exceeds the maximum virtual machine size, we recommend that you use multi-node HANA configurations. For online transaction processing (OLTP) applications, total scale-out memory capacity can be as high as 4 x 23 TiB. For online analytical processing (OLAP) applications, the scale-out memory capacity can be as high as 16 x 7.6 TiB. For example, you can deploy SAP HANA in a scale-out configuration with standby on virtual machines—running either [Red Hat Enterprise Linux](#) or [SUSE Linux Enterprise Server](#)—using [Azure NetApp Files](#) for the shared storage volumes.

## Storage

This architecture uses [Azure managed disks](#) for storage on the virtual machines or Azure NetApp Files. Guidelines for storage deployment with managed disks are in detail within the [SAP HANA Azure virtual machine storage configurations document](#). Alternatively to managed disks, [Azure NetApp Files NFS](#) volumes can be used as storage solution for SAP HANA.

To achieve high input/output operations per second (IOPS) and disk storage throughput, the common practices in storage volume [performance optimization](#) also apply to Azure storage layout. For example, combining multiple disks together with LVM to create a striped disk volume improves IO performance. Azure disk caching also plays a significant role in achieving required IO performance.

For SAP HANA log disks that run on Azure Premium SSD v1, use one of the following technologies in locations that hold `/hana/log` for production:

- [Write Accelerator](#) (on M series VMs)
- [Ultra disks](#) (on either M or E series VMs)
- [Azure NetApp Files](#) (on either M or E series VMs)

These technologies are needed to consistently meet the required storage latency of less than 1 ms.

[Azure Premium SSD v2](#) is designed for performance-critical workloads like SAP. Write Accelerator isn't required when `/hana/log` is running on Premium SSD v2. For information about this storage solution's benefits and current limitations, see [Deploy a Premium SSD v2](#).

For details about SAP HANA performance requirements, see [SAP Note 1943937 - Hardware Configuration Check Tool](#).

- **Cost-conscious storage design for non-production systems.** For SAP HANA environments that don't require maximum storage performance in all situations, you can use a storage architecture that's optimized for cost. This choice of storage

optimization can apply to little-used production systems or some non-production SAP HANA environments. The cost-optimized storage option uses a combination of Standard SSDs instead of the Premium or Ultra SSDs that are used for production environments. It also combines `/hana/data` and `/hana/log` file systems onto a single set of disks. [Guidelines and best practices](#) are available for most VM sizes. If you use Azure NetApp Files for SAP HANA, you can use size-reduced volumes to achieve the same goal.

- **Resizing storage when scaling-up.** When you resize a virtual machine because of changed business demands or because of a growing database size, the storage configuration can change. Azure supports online disk expansion, without any interruption to service. With a striped disk setup, as used for SAP HANA, a resize operation should be done equally to all disks in the volume group. The addition of more disks to a volume group can potentially unbalance the striped data. If you're adding more disks to a storage configuration, it's far preferable to create a new storage volume on new disks. Next, copy the contents during downtime and modify mountpoints. Finally, discard the old volume group and underlying disks.
- **Azure NetApp Files application volume group.** For deployments with SAP HANA files contained on Azure NetApp Files NFS volumes, application volume groups enable you to deploy all volumes according to best practices. This process also ensures optimal performance for your SAP HANA database. [Details are available](#) about how to proceed with this process. It requires manual intervention. Allow some time for the creation.

## High availability

The preceding architecture depicts a highly available deployment, with SAP HANA contained on two or more virtual machines. The following components are used.

**Load balancers.** [Azure Load Balancer](#) is used to distribute traffic to SAP HANA virtual machines. When you incorporate Azure Load Balancer in a zonal deployment of SAP, make sure you select the Standard SKU load balancer. The Basic SKU balancer doesn't support zonal redundancy. In this architecture, Load Balancer acts as the virtual IP address for SAP HANA. Network traffic is sent to the active VM where the primary database instance runs. SAP HANA active/read-enabled architecture is available ([SLES/RHEL](#)) where a second virtual IP addressed on the load balancer is used to direct network traffic to the secondary SAP HANA instance on other VM for read-intense workloads.

The Standard Load Balancer provides a layer of security by default. Virtual machines that are behind the Standard Load Balancer don't have outbound internet connectivity. To

enable outbound internet in these virtual machines, you need to update your [Standard Load Balancer](#) configuration. In addition, you can also use an [Azure NAT Gateway](#) to get outbound connectivity.

For SAP HANA database clusters, you must enable Direct Server Return (DSR), also known as Floating IP. This feature allows the server to respond to the IP address of the load balancer front end. This direct connection keeps the load balancer from becoming the bottleneck in the path of data transmission.

**Deployment options.** On Azure, SAP workload deployment can be either regional or zonal, depending on the availability and resiliency requirements of the SAP applications. Azure provides [different deployment options](#), like Virtual Machine Scale Sets with Flexible orchestration (FD=1), availability zones, and availability sets, to enhance the availability of resources. To get a comprehensive understanding of the available deployment options and their applicability across different Azure regions (including across zones, within a single zone, or in a region without zones), see [High-availability architecture and scenarios for SAP NetWeaver](#).

**SAP HANA.** For high availability, SAP HANA runs on two or more Linux virtual machines. SAP HANA System Replication (HSR) is used to replicate data between the primary and secondary (replica) SAP HANA systems. HSR is also used for cross-region or cross-zone disaster recovery. Depending on latency in the communication between your virtual machines, synchronous replication can be used within a region. HSR between regions for disaster recovery will in most cases be running in asynchronous manner.

For the Linux Pacemaker cluster, you need to decide which cluster fencing mechanism to use. Cluster fencing is the process of isolating a failed VM from the cluster and restarting it. For RedHat Enterprise Linux (RHEL), the only supported fencing mechanism for Pacemaker on Azure is Azure fence agent. For SUSE Linux Enterprise Server (SLES), you can use either Azure fence agent or STONITH Block Device (SBD). Compare the failover times for each solution and, if there's a difference, choose a solution based on your business requirements for recovery time objective (RTO).

**Azure fence agent.** This fencing method relies on the Azure ARM API, with Pacemaker querying ARM api about the status of both SAP HANA VMs in the cluster. Should one VM fail, for example OS unresponsive or VM crash, the cluster manager uses again the ARM api to restart the VM and if needed fails the SAP HANA database to the other, active node. For this purpose, a service name principal ([SPN](#)) with a custom role to query and restart VMs is used to authorize against the ARM api. No other infrastructure is needed, the SBD VMs in the architecture drawings aren't deployed in case Azure fence agent is used.

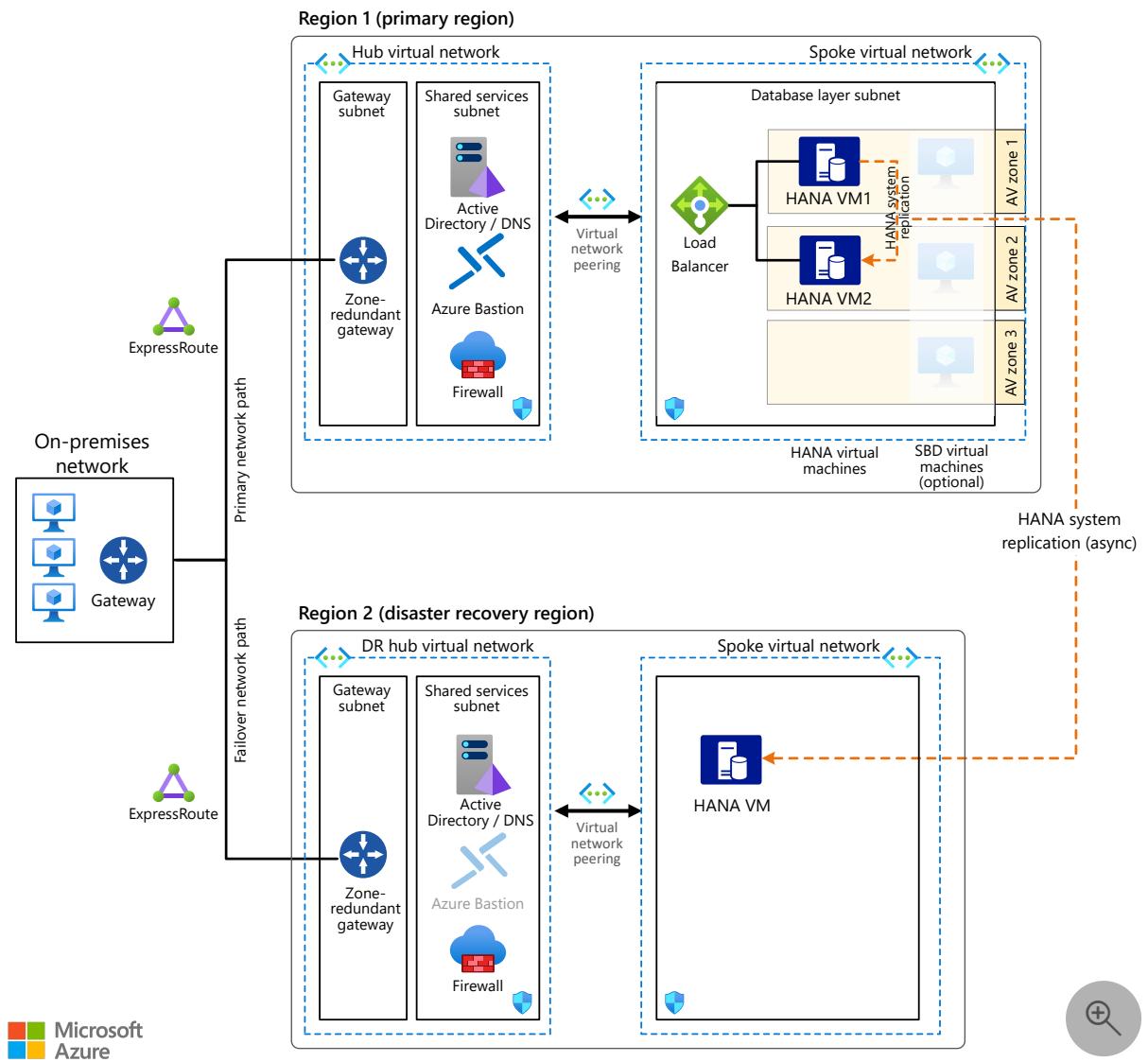
**SBD.** STONITH block device (SBD) uses a disk that is accessed as block device (raw, without filesystem) by the cluster manager. This disk, or disks if multiple, acts as a vote. Each of the two cluster nodes running SAP HANA accesses the SDB disks and reads/writes periodically to them small bits of information about status. Thus each cluster node knows the status about the other without depending only on networking between the VMs.

Preferably three small VMs are deployed in either an availability set or availability zone setup. Each VM exporting small parts of a disk as a block device which is accessed by the two SAP HANA cluster nodes. Three SBD VMs ensure sufficient voting members are available in case of planned or unplanned downtime for either SBD VM.

Alternatively to using SBD VMs, [Azure shared disk](#) can be used instead. The SAP HANA cluster nodes then [access the single shared disk](#). The shared disk can be locally ([LRS](#)) or zonally ([ZRS](#)) redundant, if ZRS is available in your Azure region.

## Disaster recovery

The following architecture shows a production HANA environment on Azure that provides disaster recovery. The architecture incorporates availability zones.



For DR strategies and implementation details, see [Disaster recovery overview](#) and [infrastructure guidelines for SAP workload](#) and [Disaster recovery guidelines for SAP application](#).

### ⓘ Note

If there's a regional disaster that causes a large failover event for many Azure customers in one region, the target region's **resource capacity** isn't guaranteed. Like all Azure services, Azure Site Recovery continues to add features and capabilities. For the latest information about Azure-to-Azure replication, see the [support matrix](#).

In addition to a local, two-node high availability implementation, HSR supports [multitier](#) and [multitarget](#) replication. HSR therefore supports inter-zone and inter-region replication. Multitarget replication is available for SAP HANA 2.0 SPS 03 and later.

Make sure to verify your target region's [resource capacity](#).

**Azure NetApp Files.** As an option, [Azure NetApp Files](#) can be used to provide a scalable and high-performance storage solution for SAP HANA data and log files. Azure NetApp Files supports snapshots for fast backup, recovery, and local replication. For cross-region content replication, Azure NetApp Files Cross-Region Replication can be used to replicate the snapshot data between two regions. [Details](#) about cross-region replication and a [whitepaper](#) describing all aspects for disaster recovery with Azure NetApp Files are available.

## Backup

SAP HANA data can be backed up in many ways. After migrating to Azure, you can continue to use any existing partner backup solutions you already have. Azure provides two native approaches: [SAP HANA file-level backup](#) and Azure Backup for SAP HANA over the Backint interface.

For SAP HANA file-level backup, you can use your tool of choice, such as hdbsql or SAP HANA Studio, and store the backup files on a local disk volume. A common mount point for this backup volume is `/hana/backup`. Your backup policies will define the data retention period on the volume. As soon as the backup is taken, a scheduled task should copy the backup files to Azure Blob storage for safekeeping. The local backup files are kept for expedient recovery.

Azure Backup offers a simple, enterprise-grade solution for workloads running on virtual machines. [Azure Backup for SAP HANA](#) provides full integration with the SAP HANA backup catalog and guarantees database-consistent, full, or point-in-time recoveries. Azure Backup is [BackInt-certified](#) by SAP. See also the [Azure Backup FAQ](#) and [support matrix](#).

**Azure NetApp Files** brings support for snapshot based backups. Integrating with SAP HANA for application consistent snapshots is through the Azure Application Consistent Snapshot tool ([AzAcSnap](#)). The snapshots created can be used for restore to a new volume for system restore or copying the SAP HANA database. Snapshots created can be used for disaster recovery, where it acts as restore point with SAP HANA logs saved on a different NFS volume.

## Monitoring

To monitor your workloads on Azure, [Azure Monitor](#) lets you comprehensively collect, analyze, and act on telemetry from your cloud and on-premises environments.

For SAP applications that run on SAP HANA and other major database solutions, see [Azure Monitor for SAP solutions](#) to learn how Azure Monitor for SAP can help you

manage the availability and performance of SAP services.

## Security

Many security measures are used to protect the confidentiality, integrity, and availability of an SAP landscape. To secure user access, for example, SAP has its own User Management Engine (UME) to control role-based access and authorization within the SAP application and databases. For more information, see [SAP HANA Security—An Overview](#).

For data at rest, different encryption functionalities provide security as follows:

- Along with the SAP HANA native encryption technology, consider using an encryption solution from a partner that supports customer-managed keys.
- To encrypt virtual machine disks, you can use functionalities described in [Disk Encryption Overview](#).
- SAP Database servers: Use Transparent Data Encryption offered by the DBMS provider (for example, *SAP HANA native encryption technology*) to help secure your data and log files and to ensure the backups are also encrypted.
- Data in Azure physical storage (Server-Side Encryption) is automatically encrypted at rest with an Azure managed key. You can also choose a customer managed key (CMK) instead of the Azure managed key.
- For information about support of Azure Disk Encryption on particular Linux distros, versions, and images, see [Azure Disk Encryption for Linux VMs](#).

### Note

Don't combine SAP HANA native encryption technology with Azure Disk Encryption or Host Based Encryption on the same storage volume. Also, operating system boot disks for Linux virtual machines don't support Azure Disk Encryption. Instead, when you use SAP HANA native encryption, combine it with Server-Side Encryption, which is automatically enabled. Be aware that the usage of customer-managed keys might affect storage throughput.

For network security, use network security groups (NSGs) and Azure Firewall or a network virtual appliance as follows:

- Use [NSGs](#) to protect and control traffic between subnets and application/database layers. Only apply NSGs to subnets. NSGs applied to both NIC and subnet very

often lead to problems during troubleshooting and should be used rarely if ever.

- Use [Azure Firewall](#) or Azure network virtual appliance to inspect and control the routing of traffic from the hub virtual network to the spoke virtual network where your SAP applications are, and also to control your outbound internet connectivity.

For User and Authorization, implement role-based access control (RBAC) and resource locks as follows:

- Follow the principle of least privilege, using [RBAC](#) for assigning administrative privileges at IaaS-level resources that host your SAP solution on Azure. The fundamental purpose of RBAC is the segregation and control of duties for your users/group. RBAC is designed to grant only the amount of access to resources that's needed to enable users to do their jobs.
- Use [resource locks](#) to help prevent accidental or malicious changes. Resource locks help prevent administrators from deleting or modifying critical Azure resources where your SAP solution is located.

More security recommendations can be found at thesees [Microsoft](#) and [SAP](#) articles.

## Communities

Communities can answer questions and help you set up a successful deployment.

Consider the following communities:

- [Azure Community Support](#)
- [SAP Community](#)
- [SAP on Stack Overflow](#)

## Contributors

*This article is maintained by Microsoft. It was originally written by the following contributors.*

Principal author:

- [Robert Biro](#) | Senior Architect

*To see non-public LinkedIn profiles, sign in to LinkedIn.*

# Next steps

Learn more about the component technologies:

- [What is Azure ExpressRoute?](#)
- [What is Azure Bastion?](#)
- [What is Power BI?](#)
- [Use the SAP Business Warehouse connector in Power BI Desktop](#)
- [SAP workload configurations with Azure Availability Zones](#)
- [What is the Azure Backup service?](#)
- [About Site Recovery](#)
- [What is Azure Load Balancer?](#)
- [Connect to SAP HANA databases in Power BI](#)
- [What is Azure NetApp Files](#)
- [Introduction to Azure managed disks](#)
- [Linux virtual machines in Azure](#)
- [Installation of SAP HANA on Azure virtual machines](#)
- [What is Azure Virtual Network?](#)
- [Network security groups](#)
- [SAP HANA Disaster Recovery with Azure NetApp Files ↗](#)

## Related resources

Explore related architectures:

- [Run a Linux VM on Azure](#)
- [Run SAP BW/4HANA with Linux virtual machines on Azure](#)
- [SAP S/4HANA in Linux on Azure](#)
- [SAP on Azure Architecture Guide](#)

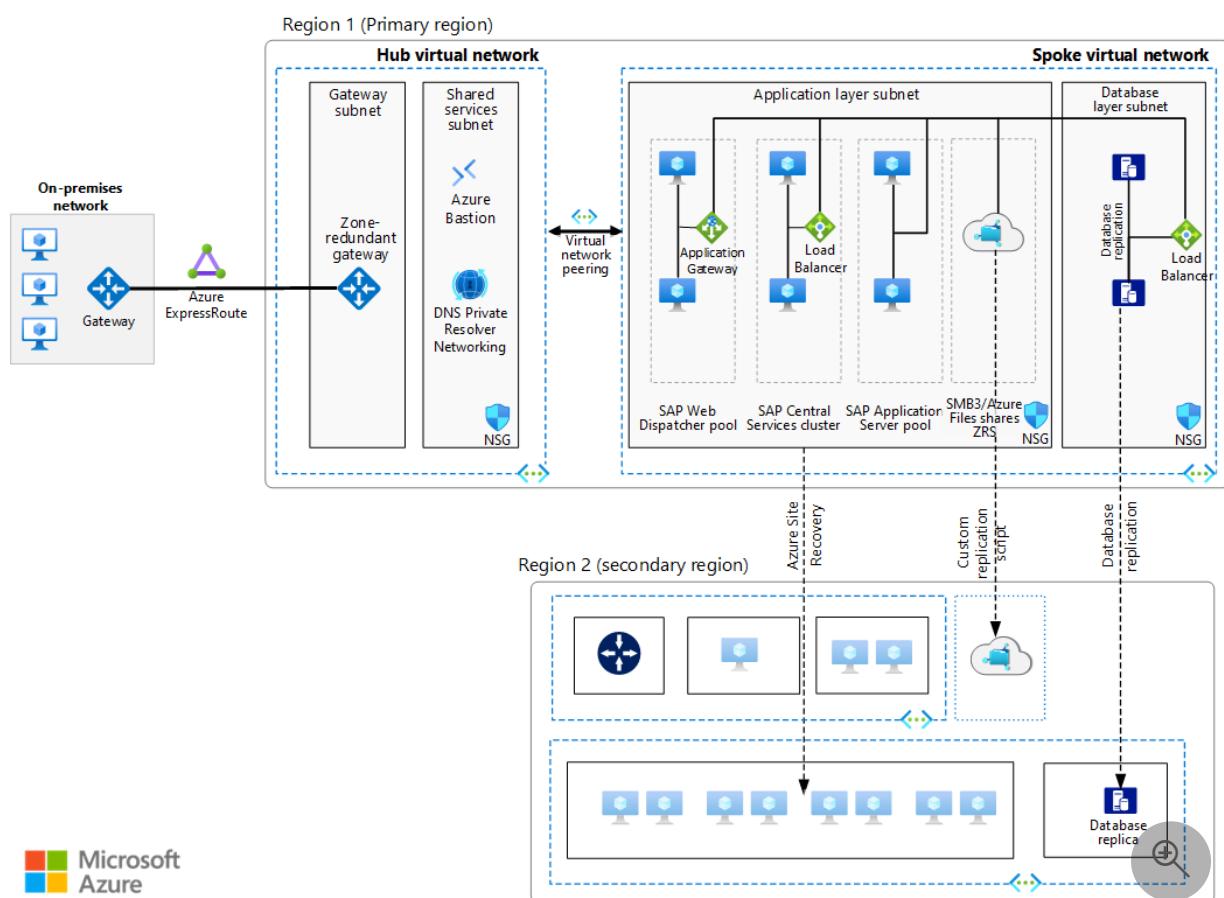
# Run SAP NetWeaver in Windows on Azure

Azure ExpressRoute    SAP HANA on Azure Large Instances    Azure Virtual Machines    Azure Virtual Network  
Azure NetApp Files

This guide presents a set of proven practices for running SAP NetWeaver in a Windows environment, on Azure, with high availability. The database is AnyDB, the SAP term for any supported database management system (DBMS) besides SAP HANA.

## Architecture

The following diagram shows SAP NetWeaver in a Windows environment.



Download a [Visio file](#) of this architecture.

Note

To deploy this architecture, you need appropriate licensing of SAP products and other non-Microsoft technologies.

This guide describes a production system. The system is deployed with specific virtual machine (VM) sizes that you can change to accommodate the needs of your organization. The system can be reduced to a single VM. In this guide, the network layout is greatly simplified to demonstrate architectural principles. It's not intended to describe a full enterprise network.

## Workflow

**Virtual networks.** The [Azure Virtual Network](#) service connects Azure resources to each other with enhanced security. In this architecture, the virtual network connects to an on-premises network via an Azure ExpressRoute gateway that's deployed in the hub of a [hub-spoke topology](#). The spoke is the virtual network that's used for the SAP applications and the database tiers. The hub virtual network is used for shared services like Azure Bastion and backup.

**Virtual network peering.** This architecture uses a hub-and-spoke networking topology with multiple virtual networks that are [peered together](#). This topology provides network segmentation and isolation for services that are deployed on Azure. Peering enables transparent connectivity between peered virtual networks through the Microsoft backbone network. It doesn't incur a performance penalty if deployed within a single region. The virtual network is divided into separate subnets for each tier: application (SAP NetWeaver), the database, and shared services like Bastion and a third-party backup solution.

**VMs.** This architecture uses VMs for the application tier and database tier, grouped in the following way:

- **SAP NetWeaver.** The application tier uses Windows VMs to run SAP Central Services and SAP application servers. For high availability, the VMs that run Central Services are configured in a Windows server failover cluster. They're supported by either Azure file shares or Azure shared disks.
- **AnyDB.** The database tier runs AnyDB as the database, which can be Microsoft SQL Server, Oracle, or IBM Db2.
- **Bastion service.** Administrators use an improved-security VM that's called a bastion host to connect to other VMs. It's typically a part of shared services, like backup services. If Secure Shell Protocol (SSH) and Remote Desktop Protocol (RDP) are the only services that are used for server administration, an [Azure Bastion](#) host

is a good solution. If you use other management tools, like SQL Server Management Studio or SAP Frontend, use a traditional, self-deployed jump box.

**Private DNS service.** [Azure Private DNS](#) provides a reliable and secure DNS service for your virtual network. Azure Private DNS manages and resolves domain names in the virtual network, without the need to configure a custom DNS solution.

**Load balancers.** To distribute traffic to VMs in the SAP application tier subnet for high availability, we recommend that you use an [Azure standard load balancer](#). Note that a standard load balancer provides a level of security by default. VMs that are behind a standard load balancer don't have outbound internet connectivity. To enable outbound internet on the VMs, you need to update your [standard load balancer configuration](#). For SAP web-based application high availability, use the built-in [SAP Web Dispatcher](#) or another commercially available load balancer. Base your selection on:

- Your traffic type, like HTTP or SAP GUI.
- The network services that you need, like Secure Sockets Layer (SSL) termination.

For some internet-facing inbound/outbound design examples, see [Inbound and outbound internet connections for SAP on Azure](#).

Standard Load Balancer supports multiple front-end virtual IPs. This support is ideal for cluster implementations that involve these components:

- Advanced Business Application Programming (ABAP) SAP Central Service (ASCS)
- Enqueue Replication Server (ERS)

The Standard SKU also supports multi-systems identifier (multi-SID) SAP clusters. In other words, [multiple SAP systems on Windows](#) can share a common high availability infrastructure to save cost. Evaluate the cost savings, and avoid placing too many systems in one cluster. Azure supports no more than five SIDs per cluster.

**Application gateway.** Azure Application Gateway is a web traffic load balancer that you can use to manage the traffic to your web applications. Traditional load balancers operate at the transport layer (OSI layer 4 - TCP and UDP). They route traffic based on the source IP address and the port to a destination IP address and port. Application Gateway can make routing decisions based on additional attributes of an HTTP request, such as the URI path or host headers. This type of routing is known as application layer (OSI layer 7) load balancing.

**Network security groups.** To restrict incoming, outgoing, and intra-subnet traffic in a virtual network, create [network security groups](#).

**Application security groups.** To define fine-grained, workload-based network security policies that are centered on applications, use [application security groups](#) instead of explicit IP addresses. Application security groups provide a way to group VMs by name and help you secure applications by filtering traffic from trusted segments of your network.

**Gateway.** A gateway connects distinct networks, extending your on-premises network to the Azure virtual network. We recommend that you use [ExpressRoute](#) to create private connections that don't go over the public internet, but you can also use a [site-to-site](#) connection. To reduce latency or increase throughput, consider [ExpressRoute Global Reach](#) and [ExpressRoute FastPath](#), as discussed later in this article.

**Azure Storage.** Storage provides data persistence for a VM in the form of a virtual hard disk. We recommend [Azure managed disks](#).

## Recommendations

This architecture describes a small production-level deployment. Deployments differ based on business requirements, so consider these recommendations as a starting point.

## VMs

In application server pools and clusters, adjust the number of VMs based on your requirements. For detailed information about running SAP NetWeaver on VMs, see [Azure Virtual Machines planning and implementation for SAP NetWeaver](#).

For details about SAP support for Azure VM types and throughput metrics (SAPS), see [SAP note 1928533](#). To access SAP notes, you need an SAP Service Marketplace account.

## SAP Web Dispatcher

The Web Dispatcher component is used for load-balancing SAP traffic among the SAP application servers. To achieve high availability for the Web Dispatcher component, Load Balancer is used to implement either the failover cluster of Web Dispatcher instances or the parallel Web Dispatcher setup. For a detailed description of the solution, see [High Availability of SAP Web Dispatcher](#).

## Application servers pool

The SAP SMLG transaction is commonly used to manage logon groups for ABAP application servers and to load balance logon users. Other transactions, like SM61 for batch server groups and RZ12 for remote function call (RFC) groups, also load balance logon users. These transactions use the load-balancing capability within the message server of SAP Central Services to distribute incoming sessions or workloads among the SAP application servers pool for SAP GUIs and RFC traffic.

## SAP Central Services cluster

This architecture runs Central Services on VMs in the application tier. Central Services is a potential single point of failure (SPOF) when it's deployed to a single VM. To implement a highly available solution, use either a file-share cluster or a shared-disk cluster.

For highly available file shares, there are several options. We recommend that you use [Azure Files](#) shares as fully managed, cloud-native Server Message Block (SMB) or Network File System (NFS) shares. An alternative to Azure Files is [Azure NetApp Files](#), which provides high-performance NFS and SMB shares.

You can also implement the highly available file shares on the Central Services instances by using a Windows server failover cluster with [Azure Files](#). This solution also supports a Windows cluster with shared disks by using an Azure shared disk as the cluster shared volume. If you prefer to use shared disks, we recommend that you use [Azure shared disks](#) to set up a [Windows Server failover cluster for SAP Central Services cluster](#).

There are also partner products like [SIOS DataKeeper Cluster Edition](#) from SIOS Technology Corp. This add-on replicates content from independent disks that are attached to the ASCS cluster nodes and then presents the disks as a cluster shared volume to the cluster software.

If you use cluster network partitioning, the cluster software uses quorum votes to select a segment of the network and its associated services to serve as the brain of the fragmented cluster. Windows offers many quorum models. This solution uses [Cloud Witness](#) because it's simpler and provides more availability than a compute node witness. The [Azure file share witness](#) is another alternative for providing a cluster quorum vote.

On an Azure deployment, the application servers connect to the highly available Central Services by using the virtual host names of the ASCS or ERS services. These host names are assigned to the cluster front-end IP configuration of the load balancer. Load Balancer supports multiple front-end IPs, so both the ASCS and ERS virtual IPs (VIPs) can be bound to one load balancer.

# Networking

This architecture uses a hub-spoke topology. The hub virtual network acts as a central point of connectivity to an on-premises network. The spokes are virtual networks that peer with the hub and isolate the SAP workloads. Traffic flows between the on-premises datacenter and the hub through a gateway connection.

## Network interface cards (NICs)

NICs enable all communication among VMs on a virtual network. Traditional on-premises SAP deployments implement multiple NICs per machine to segregate administrative traffic from business traffic.

On Azure, the virtual network is a software-defined network that sends all traffic through the same network fabric. So it's not necessary to use multiple NICs for performance reasons. But if your organization needs to segregate traffic, you can deploy multiple NICs per VM and connect each NIC to a different subnet. You can then use network security groups to enforce different access control policies.

Azure NICs support multiple IPs. This support conforms with the practice that SAP recommends of using virtual host names for installations. For a complete outline, see [SAP note 962955](#). To access SAP notes, you need an SAP Service Marketplace account.

## Subnets and network security groups

This architecture subdivides the virtual network address space into subnets. You can associate each subnet with a network security group that defines the access policies for the subnet. Place application servers on a separate subnet. By doing so, you can secure them more easily by managing the subnet security policies rather than the individual servers.

When you associate a network security group with a subnet, the network security group applies to all the servers within the subnet and offers fine-grained control over the servers. Set up network security groups by using the [Azure portal](#), [PowerShell](#), or the [Azure CLI](#).

## ExpressRoute Global Reach

If your network environment includes two or more ExpressRoute connections, [ExpressRoute Global Reach](#) can help you reduce network hops and latency. This technology is a Border Gateway Protocol (BGP) route peering that's set up between two or more ExpressRoute connections to bridge two ExpressRoute routing domains. Global

Reach reduces latency when network traffic traverses more than one ExpressRoute connection. It's currently available only for private peering on ExpressRoute circuits.

At this time, there are no network access control lists or other attributes that can be changed in Global Reach. So all routes learned by a given ExpressRoute circuit (from on-premises and Azure) are advertised across the circuit peering to the other ExpressRoute circuit. We recommend that you establish network traffic filtering on-premises to restrict access to resources.

## ExpressRoute FastPath

[FastPath](#) is designed to improve the data path performance between your on-premises network and your virtual network. When it's enabled, FastPath sends network traffic directly to virtual machines in the virtual network, bypassing the gateway.

For all new ExpressRoute connections to Azure, FastPath is the default configuration. For existing ExpressRoute circuits, contact Azure support to activate FastPath.

FastPath doesn't support virtual network peering. If other virtual networks are peered with one that's connected to ExpressRoute, the network traffic from your on-premises network to the other spoke virtual networks is sent to the virtual network gateway. The workaround is to connect all virtual networks to the ExpressRoute circuit directly. This feature is currently in public preview.

## Load balancers

[SAP Web Dispatcher](#) handles load balancing of HTTP(S) traffic to a pool of SAP application servers. This software load balancer provides application layer services (referred to as layer 7 in the ISO networking model) that can perform SSL termination and other offloading functions.

[Azure Load Balancer](#) is a network transmission layer service (layer 4) that balances traffic by using a five-tuple hash from the data streams. The hash is based on source IP, source port, destination IP, destination port, and protocol type. In SAP deployments on Azure, Load Balancer is used in cluster setups to direct traffic to the primary service instance or to the healthy node if there's a fault.

We recommend that you use [Standard Load Balancer](#) for all SAP scenarios. If VMs in the back-end pool require public outbound connectivity, or if they're used in an Azure zone deployment, Standard Load Balancer requires [additional configurations](#) because they're secure by default. They don't allow outbound connectivity unless you explicitly configure it.

For traffic from SAP GUI clients that connect to an SAP server via DIAG protocol or RFC, the Central Services message server balances the load through SAP application server [logon groups](#). For this type of setup, you don't need another load balancer.

## Storage

Some organizations use standard storage for their application servers. Standard managed disks aren't supported. See [SAP note 1928533](#). To access SAP notes, you need an SAP Service Marketplace account. We recommend that you use premium [Azure managed disks](#) in all cases. A recent update to [SAP note 2015553](#) excludes the use of Standard HDD storage and Standard SSD storage for a few specific use cases.

Application servers don't host business data. So you can also use the smaller P4 and P6 premium disks to help minimize costs. By doing so, you can benefit from the [single-instance VM SLA](#) if you have a central SAP stack installation.

For high-availability scenarios, you can use [Azure file shares](#) and [Azure shared disks](#). [Azure Premium SSD managed disks](#) and [Azure Ultra Disk Storage](#) are available for Azure shared disks, and Premium SSD is available for Azure file shares.

Storage is also used by [Cloud Witness](#) to maintain quorum with a device in a remote Azure region, away from the primary region where the cluster resides.

For the backup data store, we recommend Azure [cool and archive access tiers](#). These storage tiers provide a cost-effective way to store long-lived data that's infrequently accessed.

[Azure Premium SSD v2 disk storage](#) is designed for performance-critical workloads like online transaction processing systems that consistently need sub-millisecond latency combined with high IOPS and throughput.

[Ultra Disk Storage](#) greatly reduces disk latency. As a result, it benefits performance-critical applications like the SAP database servers. To compare block storage options in Azure, see [Azure managed disk types](#).

For a high-availability, high-performance shared data store, use [Azure NetApp Files](#). This technology is particularly useful for the database tier when you use [Oracle](#), and also when you [host application data](#).

## Performance considerations

SAP application servers communicate constantly with the database servers. For performance-critical applications that run on database platforms, enable [Write](#)

Accelerator for the log volume if you're using Premium SSD v1. Doing so can improve log-write latency. Write Accelerator is available for M-series VMs.

To optimize inter-server communications, use [Accelerated Networking](#). Most general-purpose and compute-optimized VM instance sizes that have two or more vCPUs support Accelerated Networking. On instances that support hyperthreading, VM instances with four or more vCPUs support Accelerated Networking.

To achieve high IOPS and disk throughput, follow the common practices in storage volume [performance optimization](#), which apply to Azure storage layout. For example, you can position multiple disks together to create a striped disk volume to improve I/O performance. Enabling the read cache on storage content that changes infrequently enhances the speed of data retrieval.

[Premium SSD v2](#) provides higher performance than Premium SSDs and is generally less expensive. You can set a Premium SSD v2 disk to any supported size you prefer and make granular adjustments to the performance without downtime.

[Ultra Disk Storage](#) is available for I/O-intensive applications. Where these disks are available, we recommend them over [Write Accelerator](#) premium storage. You can individually increase or decrease performance metrics like IOPS and MBps without needing to reboot.

For guidance about optimizing Azure storage for SAP workloads on SQL Server, see [Azure Virtual Machines planning and implementation for SAP NetWeaver](#).

The placement of a network virtual appliance (NVA) between the application and the database layers for any SAP application stack isn't supported. This practice introduces significant processing time for data packets, which leads to unacceptable application performance.

## Proximity placement groups

Some SAP applications require frequent communication with the database. The physical proximity of the application and the database layers affects network latency, which can adversely affect application performance.

To optimize network latency, you can use [proximity placement groups](#), which set a logical constraint on the VMs that are deployed in availability sets. Proximity placement groups favor co-location and performance over scalability, availability, or cost. They can greatly improve the user experience for most SAP applications. For scripts that are available on GitHub from the SAP deployment team, see [Scripts](#).

## Availability zones

[Availability zones](#) provide a way for you to deploy VMs across datacenters, which are physically separated locations within a specific Azure region. Their purpose is to enhance service availability. But deploying resources across zones can increase latency, so keep performance considerations in mind.

Administrators need a clear network latency profile between all zones of a target region before they can determine the resource placement with minimum inter-zone latency. To create this profile, deploy small VMs in each zone for testing. Recommended tools for these tests include [PsPing](#) and [Iperf](#). When the tests are done, remove the VMs that you used for testing. As an alternative, consider using an [Azure inter-zone latency check tool](#).

## Scalability considerations

For the SAP application layer, Azure offers a wide range of VM sizes for scaling up and scaling out. For an inclusive list, see [SAP note 1928533 - SAP Applications on Azure: Supported Products and Azure VM Types](#). To access SAP notes, you need an SAP Service Marketplace account.

You can scale SAP application servers and the Central Services clusters up and down. You can also scale them out or in by changing the number of instances that you use. The AnyDB database can scale up and down but doesn't scale out. The SAP database container for AnyDB doesn't support sharding.

## Availability considerations

Resource redundancy is the general theme in highly available infrastructure solutions. For single-instance VM availability SLAs for various storage types, see [SLA for virtual machines](#). To increase service availability on Azure, deploy VM resources with Virtual Machine Scale Sets with Flexible orchestration, availability zones, or availability sets.

With Azure, SAP workload deployment can be either regional or zonal, depending on the availability and resiliency requirements of the SAP applications. Azure provides [different deployment options](#), like Virtual Machine Scale Sets with Flexible orchestration (FD=1), availability zones, and availability sets, to enhance the availability of resources. To get a comprehensive understanding of the available deployment options and their applicability across different Azure regions (including across zones, within a single zone, or in a region without zones), see [High-availability architecture and scenarios for SAP NetWeaver](#).

In this distributed installation of the SAP application, the base installation is replicated to achieve high availability. For each layer of the architecture, the high availability design varies.

## Web Dispatcher in the application servers tier

The Web Dispatcher component is used as a load balancer for SAP traffic among the SAP application servers. To achieve [high availability of SAP Web Dispatcher](#), Load Balancer implements either the failover cluster or the parallel Web Dispatcher setup.

For internet-facing communications, we recommend a stand-alone solution in the perimeter network, which is also known as *DMZ*, to satisfy security concerns.

[Embedded Web Dispatcher](#) on ASCS is a special option. If you use this option, consider proper sizing because of the extra workload on ASCS.

## Central Services in the application servers tier

High availability of the Central Services is implemented with a Windows server failover cluster. When the cluster storage for the failover cluster is deployed on Azure, you can configure it in two ways: as a clustered shared disk or as a clustered file share.

- We recommend that you use [Azure Files](#) as fully managed, cloud-native SMB or NFS shares. Another way is to use [Azure NetApp Files](#), which provides high-performance, enterprise-class NFS and SMB shares.
- There are two ways to set up clusters with shared disks on Azure. First, we recommend that you use [Azure shared disks](#) to set up a [Windows server failover cluster for SAP Central Services](#). For an implementation example, see [ASCS cluster using Azure shared disks](#). Another way to implement a clustered shared disk is to use SIOS DataKeeper to perform the following tasks:
  - Replicate the content of independent disks that are attached to the cluster nodes.
  - Abstract the drives as a cluster shared volume for the cluster manager.

For implementation details, see [Clustering SAP ASCS on Azure with SIOS](#).

If you use Standard Load Balancer, you can enable the [high availability port](#). By doing so, you can avoid needing to configure load-balancing rules for multiple SAP ports. Also, when you set up Azure load balancers, enable Direct Server Return (DSR), which is also called *Floating IP*. Doing so provides a way for server responses to bypass the load balancer. This direct connection keeps the load balancer from becoming a bottleneck in

the path of data transmission. We recommend that you enable DSR for the ASCS and database clusters.

## Application services in the application servers tier

High availability for the SAP application servers is achieved by load balancing traffic within a pool of application servers. You don't need cluster software, SAP Web Dispatcher, or the Azure load balancer. The SAP message server can load balance client traffic to the application servers defined in an ABAP logon group by the transaction SMLG.

## Database tier

In this architecture, the source database runs on AnyDB—a DBMS like SQL Server, SAP ASE, IBM Db2, or Oracle. The native replication feature of the database tier provides either manual or automatic failover between replicated nodes.

For implementation details about specific database systems, see [Azure Virtual Machines DBMS deployment for SAP NetWeaver](#).

## VMs deployed across availability zones

An availability zone consists of one or more datacenters. It's designed to improve workload availability and protect application services and VMs against datacenter outages. VMs in a single zone are treated as if they were in a single fault domain. When you select zonal deployment, VMs in the same zone are distributed to fault domains on a best-effort basis.

In [Azure regions](#) that support multiple zones, at least three zones are available. But the maximum distance between datacenters in these zones isn't guaranteed. To deploy a multitier SAP system across zones, you need to know the network latency within a zone and across targeted zones. You also need to know how sensitive your deployed applications are to network latency.

Take these [considerations](#) into account when you decide to deploy resources across availability zones:

- Latency between VMs in one zone
- Latency between VMs across chosen zones
- Availability of the same Azure services (VM types) in the chosen zones

### Note

Availability zones support intra-region high availability, but they aren't effective for disaster recovery (DR). The distances between zones are too short. Typical DR sites should be at least 100 miles away from the primary region.

### Active/inactive deployment example

In this example deployment, the [active/passive](#) status refers to the application service state within the zones. In the application layer, all four active application servers of the SAP system are in zone 1. Another set of four passive application servers is built in zone 2 but is shut down. They're activated only when they're needed.

The two-node clusters for Central Services and the database services are stretched across two zones. If zone 1 fails, Central Services and the database services run in zone 2. The passive application servers in zone 2 get activated. With all components of this SAP system now co-located in the same zone, network latency is minimized.

### Active/active deployment example

In an [active/active](#) deployment, two sets of application servers are built across two zones. Within each zone, two application servers in each set of servers are inactive, because they're shut down. As a result, there are active application servers in both zones during normal operations.

Central Services and the database services run in zone 1. The application servers in zone 2 might have longer network latency when they connect to Central Services and the database services because of the physical distance between zones.

If zone 1 goes offline, Central Services and the database services fail over to zone 2. You can bring the dormant application servers online to provide full capacity for application processing.

## DR considerations

Every tier in the SAP application stack uses a different approach to provide DR protection. For DR strategies and implementation details, see [Disaster recovery overview](#) and [infrastructure guidelines for SAP workload](#) and [Disaster recovery guidelines for SAP application](#).

### Note

If there's a regional disaster that causes a large failover event for many Azure customers in one region, the target region's **resource capacity** isn't guaranteed. Like all Azure services, Site Recovery continues to add features and capabilities. For the latest information about Azure-to-Azure replication, see the [support matrix](#).

## Management and operations considerations

To help keep your system running in production, consider the following points.

### Azure Center for SAP solutions

Azure Center for SAP solutions is an end-to-end solution that enables you to create and run SAP systems as a unified workload on Azure and provides a more seamless foundation for innovation. Also, the Azure Center for SAP solutions guided deployment experience creates the necessary compute, storage, and networking components that you need to run your SAP system. It then helps you automate the installation of SAP software according to Microsoft best practices. You can take advantage of the management capabilities for both new and existing Azure-based SAP systems. For more information, see [Azure Center for SAP solutions](#).

If you need more control over maintenance events or hardware isolation, for either performance or compliance, consider deploying your VMs on [dedicated hosts](#).

### Backup

Databases are critical workloads that require a low recovery point objective (RPO) and long-term retention.

- For SAP on SQL Server, one approach is to use [Azure Backup](#) to back up SQL Server databases that run on VMs. Another option is to use [Azure Files snapshots](#) to back up SQL Server database files.
- For SAP on Oracle/Windows, see the "Backup/restore" section in [Azure VM DBMS Deployment for SAP](#).
- For other databases, see the backup recommendations for your database provider. If the database supports the Windows Volume Shadow Copy Service (VSS), use VSS snapshots for application-consistent backups.

### Identity management

Use a centralized identity management system like Microsoft Entra ID and Active Directory Domain Services (AD DS) to control access to resources at all levels:

- Provide access to Azure resources by using [Azure role-based access control \(Azure RBAC\)](#).
- Grant access to Azure VMs by using Lightweight Directory Access Protocol (LDAP), Microsoft Entra ID, Kerberos, or another system.

Support access within the applications themselves by using the services that SAP provides. Or use [OAuth 2.0 and Microsoft Entra ID](#).

## Monitoring

To maximize the availability and performance of applications and services on Azure, use [Azure Monitor](#), a comprehensive solution for collecting, analyzing, and acting on telemetry from your cloud and on-premises environments. Azure Monitor shows how applications are performing and proactively identifies problems that affect them and the resources that they depend on. For SAP applications that run on SAP HANA and other major database solutions, see [Azure Monitor for SAP solutions](#) to learn how Azure Monitor for SAP can help you manage the availability and performance of SAP services.

## Security considerations

SAP has its own user management engine (UME) to control role-based access and authorization within the SAP application and databases. For detailed application security guidance, see [SAP NetWeaver Security Guide](#).

For more network security, consider using a [perimeter network](#) that uses an NVA to create a firewall in front of the subnet for Web Dispatcher.

You can deploy an NVA to filter traffic between virtual networks, but don't place it between the SAP application and the database. Also, check the routing rules that are configured on the subnet and avoid directing traffic to a single-instance NVA. Doing so can lead to maintenance downtime and network or clustered-node failures.

For infrastructure security, data is encrypted in transit and at rest. For information about network security, see the "Security recommendations" section in [Azure Virtual Machines planning and implementation for SAP NetWeaver](#). This article also specifies the network ports that you need to open on the firewalls to allow application communication.

You can use [Azure Disk Encryption](#) to encrypt Windows VM disks. This service uses the BitLocker feature of Windows to provide volume encryption for the operating system

and the data disks. The solution also works with Azure Key Vault to help you control and manage the disk-encryption keys and secrets in your key vault subscription. Data on the VM disks is encrypted at rest in your Azure storage.

For data-at-rest encryption, SQL Server transparent data encryption (TDE) encrypts SQL Server, Azure SQL Database, and Azure Synapse Analytics data files. For more information, see [SQL Server Azure Virtual Machines DBMS deployment for SAP NetWeaver](#).

To monitor threats from inside and outside the firewall, consider deploying [Microsoft Sentinel \(preview\)](#). The solution provides continuous threat detection and analytics for SAP systems that are deployed on Azure, in other clouds, or on-premises. For deployment guidance, see [Deploy Threat Monitoring for SAP in Microsoft Sentinel](#).

As always, manage security updates and patches to safeguard your information assets. Consider using an end-to-end [automation approach](#) for this task.

## Cost considerations

Use the [Azure pricing calculator](#) to estimate costs.

For more information, see the cost section in [Microsoft Azure Well-Architected Framework](#).

If your workload requires more memory and fewer CPUs, consider using one of the [constrained vCPU VM](#) sizes to reduce software licensing costs that are charged per vCPU.

## VMs

This architecture uses VMs for the application tier and the database tier. The SAP NetWeaver tier uses Windows VMs to run SAP services and applications. The database tier runs AnyDB as the database, such as SQL Server, Oracle, or IBM DB2. VMs are also used as jump boxes for management.

There are several payment options for VMs:

- For workloads that have no predictable time of completion or resource consumption, consider the pay-as-you-go option.
- Consider using [Azure Reservations](#) if you can commit to using a VM over a one-year or three-year term. VM reservations can significantly reduce costs. You might pay as little as 72 percent of the cost of a pay-as-you-go service.

Use [Azure spot VMs](#) to run workloads that can be interrupted and don't require completion within a predetermined time frame or an SLA. Azure deploys spot VMs when there's available capacity and evicts them when it needs the capacity back. Costs that are associated with spot VMs are lower than for other VMs. Consider spot VMs for these workloads:

- High-performance computing scenarios, batch processing jobs, or visual rendering applications
- Test environments, including continuous integration and continuous delivery workloads
- Large-scale stateless applications

Azure Reserved Virtual Machine Instances can reduce your total cost of ownership by combining Azure Reserved Virtual Machine Instances rates with a pay-as-you-go subscription so you can manage costs across predictable and variable workloads. For more information, see [Azure Reserved Virtual Machine Instances](#).

## Load Balancer

In this scenario, Load Balancer is used to distribute traffic to VMs in the application-tier subnet.

You're charged only for the number of configured load-balancing and outbound rules, plus the data that's processed through the load balancer. Inbound network address translation (NAT) rules are free. There's no hourly charge for Standard Load Balancer when no rules are configured.

## ExpressRoute

In this architecture, ExpressRoute is the networking service that's used to create private connections between an on-premises network and Azure virtual networks.

All inbound data transfer is free. All outbound data transfer is charged based on a pre-determined rate. For more information, see [Azure ExpressRoute pricing](#).

## Communities

Communities can answer questions and help you set up a successful deployment. Consider these resources:

- [Running SAP Applications on the Microsoft Platform blog](#)
- [Azure Community Support](#)

- [SAP Community ↗](#)
- [Stack Overflow for SAP ↗](#)

## Contributors

*This article is maintained by Microsoft. It was originally written by the following contributor.*

Principal author:

- [Ben Trinh ↗](#) | Principal Architect

*To see non-public LinkedIn profiles, sign in to LinkedIn.*

## Next steps

For more information and for examples of SAP workloads that use some of the same technologies as this architecture, see these articles:

- [Azure Virtual Machines planning and implementation for SAP NetWeaver](#)
- [Use Azure to host and run SAP workload scenarios](#)

## Related resources

- [Run SAP production workloads using an Oracle Database on Azure](#)

# SAP S/4HANA in Linux on Azure

Azure ExpressRoute

SAP HANA on Azure Large Instances

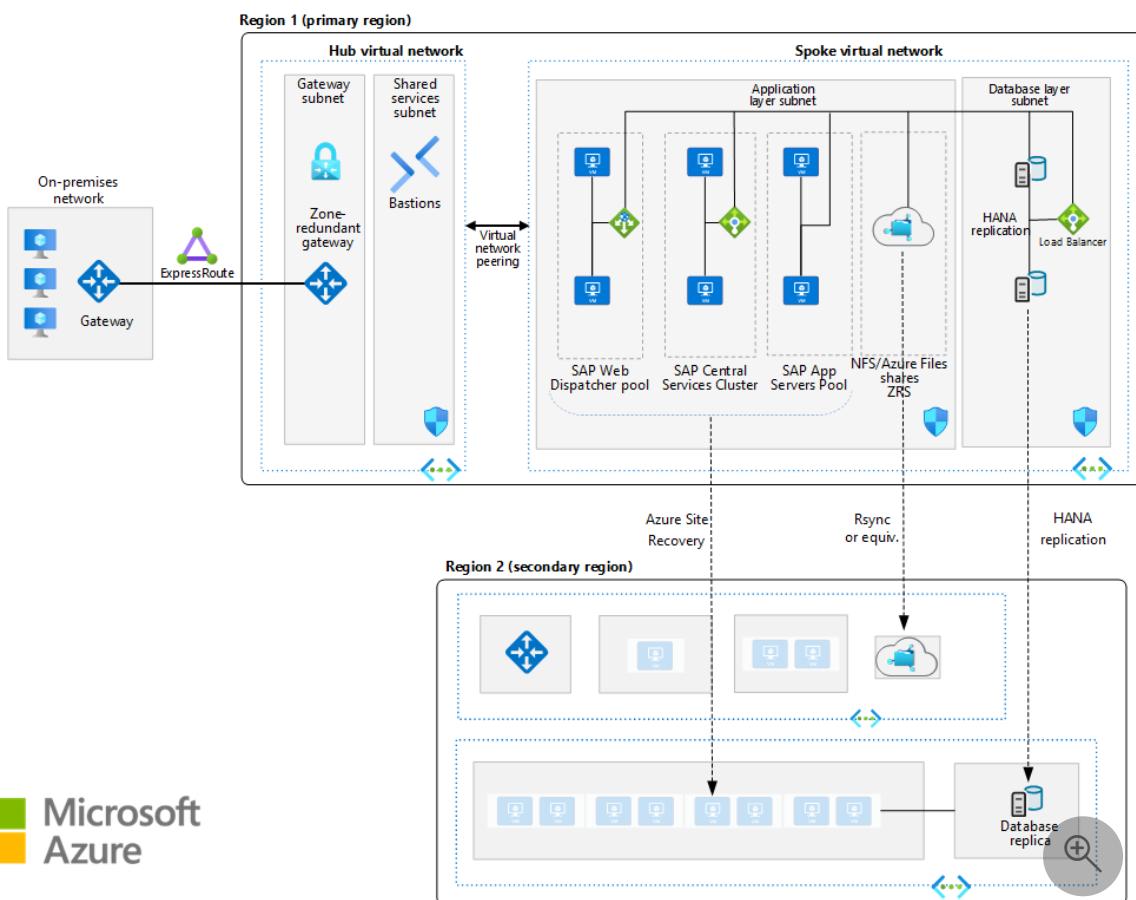
Azure Virtual Machines

Azure Virtual Network

Azure NetApp Files

This guide presents a set of proven practices for running S/4HANA and Suite on HANA in a high availability environment that supports disaster recovery (DR) on Azure. The Fiori information applies only to S/4HANA applications.

## Architecture



Download a [Visio file](#) of this architecture.

### ⓘ Note

Deploying this architecture requires appropriate licensing of SAP products and other non-Microsoft technologies.

This guide describes a common production system. This architecture is deployed with virtual machine (VM) sizes that you can change to accommodate the needs of your

organization. To suit your business needs, you can reduce this configuration to a single VM.

In this guide, the network layout is greatly simplified to demonstrate architectural principles. It's not intended to describe a full enterprise network.

The architecture uses the following components. Some shared services are optional.

**Azure Virtual Network.** The [Virtual Network](#) service securely connects Azure resources to each other. In this architecture, a virtual network connects to an on-premises environment through a gateway that's deployed in the hub of a [hub-spoke topology](#). The spoke is the virtual network that's used for the SAP applications and the database tiers.

**Virtual network peering.** This architecture uses multiple virtual networks that are [peered together](#). This topology offers network segmentation and isolation for services that are deployed on Azure. Peering connects networks transparently through the Microsoft backbone network and doesn't incur a performance penalty if implemented within a single region. Separate subnets are used for each tier application (SAP NetWeaver), database, and for shared services, such as the jump box and Windows Server Active Directory.

**VMs.** This architecture uses VMs that run Linux for the application tier and database tier, grouped in the following way:

- **Application tier.** This architectural layer includes the Fiori front-end server pool, the SAP Web Dispatcher pool, the application server pool, and the SAP Central Services cluster. For high availability of Central Services on Azure running in Linux VMs, a highly available network file share service is required, such as [NFS file shares in Azure Files](#), [Azure NetApp Files](#), clustered Network File System (NFS) servers, or [SIOS Protection Suite for Linux](#). To set up a highly available file share for the Central Services cluster on Red Hat Enterprise Linux (RHEL), you can configure [GlusterFS](#) on Azure VMs that run RHEL. On SUSE Linux Enterprise Server (SLES) 15 SP1 and later versions or SLES for SAP Applications, you can use [Azure shared disks](#) on a Pacemaker cluster to achieve high availability.
- **SAP HANA.** The database tier uses two or more Linux VMs in a cluster to achieve high availability in a scale-up deployment. HANA system replication (HSR) is used to replicate contents between primary and secondary HANA systems. Linux clustering is used to detect system failures and facilitate automatic failover. A storage-based or cloud-based fencing mechanism must be used to ensure that the failed system is isolated or shut down to avoid the cluster split-brain condition. In

HANA scale-out deployments, you can achieve database high availability by using one of the following options:

- Configure HANA standby nodes by using Azure NetApp Files without the Linux clustering component.
- Scale out without standby nodes by using Azure premium storage. Use Linux clustering for failover.
- **Azure Bastion.** This service lets you connect to a virtual machine by using your browser and the Azure portal, or via the native SSH or RDP client that's already installed on your local computer. If only RDP and SSH are used for administration, Azure Bastion is a great solution. If you use other management tools, like SQL Server Management Studio or an SAP front end, use a traditional self-deployed jump box.

**Private DNS service.** [Azure Private DNS](#) provides a reliable and secure DNS service for your virtual network. Azure Private DNS manages and resolves domain names in the virtual network without the need to configure a custom DNS solution.

**Load balancers.** To distribute traffic to VMs in the SAP application tier subnet for high availability, we recommend that you use [Azure Standard Load Balancer](#). Note that Standard Load Balancer provides a layer of security by default. VMs that are behind Standard Load Balancer don't have outbound internet connectivity. To enable outbound internet on the VMs, you need to update your [Standard Load Balancer configuration](#). You can also use [Azure NAT Gateway](#) to get outbound connectivity. For SAP web-based application high availability, use the built-in [SAP Web Dispatcher](#), or another commercially available load balancer. Base your selection on:

- Your traffic type, such as HTTP or SAP GUI.
- The network services that you need, such as Secure Sockets Layer (SSL) termination.

Standard Load Balancer supports multiple front-end virtual IPs. This support is ideal for cluster implementations that include these components:

- Advanced Business Application Programming (ABAP) SAP Central Service (ASCS)
- Enqueue Replication Server (ERS)

These two components can share a load balancer to simplify the solution.

Standard Load Balancer also supports multi-system identifier (multi-SID) SAP clusters. In other words, multiple SAP systems on [SLES](#) or [RHEL](#) can share a common high availability infrastructure to reduce costs. We recommend that you evaluate the cost

savings and avoid placing too many systems in one cluster. Azure supports no more than five SIDs per cluster.

**Application gateway.** Azure Application Gateway is a web traffic load balancer that you can use to manage the traffic to your web applications. Traditional load balancers operate at the transport layer (OSI layer 4 - TCP and UDP). They route traffic based on the source IP address and port to a destination IP address and port. Application Gateway can make routing decisions based on additional attributes of an HTTP request, such as the URI path or host headers. This type of routing is known as application layer (OSI layer 7) load balancing. S/4HANA offers web application services through Fiori. You can load balance this Fiori front end, which consists of web apps, by using Application Gateway.

**Gateway.** A gateway connects distinct networks and extends your on-premises network to an Azure virtual network. [Azure ExpressRoute](#) is the recommended Azure service for creating private connections that don't go over the public internet, but you can also use a [site-to-site](#) connection. To reduce latency, [ExpressRoute Global Reach](#) and [ExpressRoute FastPath](#) are connectivity options that are discussed later in this article.

**Zone-redundant gateway.** You can deploy ExpressRoute or virtual private network (VPN) gateways across zones to guard against zone failures. This architecture uses [zone-redundant](#) virtual network gateways for resiliency rather than a zonal deployment that's based on the same availability zone.

**Proximity placement group.** This logical group places a constraint on VMs that are deployed in an availability set or a virtual machine scale set. A [proximity placement group](#) favors co-location, which places VMs in the same datacenter to minimize application latency.

#### Note

The article [Configuration options for optimal network latency with SAP applications](#) contains an updated configuration strategy. You should read this article, especially if you intend to deploy an SAP system that has optimal network latency.

**Network security groups.** To restrict incoming, outgoing, and intra-subnet traffic in a virtual network, you can create [network security groups](#).

**Application security groups.** To define fine-grained network security policies that are based on workloads and centered on applications, use [application security groups](#)

instead of explicit IP addresses. You can group VMs by name and secure applications by filtering traffic from trusted segments of your network.

**Azure Storage.** Storage provides data persistence for a VM in the form of a virtual hard disk. We recommend [Azure managed disks](#).

## Recommendations

This architecture describes a small, production-level deployment. Deployments vary based on business requirements, so consider these recommendations as a starting point.

### VMs

In application server pools and clusters, adjust the number of VMs based on your requirements. For detailed information about running SAP NetWeaver on VMs, see [Azure Virtual Machines planning and implementation guide](#). The guide also applies to SAP S/4HANA deployments.

For details about SAP support for Azure VM types and for throughput metrics (SAPS), see [SAP Note 1928533](#). To access SAP notes, you need an SAP Service Marketplace account. For a list of certified Azure VMs for the HANA database, see [SAP Certified and Supported SAP HANA Hardware Directory](#).

### SAP Web Dispatcher

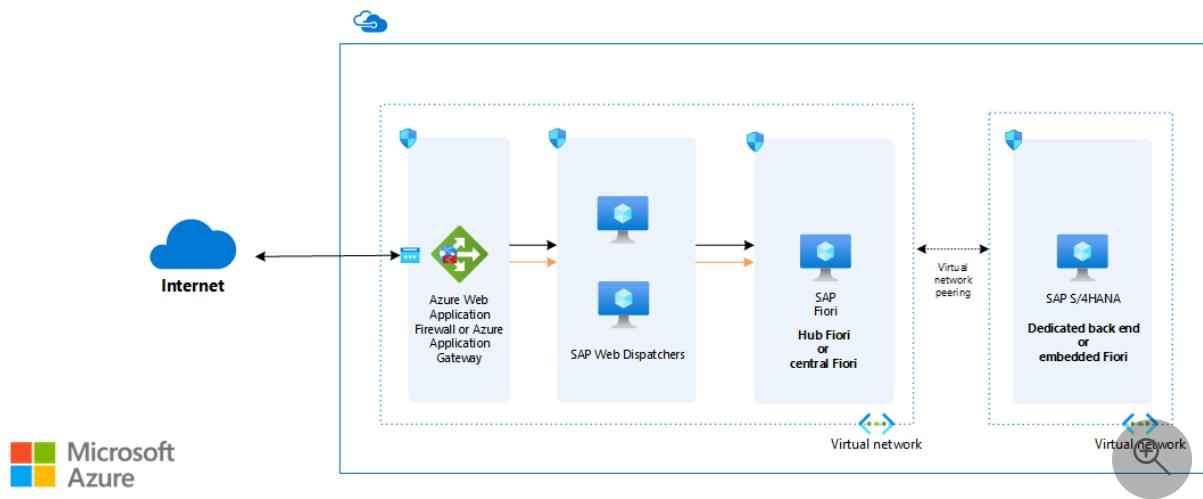
The Web Dispatcher component is used for load balancing SAP traffic among the SAP application servers. To achieve [high availability of SAP Web Dispatcher](#), Azure Load Balancer implements either a failover cluster or the parallel Web Dispatcher setup. For internet-facing communications, a stand-alone solution in the perimeter network is the recommended architecture to satisfy security concerns. [Embedded Web Dispatcher on ASCS](#) is a special option. If you use this option, consider proper sizing because of the extra workload on ASCS.

### Fiori front-end server (FES)

This architecture addresses many requirements and assumes that the embedded Fiori FES model is used. All the technology components are installed on the S/4 system itself, meaning that each S/4 system has its own Fiori launchpad. The high availability setup for this deployment model is that of the S/4 system—no extra clustering or VMs are required. For that reason, the architecture diagram doesn't show the FES component.

For a description of the primary deployment options—either embedded or hub, depending on the scenarios—see [SAP Fiori deployment options and system landscape recommendations](#). For simplicity and performance, the software releases between the Fiori technology components and the S/4 applications are tightly coupled. This setup makes a hub deployment that fits only a few, narrow use cases.

If you use the FES hub deployment, the FES is an add-on component to the classic SAP NetWeaver ABAP stack. Set up high availability the same way you protect a three-tier ABAP application stack that has clustered or multi-host capability: use a standby server database layer, a clustered ASCS layer with high availability NFS for shared storage, and at least two application servers. Traffic is load balanced via a pair of Web Dispatcher instances that can be either clustered or parallel. For internet-facing Fiori apps, we recommend an [FES hub deployment](#) in the perimeter network. Use [Azure Web Application Firewall on Application Gateway](#) as a critical component to deflect threats. Use [Microsoft Entra ID with SAML](#) for user authentication and SSO for [SAP Fiori](#).



For some internet-facing inbound/outbound design examples, see [Inbound and outbound internet connections for SAP on Azure](#).

## Application servers pool

To manage logon groups for ABAP application servers, it's common to use the SMLG transaction to load balance logon users, to use SM61 for batch server groups, to use RZ12 for remote function call (RFC) groups, and so on. These transactions use the load-balancing capability that's in the Central Services message server to distribute incoming sessions or workloads among the pool of SAP application servers that handle SAP GUIs and RFC traffic.

## SAP Central Services cluster

You can deploy Central Services to a single VM when the Azure single-instance VM availability service-level agreement (SLA) meets your requirement. However, the VM becomes a potential single point of failure (SPOF) for the SAP environment. For a highly available Central Services deployment, use either [NFS over Azure Files](#) or the [Azure NetApp Files service and a Central Services cluster](#).

Another option is to use [Azure shared disks](#) to achieve high availability. On SLES 15 SP1 and later or SLES for SAP Applications, you can set up a Pacemaker cluster by using [Azure shared disks for Linux](#).

Alternately, you can use an NFS file share for the [Linux cluster shared storage](#).

On an Azure deployment, the application servers connect to the highly available Central Services through the virtual host names of the Central Services or ERS services. These host names are assigned to the cluster front-end IP configuration of the load balancer. Load Balancer supports multiple frontend IPs, so both the Central Services and ERS virtual IPs (VIPs) can be configured to one load balancer.

Linux cluster support for [ASCS multi-SID installation](#) on Azure is now generally available. Sharing an availability cluster among multiple SAP systems simplifies the SAP landscape.

## Networking

This architecture uses a hub-spoke topology, where the hub virtual network acts as a central point of connectivity to an on-premises network. The spokes are virtual networks that peer with the hub. You can use the spokes to isolate workloads. Traffic flows between the on-premises datacenter and the hub through a gateway connection.

## Network interface cards (NICs)

Traditional on-premises SAP deployments implement multiple NICs per machine to segregate administrative traffic from business traffic. On Azure, the virtual network is a software-defined network that sends all traffic through the same network fabric. Therefore, the use of multiple NICs is unnecessary for performance considerations. However, if your organization needs to segregate traffic, you can deploy multiple NICs per VM, connect each NIC to a different subnet, and then use network security groups to enforce different access control policies.

Azure NICs support multiple IPs. This support aligns with the practice that SAP recommends of using virtual host names for installations, as outlined in [SAP note 962955](#). To access SAP notes, you need an SAP Service Marketplace account.

## Subnets and network security groups

This architecture divides the virtual network address space into subnets. You can associate each subnet with a network security group that defines the access policies for the subnet. Place application servers on a separate subnet. By doing so, you can secure them more easily by managing the subnet security policies rather than the individual servers.

When you associate a network security group with a subnet, the network security group applies to all the servers within the subnet and offers fine-grained control over the servers. Set up network security groups by using the [Azure portal](#), [PowerShell](#), or the [Azure CLI](#).

## ExpressRoute Global Reach

If your network environment includes two or more ExpressRoute circuits, [ExpressRoute Global Reach](#) can help you reduce network hops and lower latency. This technology is a Border Gateway Protocol (BGP) route peering that's set up between two or more ExpressRoute circuits to bridge two ExpressRoute routing domains. Global Reach lowers latency when network traffic traverses more than one ExpressRoute circuit. It's currently available only for private peering on ExpressRoute circuits.

Currently there are no network access control lists or other attributes that can be changed in Global Reach. So all routes learned by a given ExpressRoute circuit (from on-premises and Azure) are advertised across the circuit peering to the other ExpressRoute circuit. We recommend that you establish network traffic filtering on-premises to restrict access to resources.

## ExpressRoute FastPath

[FastPath](#) implements Microsoft Edge exchanges at the entry point of the Azure network. FastPath reduces network hops for most data packets. As a result, FastPath lowers network latency, improves application performance, and is the default configuration for new ExpressRoute connections to Azure.

For existing ExpressRoute circuits, contact Azure support to activate FastPath.

FastPath doesn't support virtual network peering. If other virtual networks are peered with one that's connected to ExpressRoute, the network traffic from your on-premises network to the other spoke virtual networks gets sent to the virtual network gateway. The workaround is to connect all virtual networks to the ExpressRoute circuit directly.

## Load balancers

[SAP Web Dispatcher](#) handles load balancing of HTTP(S) traffic to a pool of SAP application servers. This software load balancer offers application layer services (referred to as layer 7 in the ISO networking model) that are capable of SSL termination and other offloading functions.

[Load Balancer](#) is a network transmission layer service (layer 4) that balances traffic by using a five-tuple hash from data streams. The hash is based on source IP, source port, destination IP, destination port, and protocol type. Load Balancer is used in cluster setups to direct traffic to the primary service instance or the healthy node if there's a fault. We recommend that you use [Azure Standard Load Balancer](#) for all SAP scenarios. It's important to note that Standard Load Balancer is secure by default, and no VMs behind Standard Load Balancer have outbound internet connectivity. To enable outbound internet in the VMs, you must adjust your [Standard Load Balancer](#) configuration.

For traffic from SAP GUI clients that connect to an SAP server via the DIAG protocol or RFC, the Central Services message server balances the load through SAP application server [logon groups](#). No extra load balancer is needed.

## Storage

Some customers use standard storage for their application servers. Because standard managed disks aren't supported, as stated in SAP note 1928533, we recommend using premium [Azure managed disks](#) or [Azure NetApp Files](#) in all cases. A recent update to [SAP note 2015553](#) excludes the use of standard HDD storage and standard SSD storage for a few specific use cases.

Because application servers don't host any business data, you can also use the smaller P4 and P6 premium disks to help manage costs. To understand how the storage type affects the VM availability SLA, see [SLA for Virtual Machines](#). For high-availability scenarios, [Azure shared disk](#) features are available on Azure Premium SSD and Azure Ultra Disk Storage. For more information, see [Azure managed disks](#).

You can use Azure shared disks with Windows Server, SLES 15 SP 1 and later, or SLES for SAP. When you use an Azure shared disk in Linux clusters, the Azure shared disk serves as a STONITH block device (SBD). It offers a quorum vote in a cluster network partitioning situation. This shared disk doesn't have a file system and doesn't support simultaneous writes from multiple cluster member VMs.

Azure NetApp Files has built-in file sharing functionalities for NFS and SMB.

For NFS share scenarios, [Azure NetApp Files](#) provides availability for NFS shares that can be used for `/hana/shared`, `/hana/data`, and `/hana/log` volumes. For the availability guarantee, see [SLA for Azure NetApp Files](#). If you use Azure NetApp Files-based NFS shares for the `/hana/data` and `/hana/log` volumes, you need to use the NFS v4.1 protocol. For the `/hana/shared` volume, the NFS v3 protocol is supported.

For the backup data store, we recommend that you use Azure [cool and archive access tiers](#). These storage tiers are cost-effective ways to store long-lived data that's infrequently accessed. You can also consider using [Azure NetApp Files standard tier](#) as a backup target or [Azure NetApp Files backup option](#). For a managed disk, the recommended backup data tier is the Azure cool or archive access tier.

[Ultra Disk Storage](#) and Azure NetApp Files [ultra performance tier](#) greatly reduce disk latency and benefit performance-critical applications and the SAP database servers.

[Azure Premium SSD v2](#) is designed for performance-critical workloads like SAP. See [Deploy a Premium SSD v2](#) for information about this storage solution's benefits and its current limitations.

## Performance considerations

SAP application servers communicate constantly with the database servers. For performance-critical applications that run on any database platform, including SAP HANA, enable [Write Accelerator](#) for the log volume if you're using Premium SSD v1. Doing so can improve log-write latency. Premium SSD v2 doesn't use Write Accelerator. Write Accelerator is available for M-series VMs.

To optimize inter-server communications, use [Accelerated Networking](#). Most general-purpose and compute-optimized VM instance sizes that have two or more vCPUs support Accelerated Networking. On instances that support hyperthreading, VM instances with four or more vCPUs support Accelerated Networking.

For details about SAP HANA performance requirements, see [SAP note 1943937 - Hardware Configuration Check Tool](#). To access SAP notes, you need an SAP Service Marketplace account.

To achieve high IOPS and disk bandwidth throughput, the common practices in storage volume [performance optimization](#) apply to your Storage layout. For example, if you combine multiple disks to create a striped disk volume, you can improve IO performance. By enabling the read cache on storage content that changes infrequently, you can enhance the speed of data retrieval. For recommendations about storage

configurations for various VM sizes when you run SAP HANA, see [SAP HANA Azure virtual machine storage configurations](#).

Azure Premium SSD v2 is designed for performance-critical workloads like SAP. See [Azure managed disk types](#) to learn about its benefits and limitations and optimal usage scenarios.

[Ultra Disk Storage](#) is a new generation of storage that meets intensive IOPS and the transfer bandwidth demands of applications such as SAP HANA. You can dynamically change the performance of ultra disks and independently configure metrics like IOPS and MB/s without rebooting your VM. When Ultra Disk Storage is available, we recommend Ultra Disk Storage instead of Write Accelerator.

Some SAP applications require frequent communication with the database. Network latency between the application and database layers, due to distance, can adversely impact application performance. Azure [proximity placement groups](#) set a placement constraint for VMs that are deployed in availability sets. Within the logical construct of a group, co-location and performance are favored over scalability, availability, and cost. Proximity placement groups can greatly improve the user experience for most SAP applications. For scripts and utilities that are available on GitHub for proximity placement groups, see [Azure Proximity Placement Groups](#).

The placement of a network virtual appliance (NVA) between the application and the database layers of any SAP application stack isn't supported. The NVA requires a significant amount of time to process data packets. As a result, it unacceptably slows application performance.

We also recommend that you consider performance when you deploy resources with [availability zones](#), which can enhance service availability, as described later in this article. Consider creating a clear network latency profile between all zones of a target region. This approach helps you decide on the resource placement for minimum latency between zones. To create this profile, run a test by deploying small VMs in each zone. Recommended tools for the test include [PsPing](#) and [Iperf](#). After testing, remove these VMs. For a public domain network latency test tool that you can use instead, see [Availability Zone Latency Test](#).

Azure NetApp Files has unique performance features that make real-time tuning possible that meets the needs of the most demanding SAP environments. For performance considerations to keep in mind when you use Azure NetApp Files, see [Sizing for HANA database on Azure NetApp Files](#).

## Scalability considerations

At the SAP application layer, Azure offers a wide range of VM sizes for scaling up and scaling out. For an inclusive list, see "SAP Applications on Azure: Supported Products and Azure VM types" in [SAP Note 1928533](#). To access SAP notes, you need an SAP Service Marketplace account. More VM types are continually being certified, so you can scale up or down in the same cloud deployment.

On the database layer, this architecture runs SAP HANA S/4 applications on Azure VMs that can scale up to 24 terabytes (TB) in one instance. If your workload exceeds the maximum VM size, you can use a multi-node configuration for as much as 96 TBs (4 x 24) for online transaction processing (OLTP) applications. For more information, see [Certified and Supported SAP HANA Hardware Directory](#).

## Availability considerations

Resource redundancy is the general theme in highly available infrastructure solutions. For single-instance VM availability SLAs for various storage types, see [SLA for Virtual Machines](#). To increase service availability on Azure, deploy VM resources with Virtual Machine Scale Sets with flexible orchestration, availability zones, or availability sets.

In this distributed installation of the SAP application, the base installation is replicated to achieve high availability. For each layer of the architecture, the high availability design varies.

## Deployment approaches

On Azure, SAP workload deployment can be either regional or zonal, depending on the availability and resiliency requirements of the SAP applications. Azure provides [different deployment options](#), like Virtual Machine Scale Sets with Flexible orchestration (FD=1), availability zones, and availability sets, to enhance the availability of the resources. To get a comprehensive understanding of the available deployment options and their applicability across different Azure regions (including across zones, within a single zone, or in a region without zones), see [High-availability architecture and scenarios for SAP NetWeaver](#).

## Web Dispatcher in the application servers tier

You can achieve high availability by using redundant Web Dispatcher instances. For more information, see [SAP Web Dispatcher](#) in the SAP documentation. The availability level depends on the size of the application that's behind Web Dispatcher. In small deployments with few scalability concerns, you can co-locate Web Dispatcher with the

ASCS VMs. This way, you save on independent OS maintenance and gain high availability at the same time.

## Central Services in the application servers tier

For high availability of Central Services on Azure Linux VMs, use the appropriate high availability extension for the selected Linux distribution. It's customary to place the shared file systems on highly available NFS storage by using SUSE DRBD or Red Hat GlusterFS. To provide a highly available NFS and eliminate the need for an NFS cluster, you can use other cost-effective or robust solutions like [NFS over Azure Files](#) or [Azure NetApp Files](#) instead. As a side note, Azure NetApp Files shares can host the SAP HANA data and log files. This setup enables the [HANA scale-out](#) deployment model with standby nodes, while NFS over Azure Files is good for highly available non-database file sharing.

NFS over Azure Files now supports the highly available file shares for both [SLES](#) and [RHEL](#). This solution works well for highly available file shares like those of `/sapmnt`, `/saptrans` in SAP installations.

Azure NetApp Files supports high availability of [ASCS on SLES](#). For detailed information about ASCS on RHEL high availability, see [SIOS Protection Suite for Linux](#).

The improved Azure Fence Agent is available for both [SUSE](#) and [Red Hat](#) and provides significantly faster service failover than the previous version of the agent.

Another option is to use [Azure shared disks](#) to achieve high availability. On SLES 15 SP 1 and later or SLES for SAP, you can set up a Pacemaker cluster by using [Azure shared disks](#) to achieve high availability.

On Azure Standard Load Balancer, you can enable the [high availability port](#) and avoid the need to configure load balancing rules for many SAP ports. In general, if you enable the direct server return (DSR) feature when you set up a load balancer, server responses to client inquiries can bypass the load balancer. This feature is also known as Floating IP. The load balancer can be on-premises or on Azure. This direct connection keeps the load balancer from becoming the bottleneck in the path of data transmission. For the ASCS and HANA DB clusters, we recommend that you enable DSR. If VMs in the back-end pool require public outbound connectivity, more [configuration](#) is required.

For traffic from SAP GUI clients that connect to an SAP server via DIAG protocol or RFC, the Central Services message server balances the load by using SAP application server [logon groups](#). No extra load balancer is needed.

# Application servers in the application servers tier

You can achieve high availability by load balancing traffic within a pool of application servers.

## ASCS tier

As with the application servers layer, the commonly deployed HANA high availability solution for Linux is Pacemaker.

## Database tier

The architecture in this guide depicts a highly available SAP HANA database system that consists of two Azure VMs. The native system replication feature of the database tier provides either manual or automatic failover between replicated nodes:

- For manual failover, deploy more than one HANA instance and use HSR.
- For automatic failover, use both HSR and Linux high availability extension (HAE) for your Linux distribution. Linux HAE provides the cluster services to the HANA resources, detecting failure events and orchestrating the failover of errant services to the healthy node.

## Deploy VMs across availability zones

[Availability zones](#) can enhance service availability. Zones refer to physically separated locations within a specific Azure region. They improve workload availability and protect application services and VMs against datacenter outages. VMs in a single zone are treated as if they were in a single update or fault domain. When zonal deployment is selected, VMs in the same zone are distributed to fault and upgrade domains on a best-effort basis.

In [Azure regions](#) that support this feature, at least three zones are available. However, the maximum distance between datacenters in these zones isn't guaranteed. To deploy a multi-tier SAP system across zones, you must know the network latency within a zone and across targeted zones, and how sensitive your deployed applications are to network latency.

Take these [considerations](#) into account when you decide to deploy resources across availability zones:

- Latency between VMs in one zone
- Latency between VMs across chosen zones

- Availability of the same Azure services (VM types) in the chosen zones

#### Note

We don't recommend availability zones for disaster recovery. A disaster recovery site should be at least 100 miles from the primary site, in case of a natural disaster. There is no certainty of the distance between the datacenters.

### Active/passive deployment example

In this example deployment, the [active/passive](#) status refers to the application service state within the zones. In the application layer, all four active application servers of the SAP system are in zone 1. Another set of four passive application servers is built in zone 2 but is shut down. They're activated only when they're needed.

The two-node clusters for Central Services and the database are stretched across two zones. If zone 1 fails, Central Services and database services run in zone 2. The passive application servers in zone 2 get activated. With all components of this SAP system co-located in the same zone, network latency is minimized.

### Active/active deployment example

In an [active/active](#) deployment, two sets of application servers are built across two zones. In each zone, two application servers in each set are inactive, or shut down. As a result, there are active application servers in both zones in normal operations.

The ASCS and database services run in zone 1. The application servers in zone 2 might have longer network latency when they connect to the ASCS and database services due to the physical distance between zones.

If zone 1 goes offline, the ASCS and database services fail over to zone 2. The dormant application servers can be brought online to provide full capacity for application processing.

## DR considerations

Every tier in the SAP application stack uses a different approach to provide DR protection. For DR strategies and implementation details, see [Disaster recovery overview and infrastructure guidelines for SAP workload](#) and [Disaster recovery guidelines for SAP application](#).

#### Note

If there's a regional disaster that causes a mass failover event for many Azure customers in one region, the target region's **resource capacity** isn't guaranteed. Like all Azure services, Site Recovery continues to add features and capabilities. For the latest information about Azure-to-Azure replication, see the [support matrix](#).

## Cost considerations

Use the [Azure pricing calculator](#) to estimate costs.

For more information, see the cost section in [Microsoft Azure Well-Architected Framework](#).

## VMs

This architecture uses VMs that run Linux for the management, SAP application, and database tiers.

There are several payment options for VMs in general:

- For workloads with no predictable time of completion or resource consumption, consider the pay-as-you-go option.
- Consider using [Azure Reservations](#) if you can commit to using a VM over a one-year or three-year term. VM reservations can significantly reduce costs. You might pay as little as 72 percent of the cost of a pay-as-you-go service.
- Use [Azure spot VMs](#) to run workloads that can be interrupted and don't require completion within a predetermined time-frame or SLA. Azure deploys spot VMs when there's available capacity and evicts them when it needs the capacity back. Costs that are associated with spot VMs are lower than for other VMs. Consider spot VMs for these workloads:
  - High-performance computing scenarios, batch processing jobs, or visual rendering applications
  - Test environments, including continuous integration and continuous delivery workloads
  - Large-scale stateless applications
- Azure Reserved Virtual Machine Instances can reduce your total cost of ownership by combining Azure Reserved Virtual Machine Instances rates with a pay-as-you-go subscription so you can manage costs across predictable and variable workloads. For more information about this offering, see [Azure Reserved Virtual Machine Instances](#).

For an overview of pricing, see [Linux Virtual Machines Pricing](#).

## Load Balancer

In this scenario, Azure load balancers are used to distribute traffic to VMs in the application tier subnet.

You're charged only for the number of configured load-balancing and outbound rules. Inbound network address translation (NAT) rules are free. There's no hourly charge for Standard Load Balancer when no rules are configured.

## ExpressRoute

In this architecture, ExpressRoute is the networking service that's used for creating private connections between an on-premises network and Azure virtual networks.

All inbound data transfer is free. All outbound data transfer is charged based on a pre-determined rate. For more information, see [Azure ExpressRoute pricing](#).

## Management and operations considerations

To help keep your system running in production, consider the following points.

### Azure Center for SAP solutions

Azure Center for SAP solutions is an end-to-end solution that enables you to create and run SAP systems as a unified workload on Azure and provides a more seamless foundation for innovation. Also, the Azure Center for SAP solutions guided deployment experience creates the necessary compute, storage, and networking components that you need to run your SAP system. It then helps you automate the installation of the SAP software according to Microsoft best practices. You can take advantage of the management capabilities for both new and existing Azure-based SAP systems. For more information, see [Azure Center for SAP solutions](#).

## Backup

You can back up SAP HANA data in many ways. After you migrate to Azure, continue using any existing backup solutions that you have. Azure provides two native approaches to backup. You can back up [SAP HANA on VMs or use Azure Backup at the file level](#). Azure Backup is [BackInt certified](#) by SAP. For more information, see [Azure Backup FAQ](#) and [Support matrix for backup of SAP HANA databases on Azure VMs](#).

## Note

Currently only HANA single-container or scale-up deployments support Azure storage snapshots.

## Identity management

Use a centralized identity management system to control access to resources at all levels:

- Provide access to Azure resources through [Azure role-based access control \(Azure RBAC\)](#).
- Grant access to Azure VMs through Lightweight Directory Access Protocol (LDAP), Microsoft Entra ID, Kerberos, or another system.
- Support access within the apps themselves through the services that SAP provides, or use [OAuth 2.0 and Microsoft Entra ID](#).

## Monitoring

To maximize the availability and performance of applications and services on Azure, use [Azure Monitor](#), a comprehensive solution for collecting, analyzing, and acting on telemetry from your cloud and on-premises environments. Azure Monitor shows how applications are performing and proactively identifies issues that affect them and the resources that they depend on. For SAP applications that run on SAP HANA and other major database solutions, see [Azure Monitor for SAP solutions](#) to learn how Azure Monitor for SAP can help you manage SAP services availability and performance.

## Security considerations

SAP has its own Users Management Engine (UME) to control role-based access and authorization within the SAP application and databases. For details, see [SAP HANA Security: An Overview](#).

To improve network security, consider using a [perimeter network](#) that uses an NVA to create a firewall in front of the subnet for Web Dispatcher and the Fiori front-end server pools. The cost of data transfer is a reason to place active front-end servers that run Fiori apps in the same virtual network as the S/4 systems. The alternative is to place them in the perimeter network and connect them to S/4 through a virtual network peering.

For infrastructure security, data is encrypted in transit and at rest. The "Security considerations" section of [SAP NetWeaver on Azure Virtual Machines—Planning and Implementation Guide](#) contains information on network security that applies to S/4HANA. That guide also specifies the network ports to open on the firewalls to allow application communication.

To encrypt Linux VM disks, you have various choices, as described in [Disk encryption overview](#). For SAP HANA data-at-rest encryption, we recommend that you use the SAP HANA native encryption technology. For support of Azure disk encryption on specific Linux distributions, versions, and images, see [Azure disk encryption for Linux VMs](#).

For SAP HANA data-at-rest encryption, we recommend that you use the SAP HANA native encryption technology.

#### Note

Don't use the HANA data-at-rest encryption and Azure disk encryption on the same storage volume. For HANA, use only HANA data encryption. Also, the use of customer managed keys might affect I/O throughput.

## Communities

Communities can answer questions and help you set up a successful deployment.

Consider these resources:

- [Running SAP Applications on the Microsoft Platform Blog](#)
- [Azure Community Support ↗](#)
- [SAP Community ↗](#)
- [Stack Overflow SAP ↗](#)

## Contributors

*This article is maintained by Microsoft. It was originally written by the following contributor.*

Principal author:

- [Ben Trinh ↗](#) | Principal Architect

*To see non-public LinkedIn profiles, sign in to LinkedIn.*

# Next steps

For more information and for examples of SAP workloads that use some of the same technologies as this architecture, see these articles:

- [Deploy SAP S/4HANA or BW/4HANA on Azure](#)
- [Azure Virtual Machines planning and implementation for SAP NetWeaver](#)
- [Use Azure to host and run SAP workload scenarios](#)

## Related resources

- [Run SAP production workloads using an Oracle Database on Azure](#)

# SAP NetWeaver on SQL Server

Microsoft Entra ID

Azure ExpressRoute

Azure Storage

Azure Virtual Machines

SQL Server

## 💡 Solution ideas

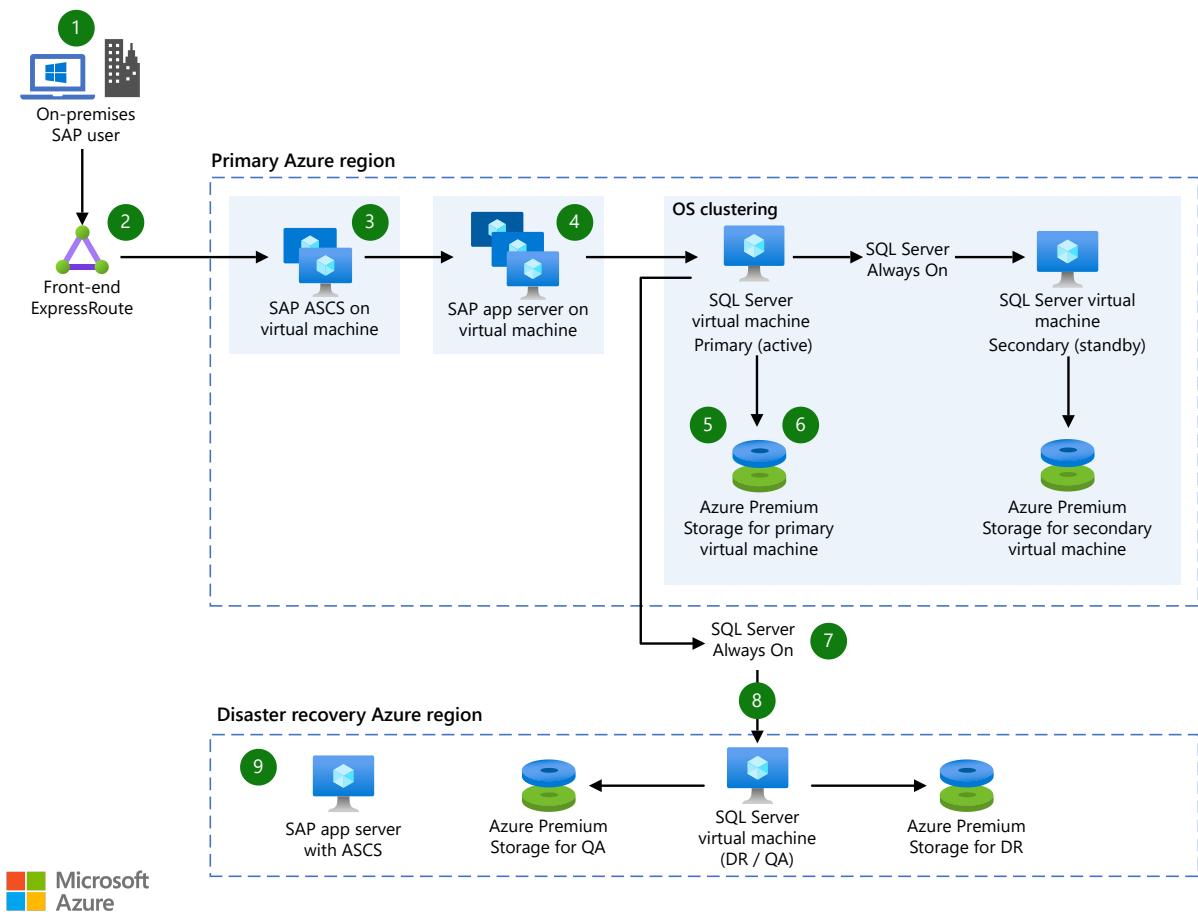
This article is a solution idea. If you'd like us to expand the content with more information, such as potential use cases, alternative services, implementation considerations, or pricing guidance, let us know by providing [GitHub feedback](#).

This NetWeaver on SQL Server application solution illustrates how a user request flows through an SAP landscape that's built on NetWeaver, by using Azure Virtual Machines to host SAP applications and a SQL Server database.

## Potential use cases

This system takes advantage of OS clustering for high availability, premium storage for faster storage performance and scalability, SQL Server Always On capability for replication, and a full disaster recovery (DR) configuration for 99.95 percent system availability.

## Architecture



Download a [Visio file](#) of this architecture.

## Dataflow

1. By using Azure Active Directory synchronized with on-premises Active Directory, SAP application user authenticates from on-premises to SAP landscape on Azure with single sign-on credentials.
2. Azure high-speed ExpressRoute Gateway connects on-premises network to Azure virtual machines and other resources securely.
3. Sales order request flows into highly available SAP ABAP SAP Central Services (ASCS), and then through SAP application servers running on Azure Virtual Machines scale out file server in an Azure VM.
4. The request moves from the SAP app server to SQL Server running on a primary high-performance Azure VM.
5. Primary (active) and secondary (standby) servers running on SAP certified virtual machines are clustered at OS level for 99.95 percent availability. Data replication is handled through SQL Server Always On in synchronous mode from primary to secondary, enabling zero Recovery Point Objective (RPO).
6. SQL Server data is persisted to high-performance Azure Premium Storage.
7. SQL Server data is replicated Disaster recovery virtual machine in another Azure region through Azure's high speed backbone network and using SQL Server's

Always On replication in asynchronous mode. The disaster recovery VM can be smaller than the production VM to save costs.

8. VMs on the disaster recovery region can be used for nonproduction work to save costs.
9. SAP app server with ASCS on disaster recovery side can be in standby shutdown mode, and can be started when needed to save costs.

## Components

- Information on [Virtual Machines](#) for SAP application servers.
- Microsoft Azure [Premium Storage](#) provides improved throughput and less variability in I/O latencies. For improved performance, [Premium Storage](#) uses solid state disk (SSD) in Azure Storage nodes, and read cache that's backed by the local SSD of an Azure compute node.

## Next steps

- [SAP Certifications for Azure](#)
- [Premium Storage: high-performance storage for Azure Virtual Machine workloads](#)

# SAP S/4 HANA for Large Instances

Azure ExpressRoute

Azure Files

SAP HANA on Azure Large Instances

Azure Virtual Machines

## 💡 Solution ideas

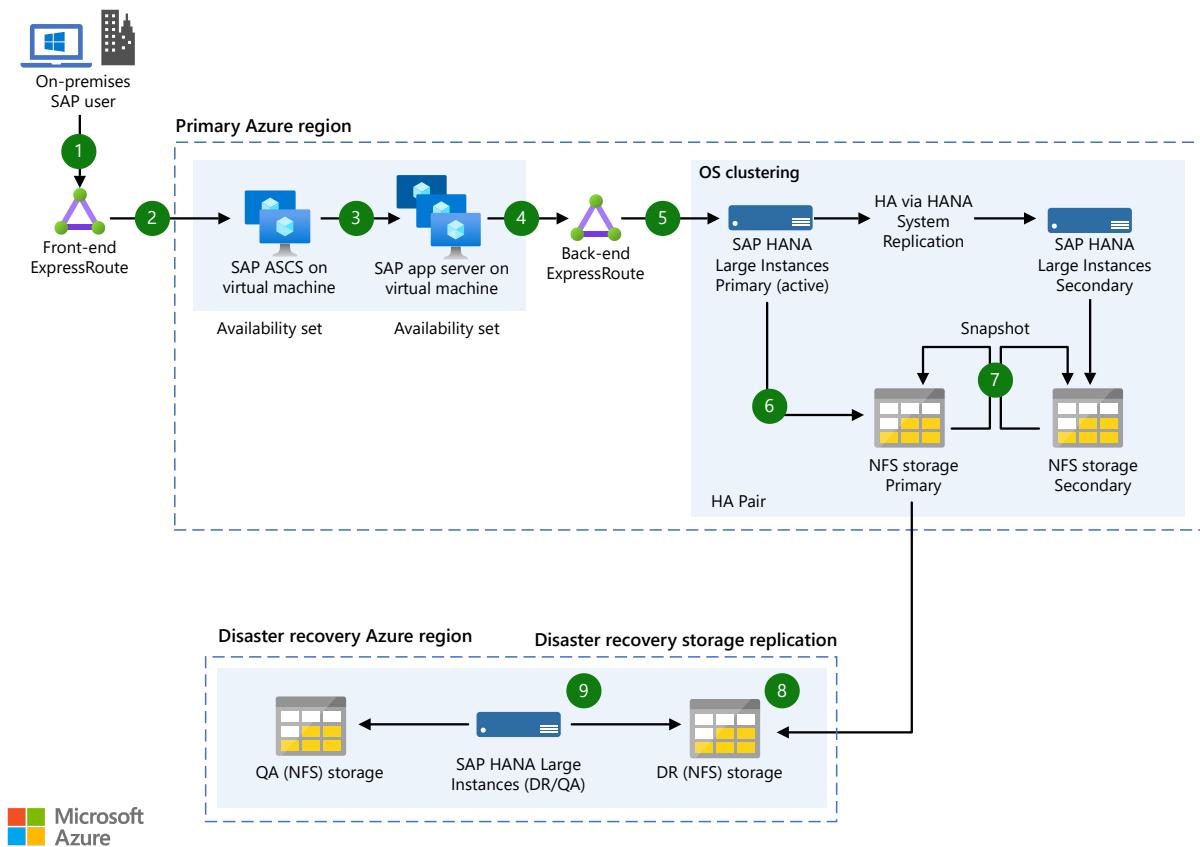
This article is a solution idea. If you'd like us to expand the content with more information, such as potential use cases, alternative services, implementation considerations, or pricing guidance, let us know by providing [GitHub feedback](#).

This solution architecture illustrates how a user request flows through an SAP landscape that's built on high-performance Azure Virtual Machines and an in-memory HANA database, which runs on HANA Large Instances for unparalleled scalability and performance.

## Potential use cases

This system takes advantage of OS clustering for database performance, high availability using HANA system replication, and a full disaster recovery (DR) configuration for guaranteed system availability.

## Architecture



Download a [Visio file](#) of this architecture.

## Dataflow

- In this example, an on-premises SAP user executes a sales order via Fiori interface, custom interface, or other.
- Azure high-speed ExpressRoute gateway is used to connect to Azure Virtual Machines.
- Request flows into highly available ABAP SAP Central Services (ASCS) and then through application servers, which run on Azure Virtual Machines. This availability set offers a 99.95 percent uptime SLA.
- Request is sent from App Server to SAP HANA running on primary large instance blades.
- Primary and secondary blades are clustered at OS level for 99.99 percent availability, and data replication is handled through HANA System Replication in synchronous mode (HSR) from primary to secondary enabling zero RPO.
- In-memory data of SAP HANA is persisted to high-performance NFS storage.
- Data from NFS storage is periodically backed up in seconds, using built-in storage snapshots on the local storage, with no impact to database performance.
- Persistent data volume on secondary storage is replicated to dedicated DR system through a dedicated backbone network for HANA storage replication.

9. Large instance on DR side can be used for nonproduction to save costs by mounting both the QA storage and DR replicated volume (read-only).

## Components

- [SAP HANA on Azure Large Instances](#): SAP HANA on Azure (Large Instances) runs on dedicated blade servers located in a Microsoft Azure datacenter. This feature is specific to the database server.
- [NFS storage for Azure HANA large instances](#): The Azure high-performance NFS storage system offers the unmatched capability to perform snapshot backups, and replication to secondary storage. In addition, HANA Large Instances are the only cloud infrastructure to provide storage volume encryption.
- SAP on Azure requires that you run your SAP workloads on certified Microsoft Azure [Virtual Machines](#). SAP requires at least two vCPUs and a ratio of 6:1 between memory and vCPU.
- Microsoft Azure [Premium Storage](#) provides improved throughput and less variability in I/O latencies. For improved performance, [Premium Storage](#) uses solid-state disk (SSD) in Azure Storage nodes and read cache that's backed by the local SSD of an Azure compute node.
- [ExpressRoute \(front end\)](#): Azure ExpressRoute used on the front end (see diagram) provides secure, high-bandwidth connectivity to establish reliable connections between your network and the Microsoft Azure network.
- [ExpressRoute \(back end\)](#): Azure ExpressRoute used on the back end (see diagram) enables you to communicate between your Azure components in the Azure Datacenter and your SAP HANA on Azure (Large Instances) systems. The cost of the back-end ExpressRoute is included in your SAP HANA on Azure (Large Instances).

## Next steps

- [Getting started](#)
- [High-performance NFS storage for SAP HANA Large Instances](#)
- [SAP Certifications for Azure](#)
- [Premium Storage: high-performance storage for Azure Virtual Machine workloads](#)
- [ExpressRoute overview](#)
- [Back end Network to HANA Large Instances](#)

# SAP workload automation using SUSE on Azure

Azure Files

Azure Cloud Shell

Azure Load Balancer

Azure Virtual Network

## 💡 Solution ideas

This article is a solution idea. If you'd like us to expand the content with more information, such as potential use cases, alternative services, implementation considerations, or pricing guidance, let us know by providing [GitHub feedback](#).

This architecture demonstrates how to use the SUSE SAP automation solution on Azure.

Since 2009, SUSE and Microsoft have partnered to provide Azure-optimized solutions for SUSE Linux Enterprise Server (SLES). SLES for SAP Applications is the leading platform for SAP solutions on Linux, with over 90 percent of SAP HANA deployments and 70 percent of SAP NetWeaver applications running on SUSE.

Automating SAP workloads in the cloud leads to better business outcomes by bolstering productivity and facilitating innovation. The task of building and manually deploying SAP infrastructures in the cloud involves a range of technical processes that can be inefficient and time-consuming. These processes also require configuration management and entail many steps. With each step, the level of complexity and the amount of specialized knowledge that's required increases if extra high availability (HA) is needed. Most SAP systems are important and require HA. The manual implementation of each step generates many opportunities for error that can render the entire infrastructure defective and delay success. Automation helps organizations streamline deployment of SAP infrastructure and accelerate customer cloud migration on Azure. Successful cloud migration allows customers to quickly and easily benefit from the power and flexibility of the cloud.

The SUSE SAP solution excels at simplifying and modernizing SAP HANA and SAP NetWeaver deployments. You can configure it to set up and monitor both environments.

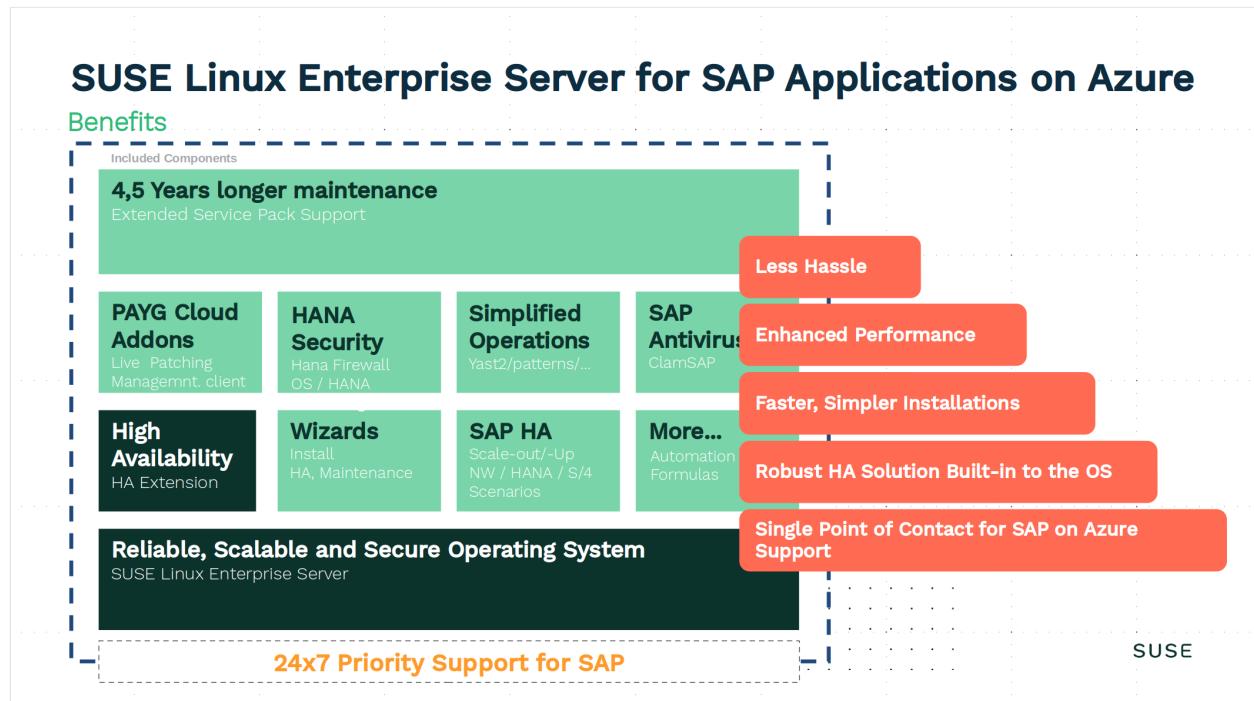
## Potential use cases

SLES4SAP is a bundle of software and services that addresses specific needs of SAP users, and delivers services faster, more efficiently, and with less risk. You can deploy SAP HANA and SAP NetWeaver applications in many different scenarios and

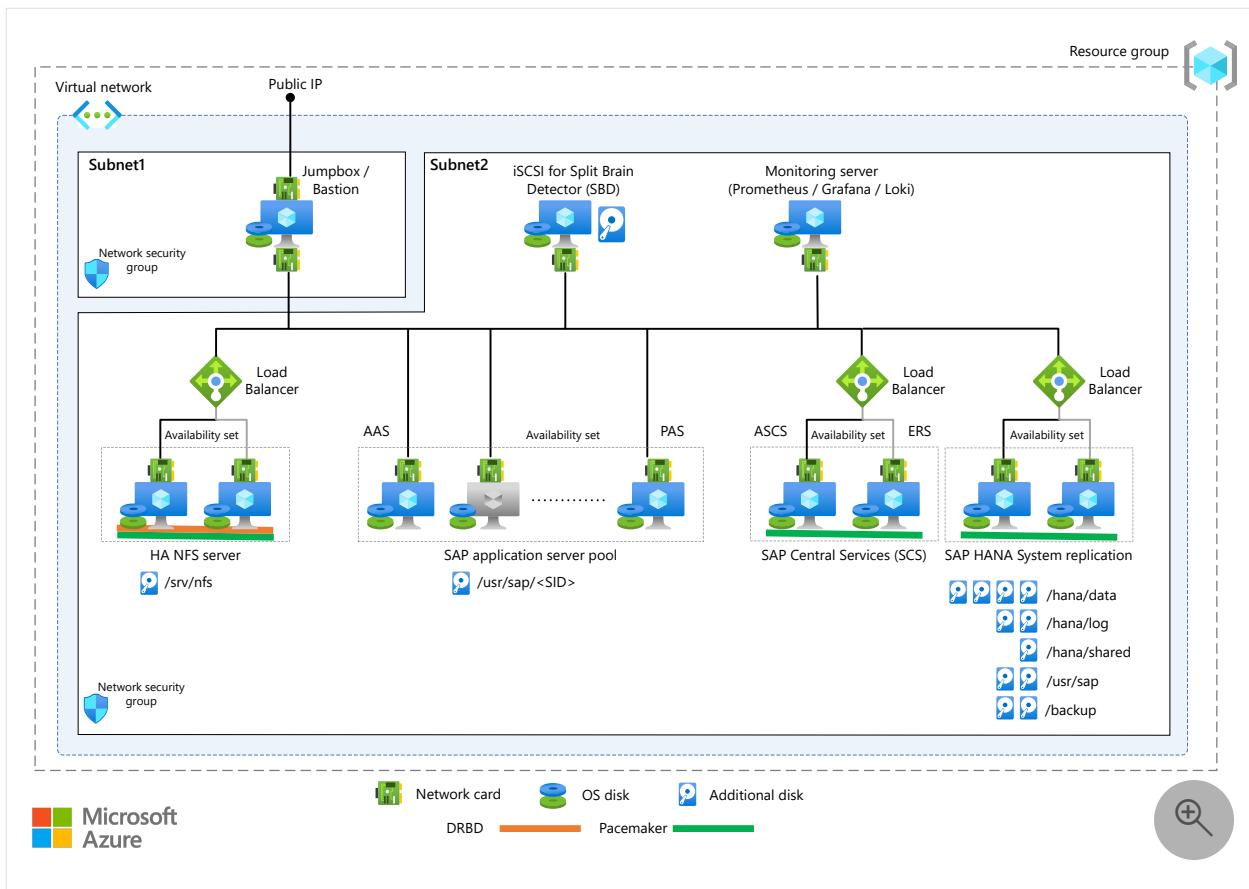
combinations. The SUSE SAP solution features modular and reusable building blocks to support use cases ranging from single install to full cluster deployment.

SUSE provides support with:

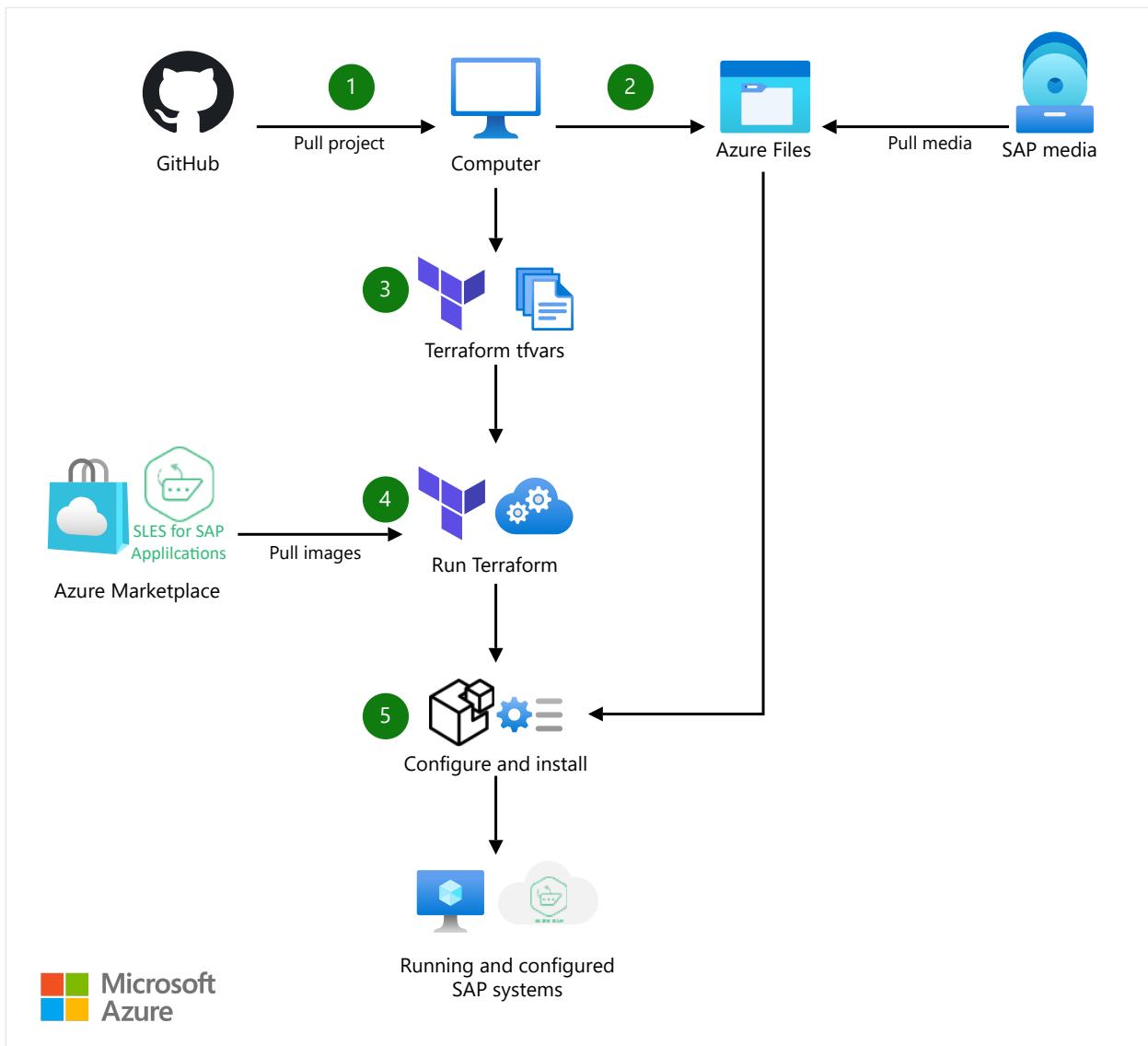
- HANA single node
- HANA HA Scale-up System replication, including performance-optimized (active/passive and active/readonly) and cost-optimized scenarios
- NetWeaver
- NetWeaver HA with Enqueue Replication Version (ENSA1)
- S/4 HANA



## Architecture



## Dataflow



Download a [Visio file](#) of diagrams in this article.

1. Download the SUSE automation git repository to your local machine or Azure Cloud Shell and install the needed Terraform version, which comes with SLES4SAP or Cloud Shell.
2. Create an Azure File Share instance and download SAP media to it.
3. Tailor the example parameters—ssh-keys, network, SID, file-share, and so on—to your needs and values.
4. Run Terraform to deploy the SAP infrastructure into Azure.
  - a. Terraform creates the infrastructure, including resource groups, networks, virtual machines, disks, availability groups, load balancers, and so on.
  - b. Terraform starts configuration with Salt.
5. Salt performs the needed OS configuration:
  - a. It installs SAP applications on the nodes.
  - b. It installs and configures clusters if HA.
  - c. It installs and configures the monitoring parts such as Prometheus, Grafana and exporters.

# Components

- [Azure Storage](#) is a set of massively scalable and secure cloud services for data, apps, and workloads. It includes [Azure Files](#), [Azure Table Storage](#), and [Azure Queue Storage](#). Azure Files is often an effective tool for migrating mainframe workloads.
- [Azure Files](#) offers simple, secure, and serverless enterprise-grade file shares in the cloud. The shares support access by the industry-standard Server Message Block (SMB) and Network File System (NFS) protocols. They can be mounted concurrently by cloud and on-premises deployments of Windows, Linux, and macOS.
- [Azure Load Balancer](#) is a layer 4 (TCP, UDP) load balancer. In this architecture, it provides load balancing options for Spring Apps and AKS.
- [Linux virtual machines in Azure](#) are on-demand, scalable Linux computing resources that give you the flexibility of virtualization but eliminate the maintenance demands of physical hardware.
- [Azure Virtual Network](#) is a secure private network in the cloud. It connects VMs to one another, to the internet, and to on-premises networks.

# Contributors

*This article is maintained by Microsoft. It was originally written by the following contributors.*

Principal author:

- [Michael Yen-Chi Ho](#) | Senior Program Manager

# Next Steps

## SAP

- [SAP on Azure Architecture Guide](#)
- [SAP workloads on Azure: planning and deployment checklist](#)
- [SAP workload configurations with Azure Availability Zones](#)
- [SAP S/4HANA in Linux on Azure](#)
- [SAP S/4 HANA for Large Instances](#)
- [Use Azure to host and run SAP workload scenarios](#)
- [SAP workloads on Azure: planning and deployment checklist](#)
- [SUSE SAP automation solution for Azure](#) (GitHub PDF document)

- [Automated SAP/HA Deployments in Public/Private Clouds with Terraform](#) ↗  
(GitHub project)
- [Deploying SUSE SAP HA Automation in Microsoft Azure](#) ↗ (Microsoft blog)

## Azure services

- [Azure premium storage: design for high performance](#)
- [Plan virtual networks](#)
- [What is Azure Load Balancer?](#)

## SUSE

- [SUSE on Azure Marketplace](#) ↗
- [Highly Available NFS Storage with DRBD and Pacemaker](#) ↗
- [Run SAP](#) ↗
- [SUSE Linux Enterprise Server for SAP Applications 15 SP3](#) ↗
- [SUSE Best Practices - all documents](#) ↗
- [Getting Started with SAP HANA High Availability Cluster Automation Operating on Azure](#) ↗
- [Monitor SAP in SLES with Grafana and Prometheus](#) ↗ (video)
- [Set up and tune your SUSE system for SAP with saptune](#) ↗
- [SAP S/4HANA in Linux on Azure](#)

## Solution templates

SUSE SAP ARM template to create the SAP infrastructure:

- [Infrastructure for SAP NetWeaver and SAP HANA](#) ↗ (Azure Marketplace)
- [SUSE and Microsoft Solution Templates for SAP Applications](#) ↗ (SUSE)
- [SUSE and Microsoft Solution templates for SAP Applications](#) ↗ (GitHub)

# Security architecture design

Microsoft Entra ID   Azure Firewall   Azure Front Door   Azure Key Vault   Azure Private Link

Information security has always been a complex subject, and it evolves quickly with the creative ideas and implementations of attackers and security researchers.

Security is one of the most important aspects of any architecture. Good security provides confidentiality, integrity, and availability assurances against deliberate attacks and abuse of your valuable data and systems. Losing these assurances can harm your business operations and revenue, and your organization's reputation.

## ⓘ Note

Learn how cloud security is an ongoing journey of incremental progress and maturity, in [Security in the Microsoft Cloud Adoption Framework for Azure](#). Learn how to build security into your solution, in the Azure Well-Architected Framework [Overview of the security pillar](#).

Here are some broad categories to consider when you design a security system:



Azure provides a wide range of security tools and capabilities. These are just some of the key security services available in Azure:

- [Microsoft Defender for Cloud](#). A unified infrastructure security management system that strengthens the security posture of your datacenters. It also provides advanced threat protection across your hybrid workloads in the cloud and on-premises.
- [Microsoft Entra ID](#). The Microsoft cloud-based identity and access management service.
- [Azure Front Door](#). A global, scalable entry-point that uses the Microsoft global edge network to create fast, highly secure, and widely scalable web applications.
- [Azure Firewall](#). A cloud-native, intelligent network firewall security service that provides threat protection for your cloud workloads that run in Azure.
- [Azure Key Vault](#). A high-security secret store for tokens, passwords, certificates, API keys, and other secrets. You can also use Key Vault to create and control the encryption keys used to encrypt your data.
- [Azure Private Link](#). A service that enables you to access Azure PaaS services, Azure-hosted services that you own, or partner services over a private endpoint in your virtual network.
- [Azure Application Gateway](#). An advanced web traffic load balancer that enables you to manage traffic to your web applications.
- [Azure Policy](#). A service that helps you enforce organizational standards and assess compliance.

For a more comprehensive description of Azure security tools and capabilities, see [End-to-end security in Azure](#).

## Introduction to security on Azure

If you're new to security on Azure, the best way to learn more is with [Microsoft Learn training](#). This free online platform provides interactive training for Microsoft products and more.

Here are two learning paths to get you started:

- [Microsoft Azure Fundamentals: Describe general security and network security features](#)
- [Microsoft Security, Compliance, and Identity Fundamentals: Describe the capabilities of Microsoft security solutions](#)

## Path to production

- To secure Azure application workloads, you use protective measures like authentication and encryption in the applications themselves. You can also add security layers to the virtual machine (VM) networks that host the applications. See [Firewall and Application Gateway for virtual networks](#) for an overview.
- Zero Trust is a proactive, integrated approach to security across all layers of the digital estate. It explicitly and continuously verifies every transaction, asserts least privilege, and relies on intelligence, advanced detection, and real-time response to threats.
  - For an implementation strategy for web apps, see [Zero Trust network for web applications with Azure Firewall and Application Gateway](#).
  - For an architecture that shows how to incorporate Microsoft Entra identity and access capabilities into an overall Zero Trust security strategy, see [Microsoft Entra IDaaS in security operations](#).
- Azure governance establishes the tooling needed to support cloud governance, compliance auditing, and automated guardrails. See [Azure governance design area guidance](#) for information about governing your Azure environment.

## Best practices

The Azure Well-Architected Framework is a set of guiding tenets, based on five pillars, that you can use to improve the quality of your architectures. For information, see [Overview of the security pillar](#) and [Security design principles in Azure](#).

The Well-Architected Framework also provides these checklists:

- Azure identity and access management considerations
- Network security
- Data protection considerations
- Governance, risk, and compliance

For information about security for sensitive IaaS workloads, see [Security considerations for highly sensitive IaaS apps in Azure](#).

## Security architectures

### Identity and access management

- [Secure OAuth 2.0 On-Behalf-Of refresh tokens for web services](#)
- [Resilient identity and access management with Microsoft Entra ID](#)
- [Microsoft Entra identity management and access management for AWS](#)

## Threat protection

- Threat indicators for cyber threat intelligence in Microsoft Sentinel
- Multilayered protection for Azure virtual machine access
- Real-time fraud detection

## Information protection

- Confidential computing on a healthcare platform
- Homomorphic encryption with SEAL
- SQL Managed Instance with customer-managed keys
- Virtual network integrated serverless microservices

## Discover and respond

- Long-term security log retention with Azure Data Explorer

## Stay current with security

Get the latest updates on [Azure security services and features](#).

## Additional resources

### Example solutions

- Hybrid Security Monitoring using Microsoft Defender for Cloud and Microsoft Sentinel
- Improved-security access to multitenant web apps from an on-premises network
- Restrict interservice communications
- Securely managed web applications
- Secure your Microsoft Teams channel bot and web app behind a firewall
- Web app private connectivity to Azure SQL database

[Browse all our security architectures.](#)

## AWS or Google Cloud professionals

- Security and identity with Azure and AWS
- AWS to Azure services comparison - Security
- Google Cloud to Azure services comparison - Security

# Next steps

Security architecture is part of a comprehensive set of security guidance that also includes:

- [Security in the Microsoft Cloud Adoption Framework for Azure](#): A high-level overview of a cloud security end state.
- [Azure Well-Architected Framework](#): Guidance on securing your workloads on Azure.
- [Azure security benchmarks](#): Prescriptive best practices and controls for Azure security.
- [End-to-end security in Azure](#): Documentation that introduces you to the security services in Azure.
- [Top 10 security best practices for Azure](#): Top Azure security best practices that Microsoft recommends based on lessons learned across customers and our own environments.
- [Microsoft Cybersecurity Architectures](#): The diagrams describe how Microsoft security capabilities integrate with Microsoft platforms and 3rd-party platforms.

# Introduction to Azure security

Article • 10/22/2023

## Overview

We know that security is job one in the cloud and how important it is that you find accurate and timely information about Azure security. One of the best reasons to use Azure for your applications and services is to take advantage of its wide array of security tools and capabilities. These tools and capabilities help make it possible to create secure solutions on the secure Azure platform. Microsoft Azure provides confidentiality, integrity, and availability of customer data, while also enabling transparent accountability.

This article provides a comprehensive look at the security available with Azure.

## Azure platform

Azure is a public cloud service platform that supports a broad selection of operating systems, programming languages, frameworks, tools, databases, and devices. It can run Linux containers with Docker integration; build apps with JavaScript, Python, .NET, PHP, Java, and Node.js; build back-ends for iOS, Android, and Windows devices.

Azure public cloud services support the same technologies millions of developers and IT professionals already rely on and trust. When you build on, or migrate IT assets to, a public cloud service provider you are relying on that organization's abilities to protect your applications and data with the services and the controls they provide to manage the security of your cloud-based assets.

Azure's infrastructure is designed from facility to applications for hosting millions of customers simultaneously, and it provides a trustworthy foundation upon which businesses can meet their security requirements.

In addition, Azure provides you with a wide array of configurable security options and the ability to control them so that you can customize security to meet the unique requirements of your organization's deployments. This document helps you understand how Azure security capabilities can help you fulfill these requirements.

### Note

The primary focus of this document is on customer-facing controls that you can use to customize and increase security for your applications and services.

For information on how Microsoft secures the Azure platform itself, see [Azure infrastructure security](#).

## Summary of Azure security capabilities

Depending on the cloud service model, there is variable responsibility for who is responsible for managing the security of the application or service. There are capabilities available in the Azure Platform to assist you in meeting these responsibilities through built-in features, and through partner solutions that can be deployed into an Azure subscription.

The built-in capabilities are organized in six functional areas: Operations, Applications, Storage, Networking, Compute, and Identity. Additional detail on the features and capabilities available in the Azure Platform in these six areas are provided through summary information.

## Operations

This section provides additional information regarding key features in security operations and summary information about these capabilities.

### Microsoft Sentinel

[Microsoft Sentinel](#) is a scalable, cloud-native, security information and event management (SIEM) and security orchestration, automation, and response (SOAR) solution. Microsoft Sentinel delivers intelligent security analytics and threat intelligence across the enterprise, providing a single solution for attack detection, threat visibility, proactive hunting, and threat response.

### Microsoft Defender for Cloud

[Microsoft Defender for Cloud](#) helps you prevent, detect, and respond to threats with increased visibility into and control over the security of your Azure resources. It provides integrated security monitoring and policy management across your Azure subscriptions, helps detect threats that might otherwise go unnoticed, and works with a broad ecosystem of security solutions.

In addition, Defender for Cloud helps with security operations by providing you a single dashboard that surfaces alerts and recommendations that can be acted upon immediately. Often, you can remediate issues with a single click within the Defender for Cloud console.

## Azure Resource Manager

[Azure Resource Manager](#) enables you to work with the resources in your solution as a group. You can deploy, update, or delete all the resources for your solution in a single, coordinated operation. You use an [Azure Resource Manager template](#) for deployment and that template can work for different environments such as testing, staging, and production. Resource Manager provides security, auditing, and tagging features to help you manage your resources after deployment.

Azure Resource Manager template-based deployments help improve the security of solutions deployed in Azure because standard security control settings and can be integrated into standardized template-based deployments. This reduces the risk of security configuration errors that might take place during manual deployments.

## Application Insights

[Application Insights](#) is an extensible Application Performance Management (APM) service for web developers. With Application Insights, you can monitor your live web applications and automatically detect performance anomalies. It includes powerful analytics tools to help you diagnose issues and to understand what users actually do with your apps. It monitors your application all the time it's running, both during testing and after you've published or deployed it.

Application Insights creates charts and tables that show you, for example, what times of day you get most users, how responsive the app is, and how well it is served by any external services that it depends on.

If there are crashes, failures or performance issues, you can search through the telemetry data in detail to diagnose the cause. And the service sends you emails if there are any changes in the availability and performance of your app. Application Insight thus becomes a valuable security tool because it helps with the availability in the confidentiality, integrity, and availability security triad.

## Azure Monitor

Azure Monitor offers visualization, query, routing, alerting, auto scale, and automation on data both from the Azure subscription ([Activity Log](#)) and each individual Azure resource ([Resource Logs](#)). You can use Azure Monitor to alert you on security-related events that are generated in Azure logs.

## Azure Monitor logs

[Azure Monitor logs](#) – Provides an IT management solution for both on-premises and third-party cloud-based infrastructure (such as AWS) in addition to Azure resources. Data from Azure Monitor can be routed directly to Azure Monitor logs so you can see metrics and logs for your entire environment in one place.

Azure Monitor logs can be a useful tool in forensic and other security analysis, as the tool enables you to quickly search through large amounts of security-related entries with a flexible query approach. In addition, on-premises [firewall and proxy logs can be exported into Azure and made available for analysis using Azure Monitor logs](#).

## Azure Advisor

[Azure Advisor](#) is a personalized cloud consultant that helps you to optimize your Azure deployments. It analyzes your resource configuration and usage telemetry. It then recommends solutions to help improve the [performance](#), [security](#), and [reliability](#) of your resources while looking for opportunities to [reduce your overall Azure spend](#). Azure Advisor provides security recommendations, which can significantly improve your overall security posture for solutions you deploy in Azure. These recommendations are drawn from security analysis performed by [Microsoft Defender for Cloud](#).

## Applications

The section provides additional information regarding key features in application security and summary information about these capabilities.

## Penetration Testing

We don't perform [penetration testing](#) of your application for you, but we do understand that you want and need to perform testing on your own applications. That's a good thing, because when you enhance the security of your applications you help make the entire Azure ecosystem more secure. While notifying Microsoft of pen testing activities is no longer required customers must still comply with the [Microsoft Cloud Penetration Testing Rules of Engagement](#).

## Web Application firewall

The web application firewall (WAF) in [Azure Application Gateway](#) helps protect web applications from common web-based attacks like SQL injection, cross-site scripting attacks, and session hijacking. It comes preconfigured with protection from threats identified by the [Open Web Application Security Project \(OWASP\)](#) as the top 10 common vulnerabilities ↴.

## Authentication and authorization in Azure App Service

[App Service Authentication / Authorization](#) is a feature that provides a way for your application to sign in users so that you don't have to change code on the app backend. It provides an easy way to protect your application and work with per-user data.

## Layered Security Architecture

Since [App Service Environments](#) provide an isolated runtime environment deployed into an [Azure Virtual Network](#), developers can create a layered security architecture providing differing levels of network access for each application tier. A common desire is to hide API back-ends from general Internet access, and only allow APIs to be called by upstream web apps. [Network Security groups \(NSGs\)](#) can be used on Azure Virtual Network subnets containing App Service Environments to restrict public access to API applications.

## Web server diagnostics and application diagnostics

[App Service web apps](#) provide diagnostic functionality for logging information from both the web server and the web application. These are logically separated into web server diagnostics and application diagnostics. Web server includes two major advances in diagnosing and troubleshooting sites and applications.

The first new feature is real-time state information about application pools, worker processes, sites, application domains, and running requests. The second new advantages are the detailed trace events that track a request throughout the complete request-and-response process.

To enable the collection of these trace events, IIS 7 can be configured to automatically capture full trace logs, in XML format, for any particular request based on elapsed time or error response codes.

# Storage

The section provides additional information regarding key features in Azure storage security and summary information about these capabilities.

## Azure role-based access control (Azure RBAC)

You can secure your storage account with [Azure role-based access control \(Azure RBAC\)](#). Restricting access based on the [need to know](#) and [least privilege](#) security principles is imperative for organizations that want to enforce Security policies for data access. These access rights are granted by assigning the appropriate Azure role to groups and applications at a certain scope. You can use [Azure built-in roles](#), such as Storage Account Contributor, to assign privileges to users. Access to the storage keys for a storage account using the [Azure Resource Manager](#) model can be controlled through Azure RBAC.

## Shared Access Signature

A [shared access signature \(SAS\)](#) provides delegated access to resources in your storage account. The SAS means that you can grant a client limited permissions to objects in your storage account for a specified period and with a specified set of permissions. You can grant these limited permissions without having to share your account access keys.

## Encryption in Transit

Encryption in transit is a mechanism of protecting data when it is transmitted across networks. With Azure Storage, you can secure data using:

- [Transport-level encryption](#), such as HTTPS when you transfer data into or out of Azure Storage.
- [Wire encryption](#), such as [SMB 3.0 encryption](#) for [Azure File shares](#).
- Client-side encryption, to encrypt the data before it is transferred into storage and to decrypt the data after it is transferred out of storage.

## Encryption at rest

For many organizations, data encryption at rest is a mandatory step towards data privacy, compliance, and data sovereignty. There are three Azure storage security features that provide encryption of data that is “at rest”:

- [Storage Service Encryption](#) allows you to request that the storage service automatically encrypt data when writing it to Azure Storage.
- [Client-side Encryption](#) also provides the feature of encryption at rest.
- [Azure Disk Encryption for Linux VMs](#) and [Azure Disk Encryption for Windows VMs](#) allows you to encrypt the OS disks and data disks used by an IaaS virtual machine.

## Storage Analytics

[Azure Storage Analytics](#) performs logging and provides metrics data for a storage account. You can use this data to trace requests, analyze usage trends, and diagnose issues with your storage account. Storage Analytics logs detailed information about successful and failed requests to a storage service. This information can be used to monitor individual requests and to diagnose issues with a storage service. Requests are logged on a best-effort basis. The following types of authenticated requests are logged:

- Successful requests.
- Failed requests, including timeout, throttling, network, authorization, and other errors.
- Requests using a Shared Access Signature (SAS), including failed and successful requests.
- Requests to analytics data.

## Enabling Browser-Based Clients Using CORS

[Cross-Origin Resource Sharing \(CORS\)](#) is a mechanism that allows domains to give each other permission for accessing each other's resources. The User Agent sends extra headers to ensure that the JavaScript code loaded from a certain domain is allowed to access resources located at another domain. The latter domain then replies with extra headers allowing or denying the original domain access to its resources.

Azure storage services now support CORS so that once you set the CORS rules for the service, a properly authenticated request made against the service from a different domain is evaluated to determine whether it is allowed according to the rules you have specified.

## Networking

The section provides additional information regarding key features in Azure network security and summary information about these capabilities.

# Network Layer Controls

Network access control is the act of limiting connectivity to and from specific devices or subnets and represents the core of network security. The goal of network access control is to make sure that your virtual machines and services are accessible to only users and devices to which you want them accessible.

## Network Security Groups

A [Network Security Group \(NSG\)](#) is a basic stateful packet filtering firewall and it enables you to control access based on a 5-tuple. NSGs do not provide application layer inspection or authenticated access controls. They can be used to control traffic moving between subnets within an Azure Virtual Network and traffic between an Azure Virtual Network and the Internet.

## Azure Firewall

[Azure Firewall](#) is a cloud-native and intelligent network firewall security service that provides threat protection for your cloud workloads running in Azure. It's a fully stateful firewall as a service with built-in high availability and unrestricted cloud scalability. It provides both east-west and north-south traffic inspection.

Azure Firewall is offered in two SKUs: Standard and Premium. [Azure Firewall Standard](#) provides L3-L7 filtering and threat intelligence feeds directly from Microsoft Cyber Security. [Azure Firewall Premium](#) provides advanced capabilities include signature-based IDPS to allow rapid detection of attacks by looking for specific patterns.

## Route Control and Forced Tunneling

The ability to control routing behavior on your Azure Virtual Networks is a critical network security and access control capability. For example, if you want to make sure that all traffic to and from your Azure Virtual Network goes through that virtual security appliance, you need to be able to control and customize routing behavior. You can do this by configuring User-Defined Routes in Azure.

[User-Defined Routes](#) allow you to customize inbound and outbound paths for traffic moving into and out of individual virtual machines or subnets to ensure the most secure route possible. [Forced tunneling](#) is a mechanism you can use to ensure that your services are not allowed to initiate a connection to devices on the Internet.

This is different from being able to accept incoming connections and then responding to them. Front-end web servers need to respond to requests from Internet hosts, and so

Internet-sourced traffic is allowed inbound to these web servers and the web servers can respond.

Forced tunneling is commonly used to force outbound traffic to the Internet to go through on-premises security proxies and firewalls.

## Virtual Network Security Appliances

While Network Security Groups, User-Defined Routes, and forced tunneling provide you a level of security at the network and transport layers of the [OSI model](#), there may be times when you want to enable security at higher levels of the stack. You can access these enhanced network security features by using an Azure partner network security appliance solution. You can find the most current Azure partner network security solutions by visiting the [Azure Marketplace](#) and searching for “security” and “network security.”

## Azure Virtual Network

An Azure virtual network (VNet) is a representation of your own network in the cloud. It is a logical isolation of the Azure network fabric dedicated to your subscription. You can fully control the IP address blocks, DNS settings, security policies, and route tables within this network. You can segment your VNet into subnets and place Azure IaaS virtual machines (VMs) and/or [Cloud services \(PaaS role instances\)](#) on Azure Virtual Networks.

Additionally, you can connect the virtual network to your on-premises network using one of the [connectivity options](#) available in Azure. In essence, you can expand your network to Azure, with complete control on IP address blocks with the benefit of enterprise scale Azure provides.

Azure networking supports various secure remote access scenarios. Some of these include:

- Connect individual workstations to an Azure Virtual Network
- Connect on-premises network to an Azure Virtual Network with a VPN
- Connect on-premises network to an Azure Virtual Network with a dedicated WAN link
- Connect Azure Virtual Networks to each other

## Azure Private Link

Azure Private Link [↗](#) enables you to access Azure PaaS Services (for example, Azure Storage and SQL Database) and Azure hosted customer-owned/partner services privately in your virtual network over a [private endpoint](#). Setup and consumption using Azure Private Link is consistent across Azure PaaS, customer-owned, and shared partner services. Traffic from your virtual network to the Azure service always remains on the Microsoft Azure backbone network.

[Private Endpoints](#) allow you to secure your critical Azure service resources to only your virtual networks. Azure Private Endpoint uses a private IP address from your VNet to connect you privately and securely to a service powered by Azure Private Link, effectively bringing the service into your VNet. Exposing your virtual network to the public internet is no longer necessary to consume services on Azure.

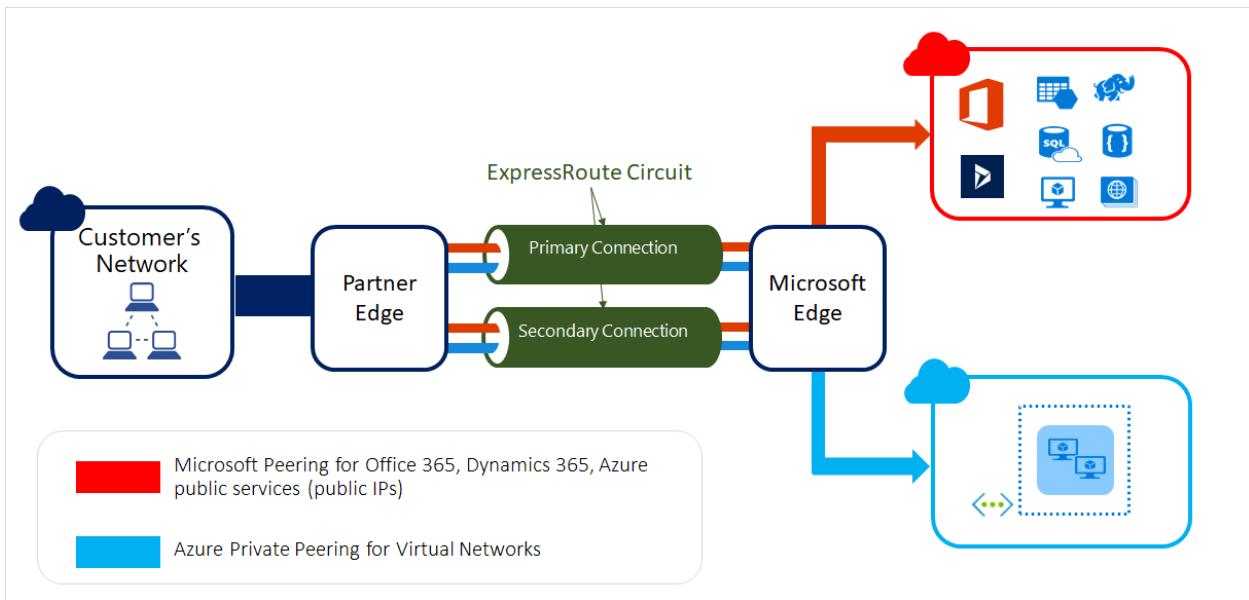
You can also create your own private link service in your virtual network. [Azure Private Link service](#) is the reference to your own service that is powered by Azure Private Link. Your service that is running behind Azure Standard Load Balancer can be enabled for Private Link access so that consumers to your service can access it privately from their own virtual networks. Your customers can create a private endpoint inside their virtual network and map it to this service. Exposing your service to the public internet is no longer necessary to render services on Azure.

## VPN Gateway

To send network traffic between your Azure Virtual Network and your on-premises site, you must create a VPN gateway for your Azure Virtual Network. A [VPN gateway](#) is a type of virtual network gateway that sends encrypted traffic across a public connection. You can also use VPN gateways to send traffic between Azure Virtual Networks over the Azure network fabric.

## Express Route

Microsoft Azure [ExpressRoute](#) is a dedicated WAN link that lets you extend your on-premises networks into the Microsoft cloud over a dedicated private connection facilitated by a connectivity provider.

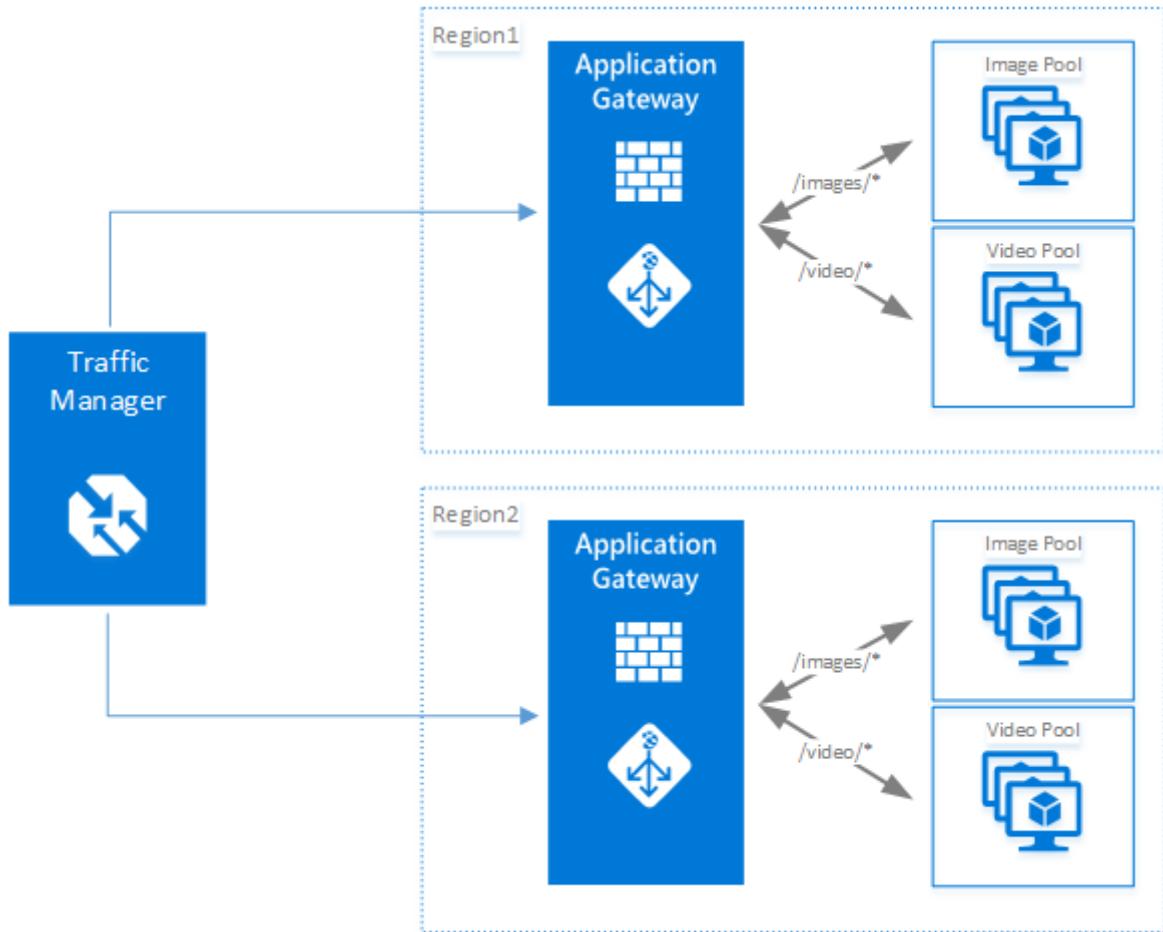


With ExpressRoute, you can establish connections to Microsoft cloud services, such as Microsoft Azure, Microsoft 365, and CRM Online. Connectivity can be from an any-to-any (IP VPN) network, a point-to-point Ethernet network, or a virtual cross-connection through a connectivity provider at a co-location facility.

ExpressRoute connections do not go over the public Internet and thus can be considered more secure than VPN-based solutions. This allows ExpressRoute connections to offer more reliability, faster speeds, lower latencies, and higher security than typical connections over the Internet.

## Application Gateway

Microsoft [Azure Application Gateway](#) provides an [Application Delivery Controller \(ADC\)](#) as a service, offering various layer 7 load balancing capabilities for your application.



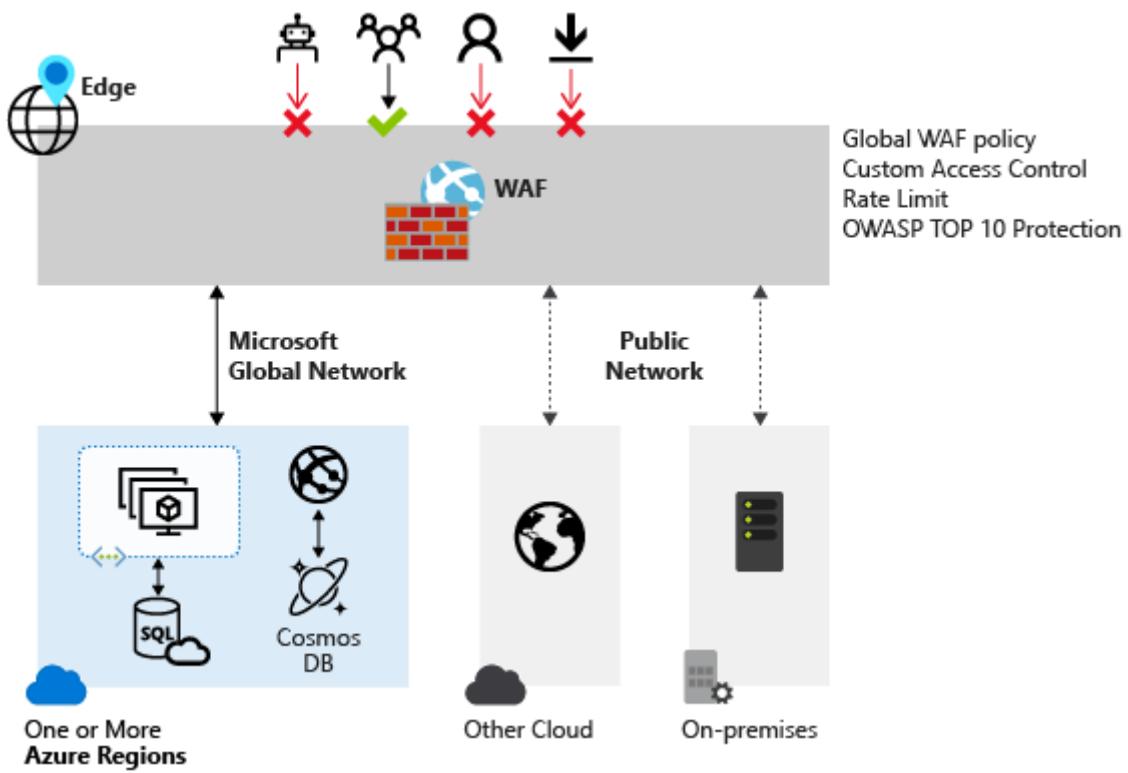
It allows you to optimize web farm productivity by offloading CPU intensive TLS termination to the Application Gateway (also known as "TLS offload" or "TLS bridging"). It also provides other Layer 7 routing capabilities including round-robin distribution of incoming traffic, cookie-based session affinity, URL path-based routing, and the ability to host multiple websites behind a single Application Gateway. Azure Application Gateway is a layer-7 load balancer.

It provides failover, performance-routing HTTP requests between different servers, whether they are on the cloud or on-premises.

Application provides many Application Delivery Controller (ADC) features including HTTP load balancing, cookie-based session affinity, **TLS offload**, custom health probes, support for multi-site, and many others.

## Web Application Firewall

Web Application Firewall is a feature of [Azure Application Gateway](#) that provides protection to web applications that use application gateway for standard Application Delivery Control (ADC) functions. Web application firewall does this by protecting them against most of the OWASP top 10 common web vulnerabilities.



- SQL injection protection
- Common Web Attacks Protection such as command injection, HTTP request smuggling, HTTP response splitting, and remote file inclusion attack
- Protection against HTTP protocol violations
- Protection against HTTP protocol anomalies such as missing host user-agent and accept headers
- Prevention against bots, crawlers, and scanners
- Detection of common application misconfigurations (that is, Apache, IIS, etc.)

A centralized web application firewall to protect against web attacks makes security management much simpler and gives better assurance to the application against the threats of intrusions. A WAF solution can also react to a security threat faster by patching a known vulnerability at a central location versus securing each of individual web applications. Existing application gateways can be converted to an application gateway with web application firewall easily.

## Traffic Manager

Microsoft [Azure Traffic Manager](#) allows you to control the distribution of user traffic for service endpoints in different data centers. Service endpoints supported by Traffic Manager include Azure VMs, Web Apps, and Cloud services. You can also use Traffic Manager with external, non-Azure endpoints. Traffic Manager uses the Domain Name

System (DNS) to direct client requests to the most appropriate endpoint based on a [traffic-routing method](#) and the health of the endpoints.

Traffic Manager provides a range of traffic-routing methods to suit different application needs, endpoint health [monitoring](#), and automatic failover. Traffic Manager is resilient to failure, including the failure of an entire Azure region.

## Azure Load Balancer

[Azure Load Balancer](#) delivers high availability and network performance to your applications. It is a Layer 4 (TCP, UDP) load balancer that distributes incoming traffic among healthy instances of services defined in a load-balanced set. Azure Load Balancer can be configured to:

- Load balance incoming Internet traffic to virtual machines. This configuration is known as [public load balancing](#).
- Load balance traffic between virtual machines in a virtual network, between virtual machines in cloud services, or between on-premises computers and virtual machines in a cross-premises virtual network. This configuration is known as [internal load balancing](#).
- Forward external traffic to a specific virtual machine

## Internal DNS

You can manage the list of DNS servers used in a VNet in the Management Portal, or in the network configuration file. Customer can add up to 12 DNS servers for each VNet. When specifying DNS servers, it's important to verify that you list customer's DNS servers in the correct order for customer's environment. DNS server lists do not work round-robin. They are used in the order that they are specified. If the first DNS server on the list is able to be reached, the client uses that DNS server regardless of whether the DNS server is functioning properly or not. To change the DNS server order for customer's virtual network, remove the DNS servers from the list and add them back in the order that customer wants. DNS supports the availability aspect of the "CIA" security triad.

## Azure DNS

The Domain Name System, or DNS, is responsible for translating (or resolving) a website or service name to its IP address. [Azure DNS](#) is a hosting service for DNS domains, providing name resolution using Microsoft Azure infrastructure. By hosting your

domains in Azure, you can manage your DNS records using the same credentials, APIs, tools, and billing as your other Azure services. DNS supports the availability aspect of the "CIA" security triad.

## Azure Monitor logs NSGs

You can enable the following diagnostic log categories for NSGs:

- Event: Contains entries for which NSG rules are applied to VMs and instance roles based on MAC address. The status for these rules is collected every 60 seconds.
- Rules counter: Contains entries for how many times each NSG rule is applied to deny or allow traffic.

## Microsoft Defender for Cloud

[Microsoft Defender for Cloud](#) continuously analyzes the security state of your Azure resources for network security best practices. When Defender for Cloud identifies potential security vulnerabilities, it creates [recommendations](#) that guide you through the process of configuring the needed controls to harden and protect your resources.

## Compute

The section provides additional information regarding key features in this area and summary information about these capabilities.

## Azure confidential computing

[Azure confidential computing](#) provides the final, missing piece, of the data protection puzzle. It allows you to keep your data encrypted at all times. While at rest, when in motion through the network, and now, even while loaded in memory and in use. Additionally, by making [Remote Attestation](#) possible, it allows you to cryptographically verify that the VM you provision has booted securely and is configured correctly, prior to unlocking your data.

The spectrum of option ranges from enabling "lift and shift" scenarios of existing applications, to a full control of security features. For Infrastructure as a Service (IaaS), you can use [confidential virtual machines powered by AMD SEV-SNP](#) or confidential application enclaves for virtual machines that run [Intel Software Guard Extensions \(SGX\)](#). For Platform as a Service, we have multiple [container based](#) options, including integrations with [Azure Kubernetes Service \(AKS\)](#).

## Antimalware & Antivirus

With Azure IaaS, you can use antimalware software from security vendors such as Microsoft, Symantec, Trend Micro, McAfee, and Kaspersky to protect your virtual machines from malicious files, adware, and other threats. [Microsoft Antimalware](#) for Azure Cloud Services and Virtual Machines is a protection capability that helps identify and remove viruses, spyware, and other malicious software. Microsoft Antimalware provides configurable alerts when known malicious or unwanted software attempts to install itself or run on your Azure systems. Microsoft Antimalware can also be deployed using Microsoft Defender for Cloud

## Hardware Security Module

Encryption and authentication do not improve security unless the keys themselves are protected. You can simplify the management and security of your critical secrets and keys by storing them in [Azure Key Vault](#). Key Vault provides the option to store your keys in hardware Security modules (HSMs) certified to FIPS 140-2 Level 2 standards. Your SQL Server encryption keys for backup or [transparent data encryption](#) can all be stored in Key Vault with any keys or secrets from your applications. Permissions and access to these protected items are managed through [Microsoft Entra ID](#).

## Virtual machine backup

[Azure Backup](#) is a solution that protects your application data with zero capital investment and minimal operating costs. Application errors can corrupt your data, and human errors can introduce bugs into your applications that can lead to security issues. With Azure Backup, your virtual machines running Windows and Linux are protected.

## Azure Site Recovery

An important part of your organization's [business continuity/disaster recovery \(BCDR\)](#) strategy is figuring out how to keep corporate workloads and apps up and running when planned and unplanned outages occur. [Azure Site Recovery](#) helps orchestrate replication, failover, and recovery of workloads and apps so that they are available from a secondary location if your primary location goes down.

## SQL VM TDE

Transparent data encryption (TDE) and column level encryption (CLE) are SQL server encryption features. This form of encryption requires customers to manage and store

the cryptographic keys you use for encryption.

The Azure Key Vault (AKV) service is designed to improve the security and management of these keys in a secure and highly available location. The SQL Server Connector enables SQL Server to use these keys from Azure Key Vault.

If you are running SQL Server with on-premises machines, there are steps you can follow to access Azure Key Vault from your on-premises SQL Server instance. But for SQL Server in Azure VMs, you can save time by using the Azure Key Vault Integration feature. With a few Azure PowerShell cmdlets to enable this feature, you can automate the configuration necessary for a SQL VM to access your key vault.

## VM Disk Encryption

[Azure Disk Encryption for Linux VMs](#) and [Azure Disk Encryption for Windows VMs](#) helps you encrypt your IaaS virtual machine disks. It applies the industry standard BitLocker feature of Windows and the DM-Crypt feature of Linux to provide volume encryption for the OS and the data disks. The solution is integrated with Azure Key Vault to help you control and manage the disk-encryption keys and secrets in your Key Vault subscription. The solution also ensures that all data on the virtual machine disks are encrypted at rest in your Azure storage.

## Virtual networking

Virtual machines need network connectivity. To support that requirement, Azure requires virtual machines to be connected to an Azure Virtual Network. An Azure Virtual Network is a logical construct built on top of the physical Azure network fabric. Each logical [Azure Virtual Network](#) is isolated from all other Azure Virtual Networks. This isolation helps ensure that network traffic in your deployments is not accessible to other Microsoft Azure customers.

## Patch Updates

Patch Updates provide the basis for finding and fixing potential problems and simplify the software update management process, both by reducing the number of software updates you must deploy in your enterprise and by increasing your ability to monitor compliance.

## Security policy management and reporting

Defender for Cloud helps you prevent, detect, and respond to threats, and provides you increased visibility into, and control over, the security of your Azure resources. It provides integrated Security monitoring and policy management across your Azure subscriptions, helps detect threats that might otherwise go unnoticed, and works with a broad ecosystem of security solutions.

## Identity and access management

Securing systems, applications, and data begins with identity-based access controls. The identity and access management features that are built into Microsoft business products and services help protect your organizational and personal information from unauthorized access while making it available to legitimate users whenever and wherever they need it.

### Secure Identity

Microsoft uses multiple security practices and technologies across its products and services to manage identity and access.

- [Multi-Factor Authentication](#) requires users to use multiple methods for access, on-premises and in the cloud. It provides strong authentication with a range of easy verification options, while accommodating users with a simple sign-in process.
- [Microsoft Authenticator](#) provides a user-friendly Multi-Factor Authentication experience that works with both Microsoft Entra ID and Microsoft accounts, and includes support for wearables and fingerprint-based approvals.
- [Password policy enforcement](#) increases the security of traditional passwords by imposing length and complexity requirements, forced periodic rotation, and account lockout after failed authentication attempts.
- [Token-based authentication](#) enables authentication via Microsoft Entra ID.
- [Azure role-based access control \(Azure RBAC\)](#) enables you to grant access based on the user's assigned role, making it easy to give users only the amount of access they need to perform their job duties. You can customize Azure RBAC per your organization's business model and risk tolerance.
- [Integrated identity management \(hybrid identity\)](#) enables you to maintain control of users' access across internal datacenters and cloud platforms, creating a single user identity for authentication and authorization to all resources.

## Secure Apps and data

Microsoft Entra ID [↗](#), a comprehensive identity and access management cloud solution, helps secure access to data in applications on site and in the cloud, and simplifies the management of users and groups. It combines core directory services, advanced identity governance, security, and application access management, and makes it easy for developers to build policy-based identity management into their apps. To enhance your Microsoft Entra ID, you can add paid capabilities using the Microsoft Entra Basic, Premium P1, and Premium P2 editions.

Free / Common Features	Basic Features	Premium P1 Features	Premium P2 Features	Microsoft Entra join – Windows 10 only related features
Directory Objects, User/Group Management (add/update/delete)/ User-based provisioning, Device registration, single sign-on (SSO), Self-Service Password Change for cloud users, Connect (Sync engine that extends on-premises directories to Microsoft Entra ID), Security / Usage Reports	Group-based access management / provisioning, Self-Service Password Reset for cloud users, Company Branding (Logon Pages/Access Panel customization), Application Proxy, SLA 99.9%	Self-Service Group and app Management/Self-Service application additions/Dynamic Groups, Self-Service Password Reset/Change/Unlock with on-premises write-back, Multi-Factor Authentication (Cloud and On-premises (MFA Server)), MIM CAL + MIM Server, Cloud App Discovery, Connect Health, Automatic password rollover for group accounts	Identity Protection, Privileged Identity Management	Join a device to Microsoft Entra ID, Desktop SSO, Microsoft Passport for Microsoft Entra ID, Administrator BitLocker recovery, MDM auto-enrollment, Self-Service BitLocker recovery, Additional local administrators to Windows 10 devices via Microsoft Entra join

- [Cloud App Discovery](#) is a premium feature of Microsoft Entra ID that enables you to identify cloud applications that are used by the employees in your organization.
- [Microsoft Entra ID Protection](#) is a security service that uses Microsoft Entra anomaly detection capabilities to provide a consolidated view into risk detections and potential vulnerabilities that could affect your organization's identities.

- [Microsoft Entra Domain Services](#) enables you to join Azure VMs to a domain without the need to deploy domain controllers. Users sign in to these VMs by using their corporate Active Directory credentials, and can seamlessly access resources.
- [Azure Active Directory B2C](#) is a highly available, global identity management service for consumer-facing apps that can scale to hundreds of millions of identities and integrate across mobile and web platforms. Your customers can sign in to all your apps through customizable experiences that use existing social media accounts, or you can create new standalone credentials.
- [Microsoft Entra B2B Collaboration](#) is a secure partner integration solution that supports your cross-company relationships by enabling partners to access your corporate applications and data selectively by using their self-managed identities.
- [Microsoft Entra joined](#) enables you to extend cloud capabilities to Windows 10 devices for centralized management. It makes it possible for users to connect to the corporate or organizational cloud through Microsoft Entra ID and simplifies access to apps and resources.
- [Microsoft Entra application proxy](#) provides SSO and secure remote access for web applications hosted on-premises.

## Next Steps

- Understand your [shared responsibility in the cloud](#).
- Learn how [Microsoft Defender for Cloud](#) can help you prevent, detect, and respond to threats with increased visibility and control over the security of your Azure resources.

# Security design principles

Article • 11/15/2023

A Well-Architected workload must be built with a zero-trust approach. A secure workload is **resilient to attacks** and incorporates the interrelated security **principles of confidentiality, integrity, and availability** (also known as the *CIA triad*) in addition to meeting business goals. Any security incident has the potential to become a major breach that damages the brand and reputation of the workload or organization. To measure the security efficacy of your overall strategy for a workload, start with these questions:

- Do your defensive investments provide meaningful cost and friction to prevent attackers from compromising your workload?
- Will your security measures be effective in restricting the blast radius of an incident?
- Do you understand how controlling the workload could be valuable for an attacker? Do you understand the impact to your business if the workload and its data are stolen, unavailable, or tampered with?
- Can the workload and operations quickly detect, respond to, and recover from disruptions?

As you design your system, use the Microsoft Zero Trust model as the compass to mitigate security risks:

- **Verify explicitly** so that **only trusted identities** perform **intended and allowed actions** that originate from **expected locations**. This safeguard makes it harder for attackers to impersonate legitimate users and accounts.
- **Use least-privilege access** for the **right identities**, with the **right set of permissions**, for the **right duration**, and to the **right assets**. Limiting permissions helps keep attackers from abusing permissions that legitimate users don't even need.
- **Assume breach** of security controls and design compensating controls that **limit risk and damage** if a primary layer of defense fails. Doing so helps you to defend your workload better by thinking like an attacker who's interested in success (regardless of how they get it).

Security isn't a one-time effort. You must implement this guidance on a recurring basis. Continuously improve your defenses and security knowledge to help keep your

workload safe from attackers who are constantly gaining access to innovative attack vectors as they're developed and added to automated attack kits.

The design principles are intended to establish an ongoing security mindset to help you continuously improve the security posture of your workload as the attempts of attackers continuously evolve. These principles should guide the security of your architecture, design choices, and operational processes. Start with the recommended approaches and **justify the benefits for a set of security requirements**. After you set your strategy, drive actions by using the [Security checklist](#) as your next step.

If these principles aren't applied properly, a negative impact on business operations and revenue can be expected. Some consequences might be obvious, like penalties for regulatory workloads. Others might not be so obvious and could lead to ongoing security problems before they're detected.

In many mission-critical workloads, security is the primary concern, alongside reliability, given that some attack vectors, like data exfiltration, don't affect reliability. Security and reliability can pull a workload in opposite directions because security-focused design can introduce points of failure and increase operational complexity. The effect of security on reliability is often indirect, introduced by way of operational constraints. Carefully consider tradeoffs between security and reliability.

By following these principles, you can improve security effectiveness, harden workload assets, and build trust with your users.

## Plan your security readiness

[ ] Expand table



**Strive to adopt and implement security practices in architectural design decisions and operations with minimal friction.**

As a workload owner, you have a shared responsibility with the organization to protect assets. Create a **security readiness plan** that's aligned with business priorities. It will lead to well-defined processes, adequate investments, and appropriate accountabilities. The plan should provide the workload requirements to the organization, which also shares responsibility for protecting assets. Security plans should factor into your strategy for reliability, health modeling, and self-preservation.

In addition to organizational assets, the workload itself needs to be protected from intrusion and exfiltration attacks. All facets of Zero Trust and the CIA triad should be factored into the plan.

Functional and non-functional requirements, budget constraints, and other considerations shouldn't restrict security investments or dilute assurances. At the same time, you need to engineer and plan security investments with those constraints and restrictions in mind.

 Expand table

Approach	Benefit
<p><b>Use segmentation as a strategy to plan security boundaries</b> in the workload environment, processes, and team structure to <b>isolate access and function</b>.</p> <p>Your segmentation strategy should be driven by business requirements. You can base it on criticality of components, division of labor, privacy concerns, and other factors.</p>	<p>You'll be able to <b>minimize operational friction</b> by defining roles and establishing <b>clear lines of responsibility</b>. This exercise also helps you <b>identify the level of access</b> for each role, especially for critical-impact accounts.</p> <p>Isolation enables you to <b>limit exposure of sensitive flows</b> to only roles and assets that need access. Excessive exposure could inadvertently lead to information flow disclosure.</p> <p>To summarize, you'll be able to <b>right-size security efforts</b> based on the needs of each segment.</p>
<p><b>Continuously build skills</b> through <b>role-based security training</b> that meets the requirements of the organization and the use cases of the workload.</p>	<p>A highly skilled team can design, implement, and monitor <b>security controls</b> that remain effective against attackers, who constantly look for new ways to exploit the system.</p> <p>Organization-wide training typically focuses on developing a broader skill set for securing the common elements. However, with role-based training, you focus on <b>developing deep expertise</b> in the platform offerings and security features that address workload concerns.</p> <p>You need to implement both approaches to defend against adversaries through <b>good design and effective operations</b>.</p>
<p><b>Make sure there's an incident response plan</b> for your workload.</p> <p>Use industry frameworks that define the standard operating procedure for preparedness, detection, containment, mitigation, and post-incident activity.</p>	<p>At the time of crisis, confusion must be avoided.</p> <p>If you have a well-documented plan, responsible roles can <b>focus on execution</b> without wasting time on uncertain actions. Also, a comprehensive plan can help you ensure that <b>all remediation requirements are fulfilled</b>.</p>
<p><b>Strengthen your security posture</b> by understanding the security compliance</p>	<p>Clarity about compliance requirements will help you <b>design for the right security assurances</b> and</p>

Approach	Benefit
<p>requirements that are imposed by influences outside the workload team, like organizational policies, regulatory compliance, and industry standards.</p>	<p><b>prevent non-compliance</b> issues, which could lead to penalties.</p> <p>Industry standards can provide a baseline and influence your choice of tools, policies, security safeguards, guidelines, risk-management approaches, and training.</p> <p>If you know that the workload adheres to compliance, you'll be able to <b>instill confidence</b> in your user base.</p>
<p><b>Define and enforce team-level security standards</b> across the lifecycle and operations of the workload.</p> <p><b>Strive for consistent practices</b> in operations like coding, gated approvals, release management, and data protection and retention.</p>	<p>Defining good security practices can <b>minimize negligence</b> and the surface area for potential errors. The team will <b>optimize efforts and the outcome will be predictable</b> because approaches are made more consistent.</p> <p>Observing security standards over time will enable you to <b>identify opportunities for improvement, possibly including automation</b>, which will streamline efforts further and increase consistency.</p>
<p>Align your incident response with the <b>Security Operation Center (SOC) centralized function</b> in your organization.</p>	<p>Centralizing incident response functions enables you to take advantage of specialized IT professionals who can detect incidents in real time to address potential threats as quickly as possible.</p>

## Design to protect confidentiality

[ ] Expand table



**Prevent exposure to privacy, regulatory, application, and proprietary information through access restrictions and obfuscation techniques.**

Workload data can be classified by user, usage, configuration, compliance, intellectual property, and more. That data can't be shared or accessed beyond the established trust boundaries. Efforts to protect confidentiality should focus on access controls, opacity, and keeping an audit trail of activities that pertain to data and the system.

[ ] Expand table

Approach	Benefit
<p>Implement <b>strong access controls</b> that grant access only on a need-to-know basis.</p>	<p><i>Least privilege.</i></p> <p>The workload will be protected from <b>unauthorized access</b> and prohibited activities. Even when access is from trusted identities, the <b>access permissions and exposure time will be minimized</b> because the communication path is open for a limited period.</p>
<p><b>Classify data based on its type, sensitivity, and potential risk.</b></p> <p>Assign a confidentiality level for each.</p> <p>Include system components that are in scope for the identified level.</p>	<p><i>Verify explicitly.</i></p> <p>This evaluation helps you right-size security measures. You'll also be able to identify data and components that have a <b>high potential impact</b> and/or exposure to risk. This exercise adds <b>clarity</b> to your information protection strategy and helps ensure <b>agreement</b>.</p>
<p>Safeguard your data at rest, in transit, and during processing by using <b>encryption</b>. Base your strategy on the assigned confidentiality level.</p>	<p><i>Assume breach.</i></p> <p>Even if an attacker gets access, they <b>won't be able to read properly encrypted sensitive data</b>. Sensitive data includes configuration information that's used to gain further access inside the system. Data encryption can help you <b>contain risks</b>.</p>
<p><b>Guard against exploits</b> that might cause unwarranted exposure of information.</p>	<p><i>Verify explicitly.</i></p> <p>It's crucial to minimize vulnerabilities in authentication and authorization implementations, code, configurations, operations, and those that stem from the social habits of the system's users.</p> <p>Up-to-date security measures enable you to <b>block known security vulnerabilities</b> from entering the system. You can also <b>mitigate new vulnerabilities</b> that can appear over time by implementing routine operations throughout the development cycle, continuously improving security assurances.</p>
<p><b>Guard against data exfiltration</b> that results from malicious or inadvertent access to data.</p>	<p><i>Assume breach.</i></p> <p>You'll be able to contain blast radius by <b>blocking unauthorized data transfer</b>. Additionally, controls applied to networking, identity, and encryption will protect data at various layers.</p>

Approach	Benefit
Maintain the level of confidentiality as data flows through various components of the system.	<p><i>Assume breach.</i></p> <p>Enforcing confidentiality levels throughout the system enables you to provide a consistent level of hardening. Doing so can <b>prevent vulnerabilities</b> that might result from moving data to a lower security tier.</p>
Maintain an <b>audit trail</b> of all types of access activities.	<p><i>Assume breach.</i></p> <p>Audit logs support <b>faster detection and recovery</b> in case of incidents and help with ongoing security monitoring.</p>

## Design to protect integrity

[+] [Expand table](#)



**Prevent corruption of design, implementation, operations, and data to avoid disruptions that can stop the system from delivering its intended utility or cause it to operate outside the prescribed limits. The system should provide information assurance throughout the workload lifecycle.**

The key is to implement controls that prevent tampering of business logic, flows, deployment processes, data, and even the lower stack components, like the operating system and boot sequence. Lack of integrity can introduce vulnerabilities that can lead to breaches in confidentiality and availability.

[+] [Expand table](#)

Approach	Benefit
Implement strong access controls that authenticate and authorize access to the system.	<p><i>Least privilege.</i></p> <p>Depending on the strength of the controls, you'll be able to <b>prevent or reduce risks from unapproved modifications</b>. This helps ensure that data is consistent and trustworthy.</p>
Minimize access based on privilege, scope, and time.	<p>Minimizing access limits the extent of potential corruption.</p>
Continuously protect against vulnerabilities and detect them in your supply chain to block attackers from injecting software faults	<p><i>Assume breach.</i></p> <p>Knowing the origin of software and verifying its</p>

Approach	Benefit
<p>into your infrastructure, build system, tools, libraries, and other dependencies.</p> <p>Supply chain should scan for vulnerabilities during <b>build time and runtime</b>.</p>	<p>authenticity throughout the lifecycle will provide <b>predictability</b>. You'll <b>know about vulnerabilities well in advance</b> so that you can proactively remediate them and keep the system secure in production.</p>
<p><b>Establish trust and verify by using cryptography techniques</b> like code signing, certificates, and encryption.</p> <p>Protect those mechanisms by allowing reputable decryption.</p>	<p><i>Verify explicitly, least privilege.</i></p> <p>You'll know that changes to data or access to the system is <b>verified by a trusted source</b>.</p> <p>Even if encrypted data is intercepted in transit by a malicious actor, the actor won't be able to unlock or decipher the content. You can use digital signatures to ensure that the data wasn't tampered with during transmission.</p>
<p><b>Ensure backup data is immutable and encrypted</b> when data is replicated or transferred.</p>	<p><i>Verify explicitly.</i></p> <p>You'll be able to recover data with confidence that <b>backup data wasn't changed at rest</b>, inadvertently or maliciously.</p>
<p><b>Avoid or mitigate system implementations that allow your workload to operate outside its intended limits and purposes.</b></p>	<p><i>Verify explicitly.</i></p> <p>When your system has strong safeguards that check whether usage aligns with its intended limits and purposes, the scope for potential abuse or tampering of your compute, networking, and data stores is reduced.</p>

## Design to protect availability

[] Expand table



**Prevent or minimize system and workload downtime and degradation in the event of a security incident by using strong security controls. You must maintain data integrity during the incident and after the system recovers.**

You need to balance availability architecture choices with security architecture choices. The system should have availability guarantees to ensure that users have access to data and that data is reachable. From a security perspective, users should operate within the allowed access scope, and the data must be trusted. Security controls should block bad actors, but they shouldn't block legitimate users from accessing the system and data.

Approach	Benefit
<p>Prevent compromised identities from misusing access to gain control of the system.</p> <p>Check for <b>overly pervasive scope and time limits</b> to minimize risk exposure.</p>	<p><i>Least privilege.</i></p> <p>This strategy <b>mitigates the risks of excessive, unnecessary, or misused access permissions</b> on crucial resources. Risks include unauthorized modifications and even the deletion of resources. Take advantage of the platform-provided just-in-time (JIT), just-enough-access (JEA), and time-based security modes to replace standing permissions wherever possible.</p>
<p>Use security controls and design patterns to <b>prevent attacks and code flaws from causing resource exhaustion</b> and blocking access.</p>	<p><i>Verify explicitly.</i></p> <p>The <b>system won't experience downtime</b> caused by malicious actions, like distributed denial of service (DDoS) attacks.</p>
<p>Implement <b>preventative measures for attack vectors that exploit vulnerabilities</b> in application code, networking protocols, identity systems, malware protection, and other areas.</p>	<p><i>Assume breach.</i></p> <p>Implement code scanners, apply the latest security patches, update software, and protect your system with effective antimalware on an ongoing basis.</p> <p>You'll be able to reduce the attack surface to ensure business continuity.</p>
<p>Prioritize security controls on the <b>critical components and flows</b> in the system that are susceptible to risk.</p>	<p><i>Assume breach, verify explicitly.</i></p> <p>Regular detection and prioritization exercises can help you <b>apply security expertise to the critical aspects</b> of the system. You'll be able to focus on the most likely and damaging threats and start your risk mitigation in areas that need the most attention.</p>
<p>Apply at least the same level of <b>security rigor in your recovery resources and processes</b> as you do in the primary environment, including security controls and frequency of backup.</p>	<p><i>Assume breach.</i></p> <p>You should have a preserved safe system state available in disaster recovery. If you do, you can fail over to a secure secondary system or location and restore backups that won't introduce a threat.</p> <p>A well-designed process can prevent a security incident from hindering the recovery process. Corrupted backup data or encrypted data that can't be deciphered can slow down recovery.</p>

# Sustain and evolve your security posture

Expand table



**Incorporate continuous improvement and apply vigilance to stay ahead of attackers who are continuously evolving their attack strategies.**

Your security posture must not degrade over time. You must continually improve security operations so that new disruptions are handled more efficiently. Strive to align improvements with the phases defined by industry standards. Doing so leads to better preparedness, reduced time to incident detection, and effective containment and mitigation. Continuous improvement should be based on lessons learned from past incidents.

It's important to measure your security posture, enforce policies to maintain that posture, and regularly validate your security mitigations and compensating controls in order to continuously improve your security posture in the face of evolving threats.

Expand table

Approach	Benefit
Create and maintain a comprehensive asset inventory that includes classified information about resources, locations, dependencies, owners, and other metadata that's relevant to security.	A well-organized inventory provides a <b>holistic view of the environment</b> , which puts you in an advantageous position against attackers, especially during post-incident activities.  It also creates a business rhythm to drive communication, upkeep of critical components, and the decommissioning of orphaned resources.
As much as possible, <b>automate</b> inventory to derive data from the system.	
Perform threat modeling to identify and mitigate potential threats.	You'll have a <b>report of attack vectors</b> prioritized by their severity level. You'll be able to identify threats and vulnerabilities quickly and set up countermeasures.
Regularly capture data to quantify your current state against your established security baseline and <b>set priorities for remediations</b> .	You need accurate reports that bring clarity and consensus to focus areas. You'll be able to immediately <b>execute technical remediations</b> , starting with the highest priority items. You'll also <b>identify gaps</b> , which provide opportunities for improvement.
Take advantage of platform-provided features for <b>security posture management</b> and the enforcement of	Implementing enforcement helps prevent violations

Approach	Benefit
<p><b>compliance imposed by external and internal organizations.</b></p>	<p>and regressions, which preserves your security posture.</p>
<p><b>Run periodic security tests</b> that are conducted by experts external to the workload team who attempt to ethically hack the system.</p>	<p>These tests enable you to validate security defenses by <b>simulating real-world attacks</b> by using techniques like penetration testing.</p>
<p>Perform routine and integrated <b>vulnerability scanning</b> to detect exploits in infrastructure, dependencies, and application code.</p>	<p>Threats can be introduced as part of your change management. Integrating scanners into the deployment pipelines enables you to automatically detect vulnerabilities and even quarantine usage until the vulnerabilities are removed.</p>
<p><b>Detect, respond, and recover</b> with swift and effective security operations.</p>	<p>The primary benefit of this approach is that it enables you to <b>preserve or restore the security assurances of the CIA triad</b> during and after an attack.</p>
	<p>You need to be alerted as soon as a threat is detected so that you can start your investigations and take appropriate actions.</p>
<p><b>Conduct post-incident activities</b> like root-cause analyses, postmortems, and incident reports.</p>	<p>These activities provide insight into the impact of the breach and into resolution measures, which drives improvements in defenses and operations.</p>
<p><b>Get current, and stay current.</b></p>	<p>You'll be able to ensure that your <b>security posture doesn't degrade over time</b>.</p>
<p>Stay current on updates, patching, and security fixes.</p>	<p>By integrating findings from real-world attacks and testing activities, you'll be able to combat attackers who continuously improve and exploit new categories of vulnerabilities.</p>
<p>Continuously evaluate the system and improve it based on audit reports, benchmarking, and lessons from testing activities. Consider automation, as appropriate.</p>	<p>Automation of repetitive tasks <b>decreases the chance of human error</b> that can create risk.</p>
<p>Use threat intelligence powered by security analytics for dynamic detection of threats.</p>	<p>SDL reviews bring clarity around security features. SDL can help you maintain an inventory of workload assets and their security reports, which cover origin, usage, operational weaknesses, and other factors.</p>
<p>At regular intervals, review the workload's conformance to Security Development Lifecycle (SDL) best practices.</p>	

## Next steps

## Security checklist

# Cloud design patterns that support security

Article • 11/14/2023

When you design workload architectures, you should use industry patterns that address common challenges. Patterns can help you make intentional tradeoffs within workloads and optimize for your desired outcome. They can also help mitigate risks that originate from specific problems, which can affect reliability, performance, cost, and operations. Those risks might be indicative of lack of security assurances, if left unattended can pose significant risks to the business. These patterns are backed by real-world experience, are designed for cloud scale and operating models, and are inherently vendor agnostic. Using well-known patterns as a way to standardize your workload design is a component of operational excellence.

Many design patterns directly support one or more architecture pillars. Design patterns that support the Security pillar prioritize concepts like segmentation and isolation, strong authorization, uniform application security, and modern protocols.

## Design patterns for security

The following table summarizes cloud design patterns that support the goals of security.

Pattern	Summary
Ambassador	Encapsulates and manages network communications by offloading cross-cutting tasks that are related to network communication. The resulting helper services initiate communication on behalf of the client. This mediation point provides an opportunity to augment security on network communications.
Backends for Frontends	Individualizes the service layer of a workload by creating separate services that are exclusive to a specific frontend interface. Because of this separation, the security and authorization in the service layer that support one client can be tailored to the functionality provided by that client, potentially reducing the surface area of an API and lateral movement among different backends that might expose different capabilities.
Bulkhead	Introduces intentional and complete segmentation between components to isolate the blast radius of malfunctions. You can also use this strategy to contain security incidents to the compromised bulkhead.
Claim Check	Separates data from the messaging flow, providing a way to separately retrieve the data related to a message. This pattern supports keeping

Pattern	Summary
	sensitive data out of message bodies, instead keeping it managed in a secured data store. This configuration enables you to establish stricter authorization to support access to the sensitive data from services that are expected to use the data, but remove visibility from ancillary services like queue monitoring solutions.
<a href="#">Federated Identity</a>	Delegates trust to an identity provider that's external to the workload for managing users and providing authentication for your application. By externalizing user management and authentication, you can get evolved capabilities for identity-based threat detection and prevention without needing to implement these capabilities in your workload. External identity providers use modern interoperable authentication protocols.
<a href="#">Gatekeeper</a>	Offloads request processing that's specifically for security and access control enforcement before and after forwarding the request to a backend node. Adding a gateway into the request flow enables you to centralize security functionality like web application firewalls, DDoS protection, bot detection, request manipulation, authentication initiation, and authorization checks.
<a href="#">Gateway Aggregation</a>	Simplifies client interactions with your workload by aggregating calls to multiple backend services in a single request. This topology often reduces the number of touch points a client has with a system, which reduces the public surface area and authentication points. The aggregated backends can stay fully network-isolated from clients.
<a href="#">Gateway Offloading</a>	Offloads request processing to a gateway device before and after forwarding the request to a backend node. Adding a gateway into the request flow enables you to centralize security functionality like web application firewalls and TLS connections with clients. Any offloaded functionality that's platform-provided already offers enhanced security.
<a href="#">Publisher/Subscriber</a>	Decouples components in an architecture by replacing direct client-to-service communication with communication via an intermediate message broker or event bus. This replacement introduces an important security segmentation boundary that enables queue subscribers to be network-isolated from the publisher.
<a href="#">Sidecar</a>	Extends the functionality of an application by encapsulating nonprimary or cross-cutting tasks in a companion process that exists alongside the main application. By encapsulating these tasks and deploying them out-of-process, you can reduce the surface area of sensitive processes to only the code that's needed to accomplish the task. You can also use sidecars to add cross-cutting security controls to an application component that's not natively designed with that functionality.
<a href="#">Throttling</a>	Imposes limits on the rate or throughput of incoming requests to a resource or component. You can design the limits to help prevent resource exhaustion that could result from automated abuse of the system.

Pattern	Summary
Valet Key	Grants security-restricted access to a resource without using an intermediary resource to proxy the access. This pattern enables a client to directly access a resource without needing long-lasting or standing credentials. All access requests start with an auditable transaction. The granted access is then limited in both scope and duration. This pattern also makes it easier to revoke the granted access.

## Next steps

Review the cloud design patterns that support the other Azure Well-Architected Framework pillars:

- [Cloud design patterns that support reliability](#)
- [Cloud design patterns that support cost optimization](#)
- [Cloud design patterns that support operational excellence](#)
- [Cloud design patterns that support performance efficiency](#)

# End-to-end security in Azure

Article • 10/12/2023

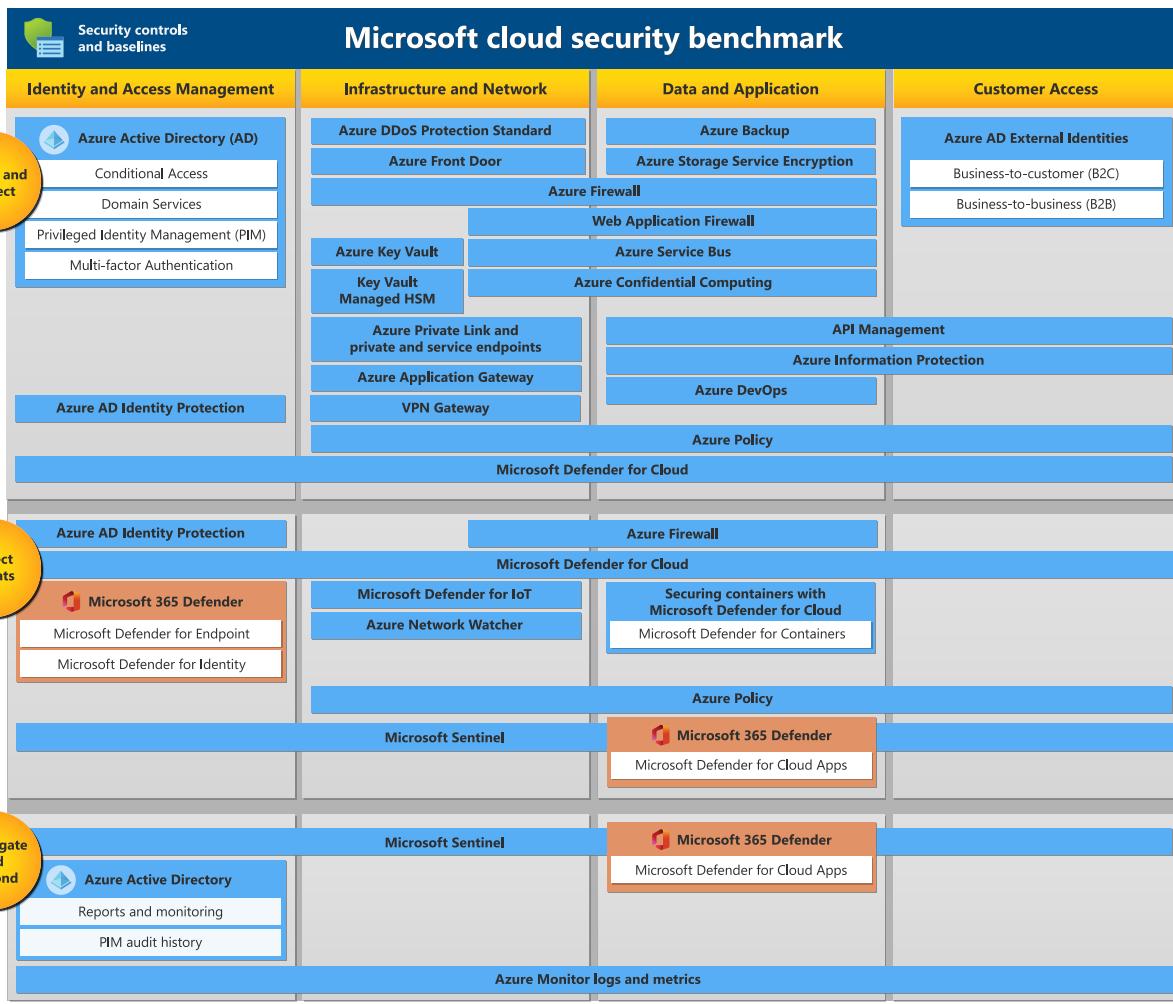
One of the best reasons to use Azure for your applications and services is to take advantage of its wide array of security tools and capabilities. These tools and capabilities help make it possible to create secure solutions on the secure Azure platform. Microsoft Azure provides confidentiality, integrity, and availability of customer data, while also enabling transparent accountability.

The following diagram and documentation introduces you to the security services in Azure. These security services help you meet the security needs of your business and protect your users, devices, resources, data, and applications in the cloud.

## Microsoft security services map

The security services map organizes services by the resources they protect (column). The diagram also groups services into the following categories (row):

- Secure and protect - Services that let you implement a layered, defense in-depth strategy across identity, hosts, networks, and data. This collection of security services and capabilities provides a way to understand and improve your security posture across your Azure environment.
- Detect threats – Services that identify suspicious activities and facilitate mitigating the threat.
- Investigate and respond – Services that pull logging data so you can assess a suspicious activity and respond.

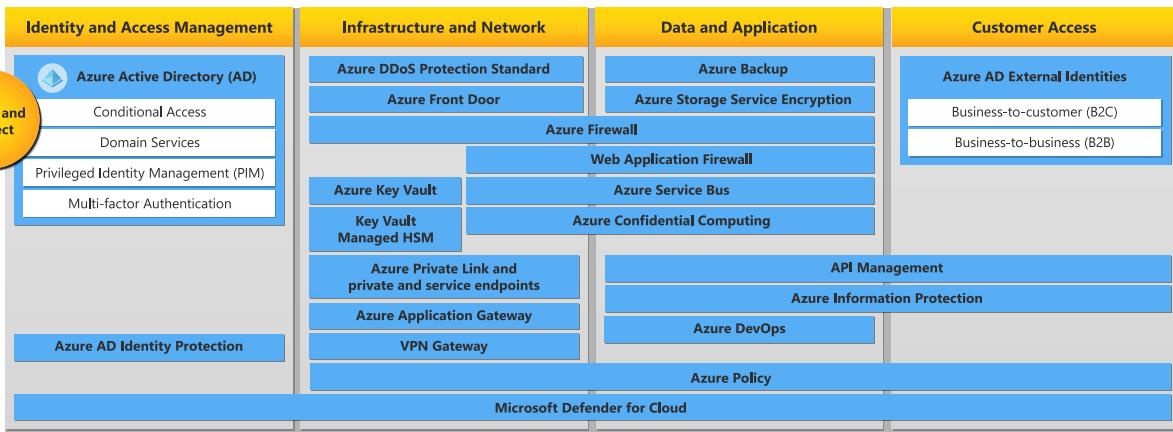


## Security controls and baselines

The [Microsoft cloud security benchmark](#) includes a collection of high-impact security recommendations you can use to help secure the services you use in Azure:

- Security controls - These recommendations are generally applicable across your Azure tenant and Azure services. Each recommendation identifies a list of stakeholders that are typically involved in planning, approval, or implementation of the benchmark.
- Service baselines - These apply the controls to individual Azure services to provide recommendations on that service's security configuration.

## Secure and protect

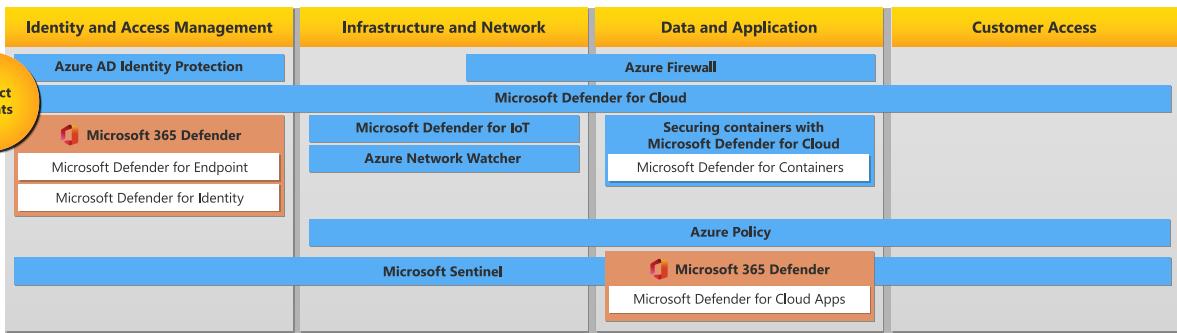


Service	Description
<a href="#">Microsoft Defender for Cloud</a>	A unified infrastructure security management system that strengthens the security posture of your data centers, and provides advanced threat protection across your hybrid workloads in the cloud - whether they're in Azure or not - as well as on premises.
<b>Identity &amp; Access Management</b>	
<a href="#">Microsoft Entra ID</a>	Microsoft's cloud-based identity and access management service.
	<a href="#">Conditional Access</a> is the tool used by Microsoft Entra ID to bring identity signals together, to make decisions, and enforce organizational policies.
	<a href="#">Domain Services</a> is the tool used by Microsoft Entra ID to provide managed domain services such as domain join, group policy, lightweight directory access protocol (LDAP), and Kerberos/NTLM authentication.
	<a href="#">Privileged Identity Management (PIM)</a> is a service in Microsoft Entra ID that enables you to manage, control, and monitor access to important resources in your organization.
	<a href="#">Multi-factor authentication</a> is the tool used by Microsoft Entra ID to help safeguard access to data and applications by requiring a second form of authentication.
<a href="#">Microsoft Entra ID Protection</a>	A tool that allows organizations to automate the detection and remediation of identity-based risks, investigate risks using data in the portal, and export risk detection data to third-party utilities for further analysis.
<b>Infrastructure &amp; Network</b>	
<a href="#">VPN Gateway</a>	A virtual network gateway that is used to send encrypted traffic between an Azure virtual network and an on-premises location

Service	Description
	over the public Internet and to send encrypted traffic between Azure virtual networks over the Microsoft network.
<a href="#">Azure DDoS Protection</a>	Provides enhanced DDoS mitigation features to defend against DDoS attacks. It is automatically tuned to help protect your specific Azure resources in a virtual network.
<a href="#">Azure Front Door</a>	A global, scalable entry-point that uses the Microsoft global edge network to create fast, secure, and widely scalable web applications.
<a href="#">Azure Firewall</a>	A cloud-native and intelligent network firewall security service that provides threat protection for your cloud workloads running in Azure. It's a fully stateful firewall as a service with built-in high availability and unrestricted cloud scalability. Azure Firewall is offered in three SKUs: <a href="#">Standard</a> , <a href="#">Premium</a> , and <a href="#">Basic</a> .
<a href="#">Azure Key Vault</a>	A secure secrets store for tokens, passwords, certificates, API keys, and other secrets. Key Vault can also be used to create and control the encryption keys used to encrypt your data.
<a href="#">Key Vault Managed HSM</a>	A fully managed, highly available, single-tenant, standards-compliant cloud service that enables you to safeguard cryptographic keys for your cloud applications, using FIPS 140-2 Level 3 validated HSMs.
<a href="#">Azure Private Link</a>	Enables you to access Azure PaaS Services (for example, Azure Storage and SQL Database) and Azure hosted customer-owned/partner services over a private endpoint in your virtual network.
<a href="#">Azure Application Gateway</a>	An advanced web traffic load balancer that enables you to manage traffic to your web applications. Application Gateway can make routing decisions based on additional attributes of an HTTP request, for example URI path or host headers.
<a href="#">Azure Service Bus</a>	A fully managed enterprise message broker with message queues and publish-subscribe topics. Service Bus is used to decouple applications and services from each other.
<a href="#">Web Application Firewall</a>	Provides centralized protection of your web applications from common exploits and vulnerabilities. WAF can be deployed with Azure Application Gateway and Azure Front Door.
<a href="#">Azure Policy</a>	Helps to enforce organizational standards and to assess compliance at-scale. Through its compliance dashboard, it provides an aggregated view to evaluate the overall state of the environment, with the ability to drill down to the per-resource, per-policy granularity. It also helps to bring your

Service	Description
	resources to compliance through bulk remediation for existing resources and automatic remediation for new resources.
<b>Data &amp; Application</b>	
<a href="#">Azure Backup</a>	Provides simple, secure, and cost-effective solutions to back up your data and recover it from the Microsoft Azure cloud.
<a href="#">Azure Storage Service Encryption</a>	Automatically encrypts data before it is stored and automatically decrypts the data when you retrieve it.
<a href="#">Azure Information Protection</a>	A cloud-based solution that enables organizations to discover, classify, and protect documents and emails by applying labels to content.
<a href="#">API Management</a>	A way to create consistent and modern API gateways for existing back-end services.
<a href="#">Azure confidential computing</a>	Allows you to isolate your sensitive data while it's being processed in the cloud.
<a href="#">Azure DevOps</a>	Your development projects benefit from multiple layers of security and governance technologies, operational practices, and compliance policies when stored in Azure DevOps.
<b>Customer Access</b>	
<a href="#">Microsoft Entra External ID</a>	With External Identities in Microsoft Entra ID, you can allow people outside your organization to access your apps and resources, while letting them sign in using whatever identity they prefer.
	You can share your apps and resources with external users via <a href="#">Microsoft Entra B2B</a> collaboration.
	<a href="#">Azure AD B2C</a> lets you support millions of users and billions of authentications per day, monitoring and automatically handling threats like denial-of-service, password spray, or brute force attacks.

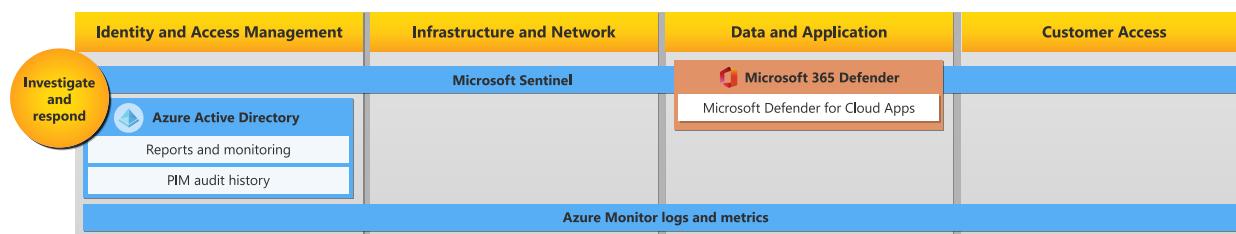
## Detect threats



Service	Description
Microsoft Defender for Cloud	Brings advanced, intelligent, protection of your Azure and hybrid resources and workloads. The workload protection dashboard in Defender for Cloud provides visibility and control of the cloud workload protection features for your environment.
Microsoft Sentinel	A scalable, cloud-native, security information event management (SIEM) and security orchestration automated response (SOAR) solution. Sentinel delivers intelligent security analytics and threat intelligence across the enterprise, providing a single solution for alert detection, threat visibility, proactive hunting, and threat response.
<b>Identity &amp; Access Management</b>	
Microsoft 365 Defender	A unified pre- and post-breach enterprise defense suite that natively coordinates detection, prevention, investigation, and response across endpoints, identities, email, and applications to provide integrated protection against sophisticated attacks.
	<b>Microsoft Defender for Endpoint</b> is an enterprise endpoint security platform designed to help enterprise networks prevent, detect, investigate, and respond to advanced threats.
	<b>Microsoft Defender for Identity</b> is a cloud-based security solution that leverages your on-premises Active Directory signals to identify, detect, and investigate advanced threats, compromised identities, and malicious insider actions directed at your organization.
Microsoft Entra ID Protection	Sends two types of automated notification emails to help you manage user risk and risk detections: Users at risk detected email and Weekly digest email.
<b>Infrastructure &amp; Network</b>	
Azure Firewall	Azure Firewall Premium provides signature-based intrusion detection and prevention system (IDPS) to allow rapid detection of attacks by looking for specific patterns, such as

Service	Description
	byte sequences in network traffic, or known malicious instruction sequences used by malware.
Microsoft Defender for IoT	A unified security solution for identifying IoT/OT devices, vulnerabilities, and threats. It enables you to secure your entire IoT/OT environment, whether you need to protect existing IoT/OT devices or build security into new IoT innovations.
Azure Network Watcher	Provides tools to monitor, diagnose, view metrics, and enable or disable logs for resources in an Azure virtual network. Network Watcher is designed to monitor and repair the network health of IaaS products which includes virtual machines, virtual networks, application gateways, and load balancers.
Azure Policy	Helps to enforce organizational standards and to assess compliance at-scale. Azure Policy uses activity logs, which are automatically enabled to include event source, date, user, timestamp, source addresses, destination addresses, and other useful elements.
<b>Data &amp; Application</b>	
Microsoft Defender for Containers	A cloud-native solution that is used to secure your containers so you can improve, monitor, and maintain the security of your clusters, containers, and their applications.
Microsoft Defender for Cloud Apps	A cloud access security broker (CASB) that operates on multiple clouds. It provides rich visibility, control over data travel, and sophisticated analytics to identify and combat cyberthreats across all your cloud services.

## Investigate and respond



Service	Description
Microsoft Sentinel	Powerful search and query tools to hunt for security threats across your organization's data sources.

Service	Description
Azure Monitor logs and metrics	Delivers a comprehensive solution for collecting, analyzing, and acting on telemetry from your cloud and on-premises environments. Azure Monitor <a href="#">collects and aggregates data</a> from a variety of sources into a common data platform where it can be used for analysis, visualization, and alerting.
<b>Identity &amp; Access Management</b>	
Azure AD reports and monitoring	<a href="#">Microsoft Entra reports</a> provide a comprehensive view of activity in your environment.
	<a href="#">Microsoft Entra monitoring</a> lets you route your Microsoft Entra activity logs to different endpoints.
<a href="#">Microsoft Entra PIM audit history</a>	Shows all role assignments and activations within the past 30 days for all privileged roles.
<b>Data &amp; Application</b>	
<a href="#">Microsoft Defender for Cloud Apps</a>	Provides tools to gain a deeper understanding of what's happening in your cloud environment.

## Next steps

- Understand your [shared responsibility in the cloud](#).
- Understand the [isolation choices in the Azure cloud](#) against both malicious and non-malicious users.

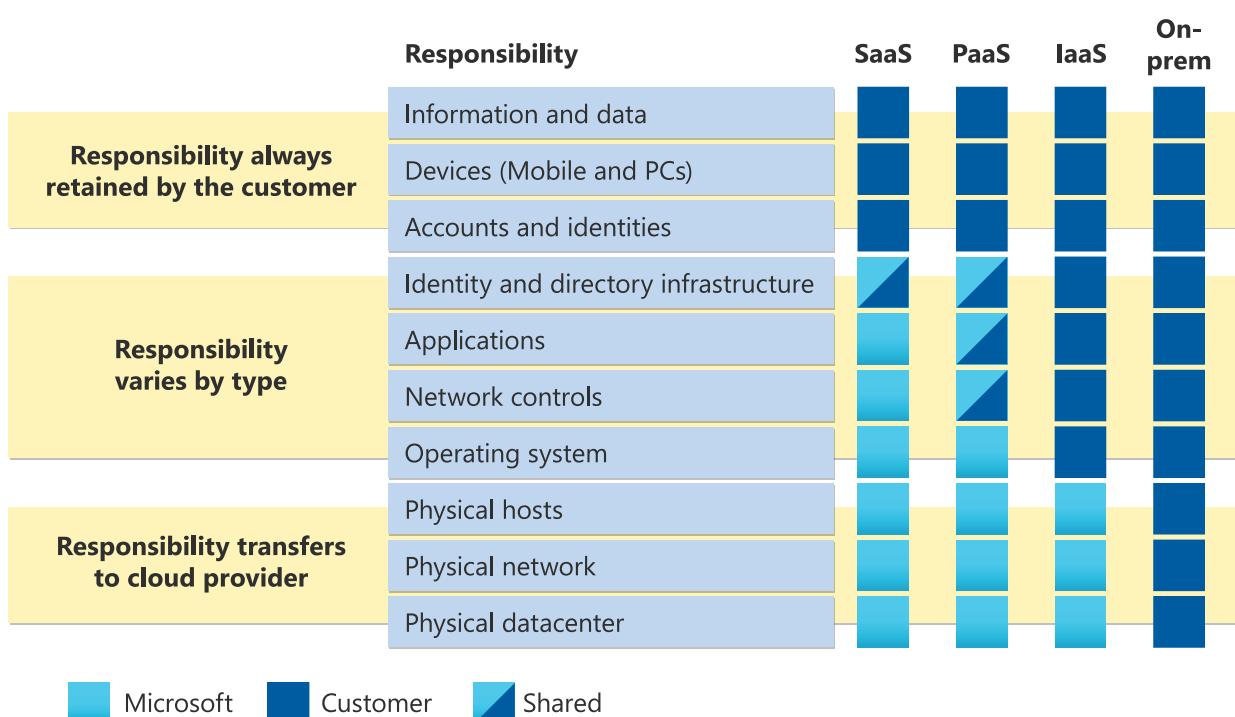
# Shared responsibility in the cloud

Article • 09/29/2023

As you consider and evaluate public cloud services, it's critical to understand the shared responsibility model and which security tasks the cloud provider handles and which tasks you handle. The workload responsibilities vary depending on whether the workload is hosted on Software as a Service (SaaS), Platform as a Service (PaaS), Infrastructure as a Service (IaaS), or in an on-premises datacenter.

## Division of responsibility

In an on-premises datacenter, you own the whole stack. As you move to the cloud some responsibilities transfer to Microsoft. The following diagram illustrates the areas of responsibility between you and Microsoft, according to the type of deployment of your stack.



For all cloud deployment types, you own your data and identities. You're responsible for protecting the security of your data and identities, on-premises resources, and the cloud components you control. Cloud components you control vary by service type.

Regardless of the type of deployment, you always retain the following responsibilities:

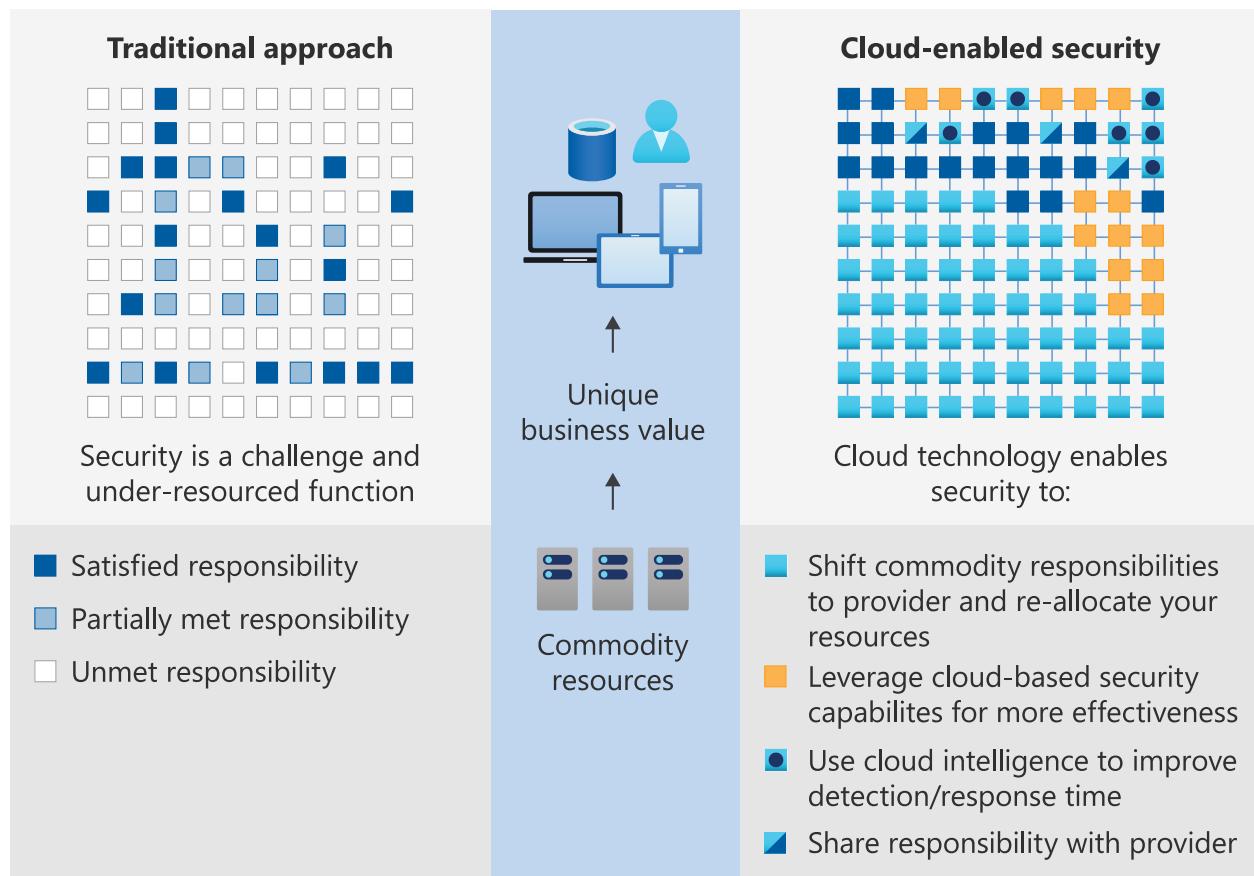
- Data
- Endpoints
- Account

- Access management

## Cloud security advantages

The cloud offers significant advantages for solving long standing information security challenges. In an on-premises environment, organizations likely have unmet responsibilities and limited resources available to invest in security, which creates an environment where attackers are able to exploit vulnerabilities at all layers.

The following diagram shows a traditional approach where many security responsibilities are unmet due to limited resources. In the cloud-enabled approach, you're able to shift day to day security responsibilities to your cloud provider and reallocate your resources.



In the cloud-enabled approach, you're also able to apply cloud-based security capabilities for more effectiveness and use cloud intelligence to improve your threat detection and response time. By shifting responsibilities to the cloud provider, organizations can get more security coverage, which enables them to reallocate security resources and budget to other business priorities.

## Next step

Learn more about shared responsibility and strategies to improve your security posture in the Well-Architected Framework's [overview of the security pillar](#).

# Security operations overview

Article • 04/24/2023

Security operations (SecOps) maintain and restore the security assurances of the system as live adversaries attack it. The NIST Cybersecurity Framework describes the SecOps functions of Detect, Respond, and Recover well.

- **Detect** - SecOps must detect the presence of adversaries in the system, who are incentivized to stay hidden in most cases, allowing them to achieve their objectives unimpeded. This can take the form of reacting to an alert of suspicious activity or proactively hunting for anomalous events in the enterprise activity logs.
- **Respond** - Upon detection of potential adversary action or campaign, SecOps must rapidly investigate to identify whether it's an actual attack (true positive) or a false alarm (false positive) and then enumerate the scope and goal of the adversary operation.
- **Recover** - The ultimate goal of SecOps is to preserve or restore the security assurances (confidentiality, integrity, availability) of business services during and after an attack.

The most significant security risk most organizations face is from human attack operators (of varying skill levels). Risk from automated/repeated attacks have been mitigated significantly for most organizations by signature and machine learning based approaches built into anti-malware. Although it must be noted that there are notable exceptions like WannaCrypt and NotPetya, which moved faster than these defenses).

While human attack operators are challenging to face because of their adaptability (vs. automated/repeated logic), they're operating at the same "human speed" as defenders, which help level the playing field.

SecOps (sometimes referred to as a Security Operations Center (SOC)) has a critical role to play in limiting the time and access an attacker can get to valuable systems and data. Each minute that an attacker has in the environment allows them to continue to conduct attack operations and access sensitive or valuable systems.

# Secure applications with Zero Trust

Article • 03/30/2023



## Background

To get the full benefit of cloud apps and services, organizations must find the right balance of providing access while maintaining control to protect critical data accessed via applications and APIs.

The **Zero Trust** model helps organizations ensure that apps, and the data they contain, are protected by:

- Applying controls and technologies to discover Shadow IT.
- Ensuring appropriate in-app permissions.
- Limiting access based on real-time analytics.
- Monitoring for abnormal behavior.
- Controlling user actions.
- Validating secure configuration options.

## Applications Zero Trust deployment objectives

Before most organizations **start the Zero Trust journey**, their on-premises apps are accessed through physical networks or VPN, and some critical cloud apps are accessible to users.

[ ] Expand table

When implementing a Zero Trust approach to managing and monitoring applications, we recommend you focus first on these **initial deployment objectives**:



- I. Gain visibility into the activities and data in your applications by connecting them via APIs.
- II. Discover and control the use of shadow IT.
- III. Protect sensitive information and activities automatically by implementing policies.

After these are completed, focus on these **additional deployment objectives**:



IV. Deploy adaptive access and session controls for all apps.

V. Strengthen protection against cyber threats and rogue apps.

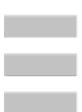
VI. Assess the security posture of your cloud environments

## Application Zero Trust deployment guide

This guide will walk you through the steps required to secure applications and APIs following the principles of a Zero Trust security framework. Our approach is aligned with these three Zero Trust principles:

1. **Verify explicitly.** Always authenticate and authorize based on all available data points, including user identity, location, device health, service or workload, data classification, and anomalies.
2. **Use least privilege access.** Limit user access with Just-In-Time and Just-Enough-Access (JIT/JEA), risk-based adaptive policies and data protection to protect both data and productivity.
3. **Assume breach.** Minimize blast radius for breaches and prevent lateral movement by segmenting access by network, user, devices, and application awareness. Verify all sessions are encrypted end to end. Use analytics to get [visibility](#), drive threat detection, and improve defenses.

  Expand table



## Initial deployment objectives

### I. Gain visibility into the activities and data in your applications by connecting them via APIs

The majority of user activities in an organization originate on cloud applications and associated resources. Most major cloud apps provide an API for consuming tenant information and receiving corresponding governance actions. Use these integrations to monitor and alert when threats and anomalies occur in your environment.

Follow these steps:

1. Adopt [Microsoft Defender for Cloud Apps](#), which works with services to optimize visibility, governance actions, and usage.
2. [Review what apps can be connected](#) with the Defender for Cloud Apps API integration, and connect the apps you need. Use the deeper visibility gained to investigate activities, files, and accounts for the apps in your cloud environment.

 **Tip**

[Learn about implementing an end-to-end identity Zero Trust strategy](#).

## II. Discover and control the use of shadow IT

On average, 1,000 separate apps are being used in your organization. 80 percent of employees use non-sanctioned apps that no one has reviewed and that may not be compliant with your security and compliance policies. And, because your employees are able to access your resources and apps from outside your corporate network, it's no longer enough to have rules and policies on your firewalls.

Focus on identifying app usage patterns, assessing risk levels and business readiness of apps, preventing data leaks to noncompliant apps, and limiting access to regulated data.

 **Tip**

[Learn about implementing an end-to-end Zero Trust strategy for data](#).

Follow these steps:

1. Set up [Cloud Discovery](#), which analyzes your traffic logs against the Microsoft Defender for Cloud Apps catalog of over 16,000 cloud apps. The apps are ranked and scored, based on more than 90 risk factors.
2. [Discover and identify shadow IT](#) to find out what apps are being used, following one of three options:
  - a. Integrate with [Microsoft Defender for Endpoint](#) to immediately start collecting data on cloud traffic across your Windows 10 devices, on and off your network.
  - b. Deploy the [Defender for Cloud Apps log collector](#) on your firewalls and other proxies to collect data from your endpoints and send it to Defender for Cloud

Apps for analysis.

- c. Integrate Defender for Cloud Apps with your proxy.

3. Identify the [risk level](#) of specific apps:

- a. In the Defender for Cloud Apps portal, under Discover, click **Discovered apps**. Filter the list of apps discovered in your organization by the risk factors you are concerned about.
- b. Drill down into the app to understand more about its compliance by clicking the app name and then clicking the **Info** tab to see details about the app's security risk factors.

4. [Evaluate compliance and analyze usage](#):

- a. In the Defender for Cloud Apps portal, under Discover, click **Discovered apps**. Filter the list of apps discovered in your organization by the compliance risk factors you are concerned about. For example, use the suggested query to filter out noncompliant apps.
- b. Drill down into the app to understand more about its compliance by clicking the app name and then clicking the **Info** tab to see details about the app's compliance risk factors.
- c. In the Defender for Cloud Apps portal, under Discover, click **Discovered apps** and then drill down by clicking on the specific app you want to investigate. The **Use** tab lets you know how many active users are using the app and how much traffic it's generating. If you want to see who, specifically, is using the app, you can drill down further by clicking **Total active users**.
- d. [Dive deeper](#) into discovered apps. View subdomains and resources to learn about specific activities, data access, and [resource usage](#) in your cloud services.

5. [Manage your apps](#):

- a. Create new custom app tags in order to classify each app according to its business status or justification. These tags can then be used for specific monitoring purposes.
- b. App tags can be managed under Cloud Discovery settings App tags. These tags can then be used later for filtering in the Cloud Discovery pages and creating policies using them.

- c. Manage discovered apps using [Microsoft Entra Gallery](#). For apps that already appear in the Microsoft Entra Gallery, apply single sign-on and manage the app with Microsoft Entra ID. To do so, on the row where the relevant app appears, choose the three dots at the end of the row, and then choose **Manage app with Microsoft Entra ID**.

 **Tip**

[Learn about implementing an end-to-end Zero Trust strategy for your network ↗](#).

### III. Protect sensitive information and activities automatically by implementing policies

Defender for Cloud Apps enables you to define the way you want users to behave in the cloud. This can be done by creating policies. There are many types: Access, activity, anomaly detection, app discovery, file policy, cloud discovery anomaly detection, and session policies.

Policies allow you to detect risky behavior, violations, or suspicious data points and activities in your cloud environment. They help you monitor trends, see security threats, and generate customized report and alerts.

Follow these steps:

1. [Use out-of-the box policies](#) that have already been tested for many activities and files. Apply governance actions such as revoking permissions and suspending users, quarantining files, and applying sensitivity labels.
2. Build new policies that Microsoft Defender for Cloud Apps suggests for you.
3. Configure policies to monitor shadow IT apps and provide control:
  - a. Create an [app discovery policy](#) that lets you know when there is a spike in downloads or traffic from an app you're concerned about. Enable **Anomalous behavior in discovered users' policy**, **Cloud storage app compliance check**, and **New risky app**.
  - b. Keep updating policies, and using the Cloud Discovery dashboard, check what (new) apps your users are using, as well as their usage and behavior patterns.
4. [Control what's sanctioned](#) and block undesirable apps using this option:
  - a. [Connect apps via API](#) for continuous monitoring.

5. Protect apps using [Conditional Access App Control](#) and [Microsoft Defender for Cloud Apps](#).

 Expand table



## Additional deployment objectives

### IV. Deploy adaptive access and session controls for all apps

Once you've accomplished your initial three objectives, you can focus on additional objectives such as ensuring that all apps are using least-privileged access with continuous verification. Dynamically adapting and restricting access as session risk changes will enable you to stop breaches and leaks in real time, before employees put your data and your organization at risk.

Take this step:

- [Enable real-time monitoring and control over access to any web app](#), based on user, location, device, and app. For example, you can create policies to protect downloads of sensitive content with sensitivity labels when using any unmanaged device. Alternatively, files can be scanned on upload to detect potential malware and block them from entering sensitive cloud environment.



[Learn about implementing an end-to-end Zero Trust strategy for endpoints](#).

### V. Strengthen protection against cyber threats and rogue apps

Bad actors have developed dedicated and unique attack tools, techniques, and procedures (TTPs) that target the cloud to breach defenses and access sensitive and business-critical information. They use tactics such as illicit OAuth consent grants, cloud ransomware, and compromising credentials for cloud identity.

Organizations can respond to such threats with tools available in Defender for Cloud Apps, such as user and entity behavioral analytics (UEBA) and anomaly detection, malware protection, OAuth app protection, incident investigation, and remediation. Defender for Cloud Apps targets numerous [security anomalies](#) out of the box, such as impossible travel, suspicious inbox rules, and ransomware.

The different detections are developed with security operations teams in mind and aim to focus the alerts on true indicators of compromise, while unlocking threat intelligence-driven investigation and remediation.

Follow these steps:

- [Take advantage of the Defender for Cloud Apps UEBA and machine learning \(ML\) capabilities that are automatically enabled out-of-the-box](#) to immediately detect threats and run advanced threat detection across your cloud environment.
- [Tune and scope](#) anomaly detection policies.

## VI. Assess the security posture of your cloud environments

Beyond SaaS applications, organizations are heavily invested in IaaS and PaaS services. Defender for Cloud Apps enables your organization to assess and strengthen your security posture and capabilities for these services by getting visibility into the security configuration and compliance status across your public cloud platforms. This enables a risk-based investigation of the entire platform configuration status.

Follow these steps:

1. Use Defender for Cloud Apps to monitor resources, subscriptions, recommendations, and corresponding severities across your cloud environments.
2. Limit the risk of a security breach by keeping cloud platforms, such as [Microsoft Azure](#), [AWS](#) and [GCP](#), compliant with your organizational configuration policy and regulatory compliance, following CIS benchmark, or the vendor's best practices for the security configuration.
3. Using Defender for Cloud Apps, the security configuration dashboard can be used to drive remediation actions to minimize the risk.

 Tip

Learn about implementing an end-to-end Zero Trust strategy for your infrastructure [↗](#).

## Products covered in this guide

Microsoft Azure

[Microsoft Entra ID ↗](#)

Microsoft 365

[Microsoft Defender for Cloud Apps ↗](#)

[Cloud Discovery](#)

[Microsoft Endpoint Manager ↗](#) (includes Microsoft Intune and Configuration Manager)

[Mobile Application Management](#)

## Conclusion

Regardless of where the cloud resource or application resides, Zero Trust principles help ensure that your cloud environments and data are protected. For further information on these processes or help with these implementations, please contact your Customer Success team.

The Zero Trust deployment guide series



[Introduction](#)



Identity



Endpoints



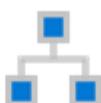
Applications

101010  
010101  
101010

Data



Infrastructure



Networks



**Visibility,  
automation,  
orchestration**

# Azure security solutions for AWS

Azure

Microsoft Sentinel

Microsoft Defender for Cloud Apps

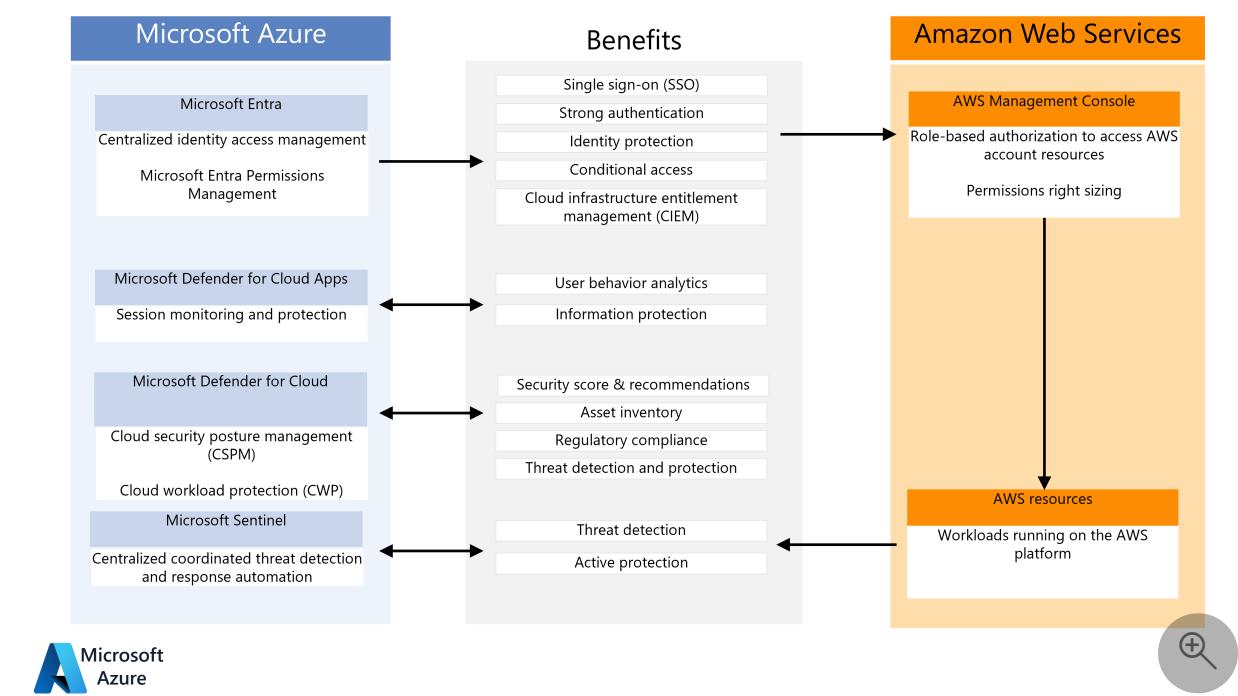
Microsoft Defender for Cloud

This guide shows how Microsoft Defender for Cloud Apps and Microsoft Sentinel can help secure and protect Amazon Web Services (AWS) account access and environments.

AWS organizations that use Microsoft Entra ID for Microsoft 365 or hybrid cloud identity and access protection can quickly and easily [deploy Microsoft Entra ID for AWS accounts](#), often without additional cost.

## Architecture

This diagram summarizes how AWS installations can benefit from key Microsoft security components:



Download a [PowerPoint file](#) of this architecture.

## Workflow

- Microsoft Entra ID provides centralized *single sign-on (SSO)* and strong authentication through *multifactor authentication* and the *conditional access* feature. Microsoft Entra ID supports AWS role-based identities and authorization for access to AWS resources. For more information and detailed instructions, see [Microsoft Entra identity and access management for AWS](#).

Permissions Management is a *cloud infrastructure entitlement management (CIEM)* product that provides comprehensive visibility and control over permissions for any AWS identity or resource. You can use Microsoft Entra Permissions Management to:

- Get a multi-dimensional view of your risk by assessing identities, permissions, and resources.
- Automate the enforcement of the [least privilege](#) policy in your entire multicloud infrastructure.
- Use anomaly and outlier detection to prevent data breaches that are caused by misuse and malicious exploitation of permissions.

For more information and detailed onboarding instructions, see [Onboard an Amazon Web Services \(AWS\) account](#).

- Defender for Cloud Apps:
  - Integrates with the Microsoft Entra Conditional Access feature to enforce additional restrictions.
  - Helps monitor and protect sessions after sign-in.
  - Uses *user behavior analytics (UBA)* and other AWS APIs to monitor sessions and users and to support information protection.
- Microsoft Defender for Cloud displays AWS security recommendations in the Defender for Cloud portal together with Azure recommendations. Defender for Cloud offers more than 160 out-of-the-box recommendations for infrastructure as a service (IaaS) and platform as a service (PaaS) services. It also provides support for regulatory standards, including Center for Internet Security (CIS) and payment card industry (PCI) standards, and for the AWS Foundational Security Best Practices standard. Defender for Cloud also provides cloud workload protection (CWP) for [Amazon EKS clusters](#), [AWS EC2 instances](#), and [SQL servers that run on AWS EC2](#).
- Microsoft Sentinel integrates with Defender for Cloud Apps and AWS to detect and automatically respond to threats. Microsoft Sentinel monitors the AWS environment for misconfiguration, potential malware, and advanced threats to AWS identities, devices, applications, and data.

## Components

- [Microsoft Defender for Cloud Apps](#)
- [Microsoft Defender for Cloud](#)
- [Microsoft Sentinel](#)
- [Microsoft Entra ID](#)

## Defender for Cloud Apps for visibility and control

When several users or roles make administrative changes, a consequence can be *configuration drift* away from intended security architecture and standards. Security standards can also change over time. Security personnel must constantly and consistently detect new risks, evaluate mitigation options, and update security architecture to prevent potential breaches. Security management across multiple public cloud and private infrastructure environments can become burdensome.

Defender for Cloud Apps is a *cloud access security broker (CASB)* platform with *cloud security posture management (CSPM)* capabilities. Defender for Cloud Apps can connect to multiple cloud services and applications to collect security logs, monitor user behavior, and impose restrictions that the platforms themselves might not offer.

Defender for Cloud Apps provides several capabilities that can integrate with AWS for immediate benefits:

- The Defender for Cloud Apps app connector uses several AWS APIs, including UBA, to search for configuration issues and threats on the AWS platform.
- AWS Access Controls can enforce sign-in restrictions that are based on application, device, IP address, location, registered ISP, and specific user attributes.
- Session Controls for AWS block potential malware uploads or downloads based on Microsoft Defender Threat Intelligence or real-time content inspection.
- Session controls can also use real-time content inspection and sensitive data detection to impose *data loss prevention (DLP)* rules that prevent cut, copy, paste, or print operations.

Defender for Cloud Apps is available standalone, or as part of Microsoft Enterprise Mobility + Security E5, which includes Microsoft Entra ID P2. For pricing and licensing information, see [Enterprise Mobility + Security pricing options](#).

## Defender for Cloud for CSPM and CWP platforms (CWPP)

With cloud workloads commonly spanning multiple cloud platforms, cloud security services must do the same. Defender for Cloud helps protect workloads in Azure, AWS, and Google Cloud Platform (GCP).

Defender for Cloud provides an agentless connection to your AWS account. Defender for Cloud also offers plans to secure your AWS resources:

- The [Defender for Cloud overview page](#) displays CSPM metrics, alerts, and insights. Defender for Cloud assesses your AWS resources according to [AWS-specific security recommendations](#) and incorporates your security posture into your secure

score. The [asset inventory](#) provides a single place to view all your protected AWS resources. The [regulatory compliance dashboard](#) reflects the status of your compliance with built-in standards that are specific to AWS. Examples include AWS CIS standards, PCI data security standards (PCI-DSS), and the AWS Foundational Security Best Practices standard.

- Microsoft Defender for Servers brings threat detection and advanced defenses to supported Windows and Linux EC2 instances.
- Microsoft Defender for Containers brings threat detection and advanced defenses to supported Amazon EKS clusters.
- Microsoft Defender for SQL brings threat detection and advanced defenses to your SQL servers that run on AWS EC2 and AWS RDS Custom for SQL Server.

## Microsoft Sentinel for advanced threat detection

Threats can come from a wide range of devices, applications, locations, and user types. DLP requires inspecting content during upload or download, because post-mortem review might be too late. AWS doesn't have native capabilities for device and application management, risk-based conditional access, session-based controls, or inline UBA.

It's critical that security solutions reduce complexity and deliver comprehensive protection regardless of whether resources are in multicloud, on-premises, or hybrid environments. Defender for Cloud provides CSPM and CWP. Defender for Cloud identifies configuration weak spots across AWS to help strengthen your overall security posture. It also helps provide threat protection for Amazon EKS Linux clusters, AWS EC2 instances, and SQL servers in AWS EC2.

[Microsoft Sentinel](#) is a *security information and event management (SIEM) and security orchestration, automation, and response (SOAR)* solution that centralizes and coordinates threat detection and response automation for modern security operations. Microsoft Sentinel can monitor AWS accounts to compare events across multiple firewalls, network devices, and servers. Microsoft Sentinel combines monitoring data with threat intelligence, analytics rules, and machine learning to discover and respond to advanced attack techniques.

You can connect AWS and Defender for Cloud Apps with Microsoft Sentinel. Then you can see Defender for Cloud Apps alerts and run additional threat checks that use multiple Defender Threat Intelligence feeds. Microsoft Sentinel can initiate a coordinated response that's outside Defender for Cloud Apps. Microsoft Sentinel can also integrate with IT service management (ITSM) solutions and retain data on a long-term basis for compliance purposes.

# Scenario details

Microsoft offers several security solutions that can help secure and protect AWS accounts and environments.

Other Microsoft security components can integrate with Microsoft Entra ID to provide additional security for AWS accounts:

- Defender for Cloud Apps backs up Microsoft Entra ID with session protection and user-behavior monitoring.
- Defender for Cloud provides threat protection to AWS workloads. It also helps proactively strengthen security for AWS environments and uses an agentless approach to connect to those environments.
- Microsoft Sentinel integrates with Microsoft Entra ID and Defender for Cloud Apps to detect and automatically respond to threats against AWS environments.

These Microsoft security solutions are extensible and offer multiple levels of protection. You can implement one or more of these solutions along with other types of protection for a full-security architecture that helps protect current and future AWS deployments.

## Potential use cases

This article provides AWS identity architects, administrators, and security analysts with immediate insights and detailed guidance for deploying several Microsoft security solutions.

## Recommendations

Keep the following points in mind when you develop a security solution.

## Security recommendations

The following principles and guidelines are important for any cloud security solution:

- Ensure that the organization can monitor, detect, and automatically protect user and programmatic access into cloud environments.
- Continually review current accounts to ensure identity and permission governance and control.
- Follow least privilege and [zero trust](#) principles. Make sure that users can access only the specific resources that they require, from trusted devices and known

locations. Reduce the permissions of every administrator and developer to provide only the rights that they need for the role that they perform. Review regularly.

- Continuously monitor platform configuration changes, especially if they provide opportunities for privilege escalation or attack persistence.
- Prevent unauthorized data exfiltration by actively inspecting and controlling content.
- Take advantage of solutions that you might already own, like Microsoft Entra ID P2, that can increase security without additional expense.

## Basic AWS account security

To ensure basic security hygiene for AWS accounts and resources:

- Review the AWS security guidance at [Best practices for securing AWS accounts and resources](#).
- Reduce the risk of uploading and downloading malware and other malicious content by actively inspecting all data transfers through the AWS Management Console. Content that you upload or download directly to resources within the AWS platform, such as web servers or databases, might need additional protection.
- Consider protecting access to other resources, including:
  - Resources created within the AWS account.
  - Specific workload platforms, like Windows Server, Linux Server, or containers.
  - Devices that administrators and developers use to access the AWS Management Console.

## Deploy this scenario

Take the steps in the following sections to implement a security solution.

## Plan and prepare

To prepare for deployment of Azure security solutions, review and record current AWS and Microsoft Entra account information. If you've deployed more than one AWS account, repeat these steps for each account.

1. In the [AWS Billing Management Console](#), record the following current AWS account information:
  - **AWS Account ID**, a unique identifier
  - **Account name**, or root user

- **Payment method**, whether assigned to a credit card or a company billing agreement
- **Alternate contacts** who have access to AWS account information
- **Security questions**, securely updated and recorded for emergency access
- **AWS regions** that are enabled or disabled to comply with data security policy

2. In the [Azure portal](#), review the Microsoft Entra tenant:

- Assess **Tenant information** to see whether the tenant has a Microsoft Entra ID P1 or P2 license. A P2 license provides [advanced Microsoft Entra identity management](#) features.
- Assess **Enterprise applications** to see whether any existing applications use the AWS application type, as shown by `http://aws.amazon.com/` in the **Homepage URL** column.

## Deploy Defender for Cloud Apps

After you deploy the central management and strong authentication that modern identity and access management require, you can implement Defender for Cloud Apps to:

- Collect security data and carry out threat detections for AWS accounts.
- Implement advanced controls to mitigate risk and prevent data loss.

To deploy Defender for Cloud Apps:

1. Add a Defender for Cloud Apps app connector for AWS.
2. Configure Defender for Cloud Apps monitoring policies for AWS activities.
3. Create an enterprise application for SSO to AWS.
4. Create a conditional access app control application in Defender for Cloud Apps.
5. Configure Microsoft Entra session policies for AWS activities.
6. Test Defender for Cloud Apps policies for AWS.

## Add an AWS app connector

1. In the [Defender for Cloud Apps portal](#), expand **Investigate** and then select **Connected apps**.
2. On the **App connectors** page, select the **Plus Sign (+)** and then select **Amazon Web Services** from the list.
3. Use a unique name for the connector. In the name, include an identifier for the company and specific AWS account, for example *Contoso-AWS-Account1*.

4. Follow the instructions at [Connect AWS to Microsoft Defender for Cloud Apps](#) to create an appropriate AWS identity and access management (IAM) user.
  - a. Define a policy for restricted permissions.
  - b. Create a service account to use those permissions on behalf of the Defender for Cloud Apps service.
  - c. Provide the credentials to the app connector.

The time it takes to establish the initial connection depends on the AWS account log sizes. When the connection is complete, you see a connection confirmation:

App	Status	Was connected on	Last activity	Accounts	⋮
 Contoso-AWS-Account1 Cloud computing platform	Connected	Oct 5, 2020, 5:3...	Oct 13, 2020, 7:...	465	⋮

## Configure Defender for Cloud Apps monitoring policies for AWS activities

After you turn on the app connector, Defender for Cloud Apps shows new templates and options in the policy configuration builder. You can create policies directly from the templates and modify them for your needs. You can also develop a policy without using the templates.

To implement policies by using the templates:

1. In the Defender for Cloud Apps left navigation window, expand **Control** and then select **Templates**.

# Defender for Cloud Apps



Dashboard

Discover ▾

Investigate ▾

Control ^

Policies

Templates

Alerts | 40

2. Search for aws and review the available policy templates for AWS.

## Policy templates



TYPE	SEVERITY	NAME	CATEGORY	Advanced
Select type...		aws	Select risk category...	
1 - 10 of 10 Templates				
Template	Severity	Linked policies	Published	
Publicly accessible S3 buckets (AWS)		1	Oct 11, 2020, 2:29 A...	
Alert when an S3 bucket in AWS is publicly accessible.				
Virtual Private Network (VPC) changes (AWS)		1	Oct 11, 2020, 2:29 A...	
Alert on any API calls made to create, update, or delete an Amazon VPC, an A...				
IAM Policy changes (AWS)		1	Oct 11, 2020, 2:29 A...	
Alert on any API calls made to change IAM policy				
Console Sign-in Failures (AWS)		1	Oct 11, 2020, 2:29 A...	
Alert of multiple sign-in failures to AWS console.				
CloudTrail changes (AWS)		1	Oct 11, 2020, 2:29 A...	
Alert on any API call made to create, update, or delete a CloudTrail trail, or to ...				
EC2 Instance changes (AWS)		1	Oct 11, 2020, 2:29 A...	
Alert on any API call is made to create, terminate, start, stop, or reboot an Am...				
Network Gateway changes (AWS)		1	Oct 11, 2020, 2:29 A...	
Alert on API call made to create, update, or delete customer's internet gateway.				
Network Access Control List (ACL) changes (AWS)		1	Oct 11, 2020, 2:29 A...	
Alert on any configuration changes involving Network ACLs.				
S3 Bucket Activity (AWS)		1	Oct 11, 2020, 2:29 A...	
Alert when AWS S3 API call is made to PUT or DELETE bucket policy, bucket lif...				
Security Group Configuration changes (AWS)		1	Oct 11, 2020, 2:29 A...	
Alert on configuration changes which involve security groups.				

3. To use a template, select the **Plus Sign (+)** to the right of the template item.
4. Each policy type has different options. Review the configuration settings and save the policy. Repeat this step for each template.

## Create file policy



### Policy template

Publicly accessible S3 buckets (AWS)

### Policy name

Publicly accessible S3 buckets (AWS)

### Description

Alert when an S3 bucket in AWS is publicly accessible.

### Policy severity

Medium

### Category

Sharing control

### Create a filter for the files this policy will act on

FILES MATCHING ALL OF THE FOLLOWING

Edit and preview results

Access level	equals	Public (Internet), Public
App	equals	Amazon Web Services

### Apply to:

all files

### Apply to:

all file owners

### Inspection method

None

To use file policies, make sure the file monitoring setting is turned on in Defender for Cloud Apps settings:

## Files



### Enable file monitoring

This enables Defender for Cloud Apps to see files in your SaaS apps.

As Defender for Cloud Apps detects alerts, it displays them on the **Alerts** page in the Defender for Cloud Apps portal:

The screenshot shows the Microsoft Defender for Cloud Apps interface. At the top, there are filters for Resolution Status (OPEN, DISMISSED, RESOLVED), Category (Select ris...), Severity (Low, Medium, High), APP (Select ap...), User Name (Select us...), and Policy (Select p...). Below the filters, a search bar shows "1 - 6 of 6 alerts". The main area displays a table of alerts:

Alert	Resolution	Severity	Date
CloudTrail changes (AWS) CloudTrail changes (AWS) Contoso-AWS-Account1 aws root user 67.184.79.203 United States	OPEN	Low	10/12/20, ...
S3 Bucket Activity (AWS) S3 Bucket Activity (AWS) Contoso-AWS-Account1 aws root user 67.184.79.203 United States	OPEN	Low	10/12/20, ...
IAM Policy changes (AWS) IAM Policy changes (AWS) Contoso-AWS-Account1 aws root user 67.184.79.203 United States	OPEN	Low	10/12/20, ...
Activity from infrequent country Activity from infrequent co... Microsoft Defender for Cloud Apps Richard 67.184.79.203 United States	OPEN	Medium	10/12/20, ...
Suspicious administrative activity Unusual administrative acti... Contoso-AWS-Account1 aws root user 67.184.79.203	OPEN	Medium	10/9/20, 3...
Block upload of potential malware (based on Microsoft Threat Intelligence) Block upload of potential m... Amazon WorkDocs - General Richard 67.184.79.203 United States	OPEN	High	10/8/20, 8...

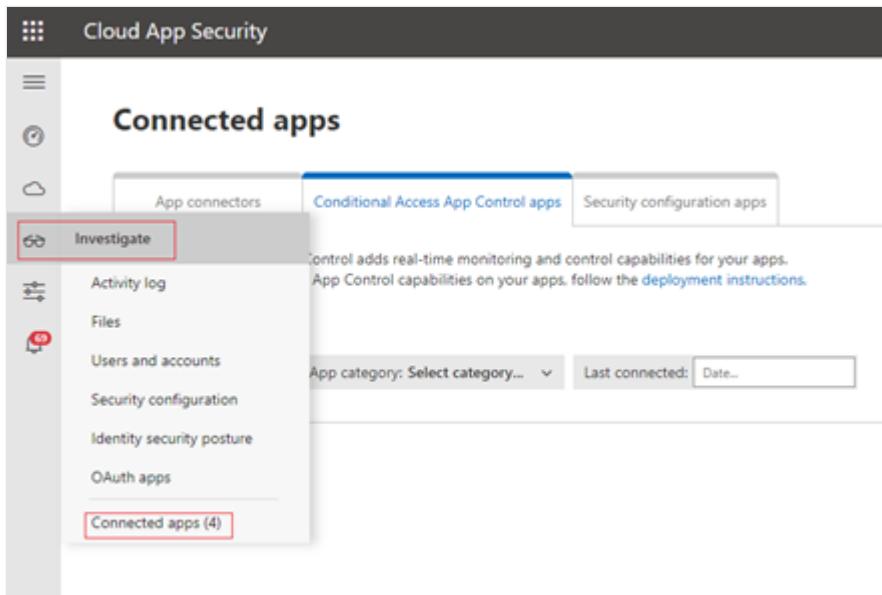
## Create an enterprise application for SSO to AWS

Follow the instructions at [Tutorial: Microsoft Entra single sign-on \(SSO\) integration with AWS single sign-on](#) to create an enterprise application for SSO to AWS. Here's a summary of the procedure:

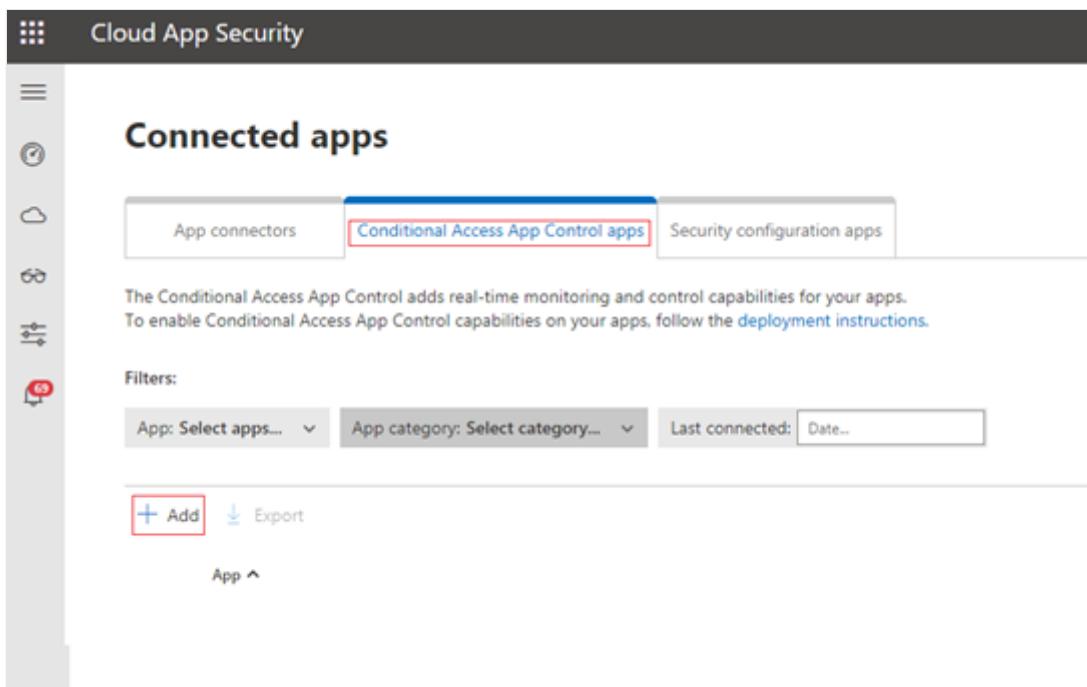
1. Add AWS SSO from the gallery.
2. Configure and test Microsoft Entra SSO for AWS SSO:
  - a. Configure Microsoft Entra SSO.
  - b. Configure AWS SSO.
  - c. Create an AWS SSO test user.
  - d. Test SSO.

## Create a conditional access app control application in Defender for Cloud Apps

1. Go to the [Defender for Cloud Apps portal](#), select **Investigate**, and then select **Connected apps**.



2. Select **Conditional Access App Control apps**, and then select **Add**.



3. In the **Search for an app** box, enter **Amazon Web Services**, and then select the application. Select **Start wizard**.

## Add a SAML application with your identity provider



Add a SAML application with your identity provider to enable real-time monitoring and control

Choose the SAML application that you want to add with your identity provider

Amazon Web Services

⚠ This will add **Amazon Web Services** and all its sub-apps to Conditional Access App Control

Start wizard

We secure your data as described in our [privacy statement](#) and [online service terms](#).

4. Select **Fill in data manually**. Enter the **Assertion consumer service URL** value that's shown in the following screenshot, and then select **Next**.

## Add a SAML application with your identity provider

APP INFORMATION

IDENTITY PROVIDER

APP CHANGES

FINISH

Get the following information from your Amazon Web Services single sign-on configuration

Upload metadata file from the app ⓘ

Fill in data manually ⓘ

Assertion consumer service URL: \* ⓘ

Use Amazon Web Services SAML certificate ⓘ

Next >

Quit ⏮

5. On the next page, ignore the **External configuration** steps. Select **Next**.



## Add a SAML application with your identity provider

APP INFORMATION

IDENTITY PROVIDER

APP CHANGES

FINISH



### External configuration

Perform the following steps in your identity provider's portal, under **Applications**

- 1 Go to your identity provider's portal and [create a new custom SAML app](#) ⓘ
- 2 [Copy the single sign-on configuration](#) of the existing Amazon Web Services app to the new custom app ⓘ
- 3 [Assign users](#) to the new custom app

< Previous

Next >

Quit ▾

6. Select **Fill in data manually**, and then take the following steps to enter the data:
- a. Under **Single sign-on service URL**, enter the **Login URL** value for the enterprise application that you created for AWS.
  - b. Under **Upload identity provider's SAML certificate**, select **Browse**.
  - c. Locate the certificate for the enterprise application that you created.
  - d. Download the certificate to your local device, and then upload it to the wizard.
  - e. Select **Next**.

**LAB1-AWS Single-Account Access | SAML-based Sign-on**

Enterprise Application

**LAB1-AWS Single-Account Access | SAML-based Sign-on**

Enterprise Application

Overview Deployment Plan Manage Properties Owners Roles and administrators (Preview) Users and groups Single sign-on Provisioning Self-service

App Federation Metadata Url [https://login.microsoftonline.com/69e13e16-254d...](https://login.microsoftonline.com/69e13e16-254d-49f5-97d...)

Certificate (Base64) [Download](#)  
Certificate (Raw) [Download](#)  
Federation Metadata XML [Download](#)

Set up LAB1-AWS Single-Account Access

You'll need to configure the application to link with Azure AD.

Login URL [https://login.microsoftonline.com/69e13e16-254d...](https://login.microsoftonline.com/69e13e16-254d-49f5-97d...)  
Azure AD identifier <https://sts.windows.net/69e13e16-254d-49f5-97d...>  
Logout URL <https://login.microsoftonline.com/69e13e16-254d...>

[View step-by-step instructions](#)

Add a SAML application with your identity provider

APP INFORMATION IDENTITY PROVIDER APP CHANGES FINISH

Get the following information from your identity provider's new custom app configuration

Upload metadata file from your identity provider [Browse...](#)

Fill in data manually

Single sign-on service URL: \* <https://login.microsoftonline.com/69e13e16-254d...>

Upload identity provider's SAML certificate: \* [LAB1-AWS Single-Account Access.cer](#) [Browse...](#)

< Previous [Next >](#) Quit

7. On the next page, ignore the **External configuration** steps. Select **Next**.

Add a SAML application with your identity provider

APP INFORMATION IDENTITY PROVIDER APP CHANGES FINISH

**External configuration**  
Perform the following steps in your identity provider's portal, under the **custom app settings**

- Copy the following URL. Go to your identity provider's portal and paste it as the **single sign-on URL** in the new custom SAML app you created: [https://eu2.saml.cas.ms/saml/sso\\_login\\_consumer?orig\\_consumer=https%3A%2F%2Fsignin.aws.amazon.com%2F](https://eu2.saml.cas.ms/saml/sso_login_consumer?orig_consumer=https%3A%2F%2Fsignin.aws.amazon.com%2F)
- Add the following **attributes and values**: [\(1\)](#)

Attribute (case sensitive):	McasSigningCert	Value:	MIIIfjCCA54CAQEcwDQYJKoZ
Attribute (case sensitive):	McasAppId	Value:	2146446671
- Verify that the **name identifier** is in email address format [\(1\)](#)
- Save your settings

< Previous [Next >](#) Quit

8. On the next page, ignore the **External configuration** steps. Select **Finish**.

Add a SAML application with your identity provider

APP INFORMATION    IDENTITY PROVIDER    APP CHANGES    FINISH

 **External configuration**  
Perform the following steps in Amazon Web Services, under [single sign-on settings](#)

- 1 Go to [Amazon Web Services single sign-on](#) settings and create a backup of your current settings (Recommended)
- 2 Copy the following URL and paste it as the **SAML single sign-on URL**: ⓘ  
[https://eu2.saml.cas.ms/saml/sso\\_login?orig\\_idp=https%3A%2F%2Flogin.microsoftonline.com%2F69e13e16-25](https://eu2.saml.cas.ms/saml/sso_login?orig_idp=https%3A%2F%2Flogin.microsoftonline.com%2F69e13e16-25) ⓘ
- 3 Download the [Cloud App Security SAML certificate](#) for the app
- 4 Upload the new certificate to Amazon Web Services instead of the previous identity provider SAML certificate
- 5 Save your settings.  
After you save your changes, all login requests will be routed through Conditional Access App Control (this may take a few minutes).

We secure your data as described in our [privacy statement](#) and [online service terms](#).

[Previous](#) [Finish >](#) [Quit ▾](#)

9. On the next page, ignore the **Verify your settings** steps. Select **Close**.

Add a SAML application with your identity provider

 Congratulations!

You successfully added Amazon Web Services with your identity provider

**Verify your settings**

- 1 Log in to [Amazon Web Services](#). If the login fails, verify that you completed all the wizard steps correctly.
- 2 Make sure your login appears in the [Activity log](#). If not, wait a few minutes and try to log in again.

**What's next?**

- 1 Add a root certificate to identify your [managed devices](#) ⓘ
- 2 Update your [access policies](#) with the new app or create a new [access policy](#) ⓘ

[Previous](#) [Close](#)

## Configure Microsoft Entra session policies for AWS activities

Session policies are a powerful combination of Microsoft Entra Conditional Access policies and the reverse proxy capability of Defender for Cloud Apps. These policies provide real-time suspicious behavior monitoring and control.

1. In Microsoft Entra ID, create a new conditional access policy with the following settings:

- Under **Name**, enter **AWS Console – Session Controls**.
- Under **Users and Groups**, select the two role groups that you created earlier:
  - **AWS-Account1-Administrators**
  - **AWS-Account1-Developers**
- Under **Cloud apps or actions**, select the enterprise application that you created earlier, such as **Contoso-AWS-Account 1**.
- Under **Session**, select **Use Conditional Access App Control**.

2. Under **Enable policy**, select **On**.

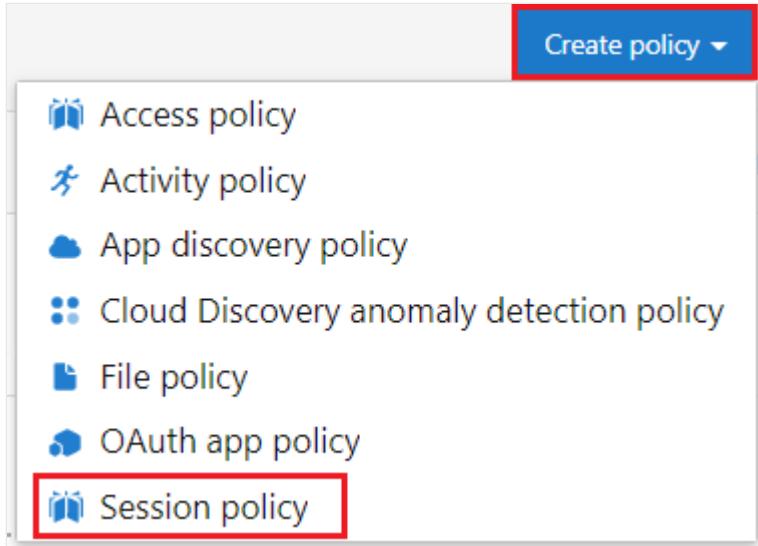
The screenshot shows the Microsoft Entra ID Conditional access policy creation interface. The left pane displays the policy details, including the name 'AWS Console - Session Controls', assignments for users and groups, and access controls for grants and sessions. The right pane shows session control options, with 'Use Conditional Access App Control' selected. A note indicates that this control only works with supported apps (Office 365, Exchange Online, SharePoint Online). Another note mentions custom policies need to be configured in the Cloud App Security portal. A warning at the bottom right states that this policy impacts the Azure AD device registration service, so session controls requiring device registration are not available.

3. Select **Create**.

After you create the Microsoft Entra Conditional Access policy, set up a Defender for Cloud Apps session policy to control user behavior during AWS sessions.

1. In the Defender for Cloud Apps portal, expand **Control** and then select **Policies**.

2. On the **Policies** page, select **Create policy** and then select **Session policy** from the list.



3. On the **Create session policy** page, under **Policy template**, select **Block upload of potential malware (based on Microsoft Threat Intelligence)**.
4. Under **Activities matching all of the following**, modify the activity filter to include **App**, **equals**, and **Amazon Web Services**. Remove the default device selection.

A screenshot of the "Create session policy" configuration screen. The top section is titled "Activity source" with the sub-instruction "Add activity filters to the policy". Below this is a panel titled "ACTIVITIES MATCHING ALL OF THE FOLLOWING". Inside this panel, there is a row of three dropdown menus: the first contains "App", the second contains "equals", and the third contains "Amazon Web Services". To the right of these dropdowns is a link "Edit and preview results". At the bottom of the panel is a small circular button with a plus sign (+).

5. Review the other settings, and then select **Create**.

## Test Defender for Cloud Apps policies for AWS

Test all policies regularly to ensure that they're still effective and relevant. Here are a few recommended tests:

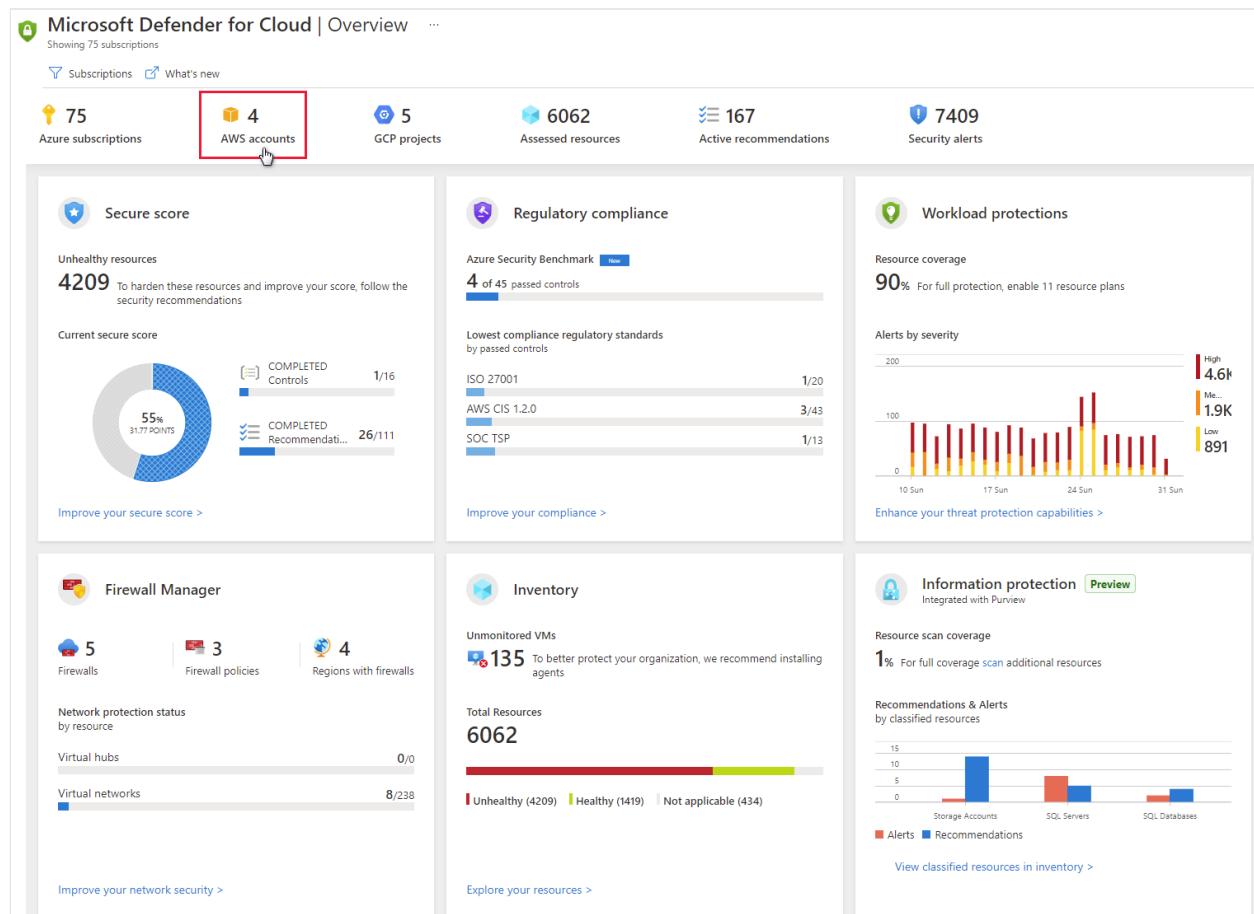
- **IAM policy changes:** This policy is triggered each time that you attempt to modify the settings within AWS IAM. For instance, when you follow the procedure later in this deployment section to create a new IAM policy and account, you see an alert.
- **Console sign-in failures:** Any failed attempts to sign in to one of the test accounts trigger this policy. The alert details show that the attempt came from one of the Azure regional datacenters.

- S3 bucket activity policy: When you attempt to create a new AWS S3 storage account and set it to be publicly available, you trigger this policy.
- Malware detection policy: If you configure malware detection as a session policy, you can test it by following these steps:
  1. Download a safe test file from the [European Institute for Computer Anti-Virus Research \(EICAR\)](#).
  2. Try to upload that file to an AWS S3 storage account.

The policy immediately blocks the upload attempt, and an alert appears in the Defender for Cloud Apps portal.

## Deploy Defender for Cloud

You can use a native cloud connector to connect an AWS account to Defender for Cloud. The connector provides an agentless connection to your AWS account. You can use this connection to gather CSPM recommendations. By using Defender for Cloud plans, you can secure your AWS resources with CWP.



To protect your AWS-based resources, take these steps, which the following sections describe in detail:

1. Connect an AWS account.
2. Monitor AWS.

## Connect your AWS account

To connect your AWS account to Defender for Cloud by using a native connector, follow these steps:

1. Review the [prerequisites](#) for connecting an AWS account. Ensure that you complete them before you proceed.
2. If you have any classic connectors, remove them by following the steps in [Remove classic connectors](#). Using both the classic and native connectors can produce duplicate recommendations.
3. Sign in to the [Azure portal](#).
4. Select **Microsoft Defender for Cloud**, and then select **Environment settings**.
5. Select **Add environment > Amazon Web Services**.

The screenshot shows the Microsoft Defender for Cloud Environment settings page. On the left, there's a navigation sidebar with sections like General, Recommendations, Security alerts, Inventory, Workbooks, Community, Diagnose and solve problems, Cloud Security, Management, and Environment settings (which is currently selected). The main area has a search bar and a 'Welcome' message. It displays three cloud provider icons: AWS (9 accounts), Google Cloud Platform (6 projects), and Microsoft Azure. Below these are filters for 'Search by name', 'Environments == All', 'Standards == All', and 'Coverage == All'. A 'Name ↑' dropdown menu lists 'Azure', 'AWS', and 'GCP', with 'AWS' currently highlighted. The top navigation bar includes links for Home, Microsoft Defender for Cloud, and a 'Showing 53 subscriptions' count.

6. Enter the details of the AWS account, including the storage location of the connector resource. Optionally, select **Management account** to create a connector to a management account. Connectors are created for each member account that's

discovered under the provided management account. Auto-provisioning is turned on for all newly onboarded accounts.

The screenshot shows the 'Add account' wizard in Microsoft Azure (Preview). The current step is 'Account details' (step 1). The page includes fields for 'Connector name' (with a placeholder 'Select a name'), 'Onboard' (radio buttons for 'Management account' and 'Single account' - 'Management account' is selected), 'Subscription' (dropdown set to 'Contoso Hotels Tenant - Production'), 'Resource group' (dropdown with 'Create new' option), 'Location' (dropdown set to 'East US'), 'AWS account Id' (text input placeholder 'Enter Id'), and 'Excluded accounts' (text input placeholder 'Insert accounts to exclude - separated by ","').

## 7. Select Next: Select plans.

The screenshot shows the 'Add account' wizard in Microsoft Azure (Preview). The current step is 'Select plans' (step 2). The page displays a table of available plans:

Plan name & Description	Configurations	Pricing	Plan status
<span>🛡️</span> Security posture management	Permissions: Read (SecurityAudit)	Free (preview)	<input type="button" value="On"/> <input type="button" value="Off"/>
<span>💻</span> Servers	<span>✓</span> Auto-provisioning enabled <a href="#">Configure &gt;</a>	X\$/Server/Month	<input checked="" type="button" value="On"/> <input type="button" value="Off"/>
<span>📦</span> Containers	<span>✓</span> Audit logs enabled Will incur additional AWS costs <small>(i)</small> <a href="#">Configure &gt;</a>	Free (preview)	<input checked="" type="button" value="On"/> <input type="button" value="Off"/>

At the bottom, there are navigation buttons: '< Previous' and 'Next : Configure access >'.

## 8. By default, the servers plan is turned on. This setting is necessary to extend Defender for Servers coverage to your AWS EC2. Ensure you've fulfilled the [network requirements for Azure Arc](#). Optionally, to edit the configuration, select [Configure](#).

9. By default, the containers plan is turned on. This setting is necessary to have Defender for Containers protection for your AWS EKS clusters. Ensure you've fulfilled the [network requirements](#) for the Defender for Containers plan. Optionally, to edit the configuration, select **Configure**. If you disable this configuration, the threat detection feature for the control plane is disabled. To view a list of features, see [Defender for Containers feature availability](#).

10. By default, the databases plan is turned on. This setting is necessary to extend Defender for SQL coverage to your AWS EC2 and RDS Custom for SQL Server. Optionally, to edit the configuration, select **Configure**. We recommend that you use the default configuration.

11. Select **Next: Configure access**.

12. Download the CloudFormation template.

13. Follow the on-screen instructions to use the downloaded CloudFormation template to create the stack in AWS. If you onboard a management account, you need to run the CloudFormation template as Stack and as StackSet. Connectors are created for the member accounts within 24 hours of onboarding.

14. Select **Next: Review and generate**.

15. Select **Create**.

Defender for Cloud immediately starts scanning your AWS resources. Within a few hours, you see security recommendations. For a list of all the recommendations Defender for Cloud can provide for AWS resources, see [Security recommendations for AWS resources - a reference guide](#).

## Monitor your AWS resources

The Defender for Cloud security recommendations page displays your AWS resources. You can use the environments filter to take advantage of the multicloud capabilities of Defender for Cloud, such as viewing the recommendations for Azure, AWS, and GCP resources together.

To view all the active recommendations for your resources by resource type, use the Defender for Cloud asset inventory page. Set the filter to display the AWS resource type that you're interested in.

The screenshot shows the Microsoft Defender for Cloud Inventory interface. On the left, there's a sidebar with navigation links like Overview, Getting started, Recommendations, Security alerts, Inventory (which is selected), Workbooks, Community, and Diagnose and solve problems. The main area displays 'Total resources' (520) and 'Unhealthy' resources (45). A search bar at the top has 'aws' typed into it. Below the search bar is a 'Resource types' filter dialog with fields for Filter (Resource types), Operator (==), and Value (0 selected). The dropdown menu lists various AWS services with their counts: aws ec2 subnet (165), aws ec2 vpc (51), aws ec2 security group (51), aws ec2 network acl (51), aws kms key (6), aws redshift cluster (6), aws account (6), aws iam user (2), aws s3 bucket (2), and aws lambda function (1). The main pane shows a table of resources with columns for Name, Type, Region, Status, and Actions.

## Deploy Microsoft Sentinel

If you connect an AWS account and Defender for Cloud Apps to Microsoft Sentinel, you can use monitoring capabilities that compare events across multiple firewalls, network devices, and servers.

### Enable the Microsoft Sentinel AWS connector

After you enable the Microsoft Sentinel connector for AWS, you can monitor AWS incidents and data ingestion.

As with the Defender for Cloud Apps configuration, this connection requires configuring AWS IAM to provide credentials and permissions.

1. In AWS IAM, follow the steps at [Connect Microsoft Sentinel to AWS CloudTrail](#).
2. To complete the configuration in the Azure portal, under **Microsoft Sentinel > Data connectors**, select the **Amazon Web Services** connector.

Microsoft Azure

Search resources, services, and docs (G+)

Home > Microsoft Sentinel

## Microsoft Sentinel | Data connectors

Selected workspace: 'ispectre-loganalytics-azuresentinel'

Search (Ctrl+ /) Refresh

General

- Overview
- Logs
- News & guides

Threat management

- Incidents
- Workbooks
- Hunting
- Notebooks (Preview)
- Entity behavior (Preview)
- Threat intelligence (Preview)

Configuration

- Data connectors
- Analytics
- Watchlist (Preview)
- Playbooks
- Community
- Settings

60 Connectors    9 Connected    0 Coming soon

Search by name or provider

Providers : All

Status ↑↓	Connector name ↑↓
	AI Vectra Detect (Preview) Vectra AI
	Alcide kAudit (Preview) Alcide
	Amazon Web Services Amazon
	Azure Active Directory Microsoft
	Azure Active Directory Identity Protection Microsoft
	Azure Activity Microsoft
	Azure Advanced Threat Protection (Preview) Microsoft
	Azure DDoS Protection (Preview) Microsoft

3. Select **Open connector page**.
4. Under **Configuration**, enter the **Role ARN** value from the AWS IAM configuration in the **Role to add** field, and select **Add**.
5. Select **Next steps**, and then select the **AWS Network Activities** and **AWS User Activities** activities to monitor.
6. Under **Relevant analytic templates**, select **Create rule** next to the AWS analytic templates that you want to turn on.
7. Set up each rule, and select **Create**.

The following table shows the rule templates that are available for checking AWS entity behaviors and threat indicators. The rule names describe their purpose, and the potential data sources list the data sources that each rule can use.

[] Expand table

Analytic template name	Data sources
Known IRIDIUM IP	DNS, Azure Monitor, Cisco ASA, Palo Alto Networks, Microsoft Entra ID, Azure Activity, AWS
Full Admin policy created and then attached to Roles, Users, or Groups	AWS
Failed AzureAD logons but success logon to AWS Console	Microsoft Entra ID, AWS
Failed AWS Console logons but success logon to AzureAD	Microsoft Entra ID, AWS
Multifactor authentication disabled for a user	Microsoft Entra ID, AWS
Changes to AWS Security Group ingress and egress settings	AWS
Monitor AWS Credential abuse or hijacking	AWS
Changes to AWS Elastic Load Balancer security groups	AWS
Changes to Amazon VPC settings	AWS
New UserAgent observed in last 24 hours	Microsoft 365, Azure Monitor, AWS
Login to AWS Management Console without multifactor authentication	AWS
Changes to internet facing AWS RDS Database instances	AWS
Changes made to AWS CloudTrail logs	AWS

Analytic template name	Data sources
Defender Threat Intelligence map IP entity to AWS CloudTrail	Defender Threat Intelligence Platforms, AWS

Enabled templates have an **IN USE** indicator on the connector details page.

Relevant analytic templates (14)					<a href="#">Go to analytics templates</a>
SEVERITY ↑↓	NAME ↑↓	RULE TYPE ↑↓	DATA SOURCES	TACTICS	
High	<span style="background-color: #800000; color: white; padding: 2px 5px;">IN USE</span> Known IRIDIUM IP	<span style="color: #0078D4;">⌚</span> Scheduled	Office 365 +10 ⓘ	<span style="color: #0078D4;">🛡️</span> Command and ...	
Medium	<span style="background-color: #FFA500; color: white; padding: 2px 5px;">IN USE</span> Full Admin policy created and t...	<span style="color: #0078D4;">⌚</span> Scheduled	Amazon Web Servic...	<span style="color: #0078D4;">💡</span> Privilege Escalat...	
Medium	<span style="background-color: #FFA500; color: white; padding: 2px 5px;">IN USE</span> Failed AzureAD logons but suc...	<span style="color: #0078D4;">⌚</span> Scheduled	Azure Active ... +1 ⓘ	<span style="color: #0078D4;">💻</span> <span style="color: #FFA500;">💻</span>	

## Monitor AWS incidents

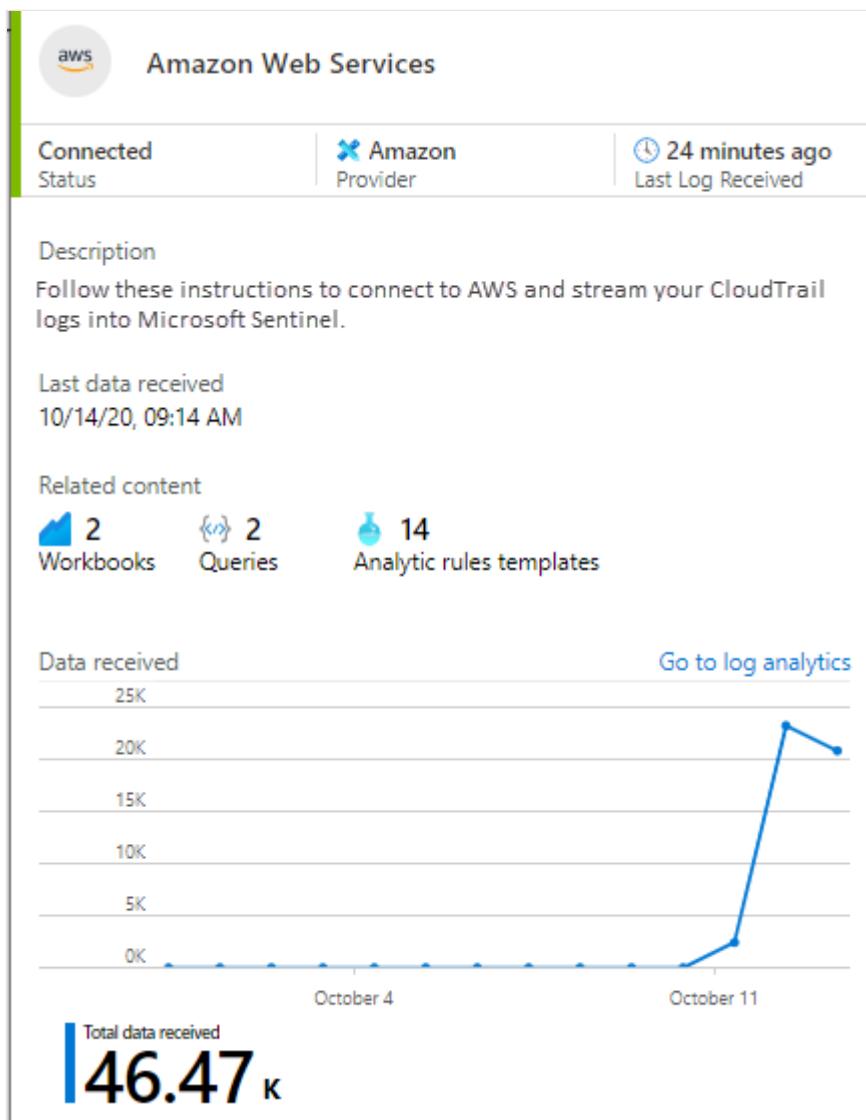
Microsoft Sentinel creates incidents based on the analyses and detections that you turn on. Each incident can include one or more events, which reduces the overall number of investigations that are necessary to detect and respond to potential threats.

Microsoft Sentinel shows incidents that Defender for Cloud Apps generates, if it's connected, and incidents that Microsoft Sentinel creates. The **Product names** column shows the incident source.

Incident id	Title	Alerts	Product names	Created time
145	Login to AWS Management Console...	1	Microsoft Sentinel	10/14/20, 09:51 AM
144	Suspicious administrative activity	1	Microsoft Cloud Ap...	10/14/20, 04:41 AM
143	Console Sign-in Failures (AWS)	1	Microsoft Cloud Ap...	10/14/20, 04:36 AM

## Check data ingestion

Check that data is continuously ingested into Microsoft Sentinel by regularly viewing the connector details. The following chart shows a new connection.



If the connector stops ingesting data and the line chart value drops, check the credentials that you use to connect to the AWS account. Also check that AWS CloudTrail can still collect the events.

## Contributors

*This article is maintained by Microsoft. It was originally written by the following contributor.*

Principal author:

- [Lavanya Murthy](#) | Senior Cloud Solution Architect

*To see non-public LinkedIn profiles, sign in to LinkedIn.*

## Next steps

- For security guidance from AWS, see [Best practices for securing AWS accounts and resources](#).
- For the latest Microsoft security information, see [Microsoft Security](#).
- For full details of how to implement and manage Microsoft Entra ID, see [Securing Azure environments with Microsoft Entra ID](#).
- For an overview of AWS asset threats and corresponding protective measures, see [How Defender for Cloud Apps helps protect your Amazon Web Services \(AWS\) environment](#).
- For information about connectors and how to establish connections, see these resources:
  - [Connect your AWS accounts to Microsoft Defender for Cloud](#)
  - [New AWS connector in Microsoft Defender for Cloud](#)
  - [Connect AWS to Microsoft Defender for Cloud Apps](#)
  - [Connect Microsoft Sentinel to AWS CloudTrail](#)

## Related resources

- For in-depth coverage and comparison of Azure and AWS features, see the [Azure for AWS professionals](#) content set.
- For guidance for deploying Microsoft Entra identity and access solutions for AWS, see [Microsoft Entra identity and access management for AWS](#).

# Security considerations for highly sensitive IaaS apps in Azure

Azure Disk Encryption

Azure Firewall

Azure Key Vault

Microsoft Defender for Cloud

Azure Virtual Network

There are many security considerations for deploying *infrastructure-as-a-service (IaaS)* apps to Azure. This article builds on reference architectures for virtual machine-based workloads and hybrid network infrastructures to focus on security for highly sensitive IaaS workloads in Azure, based on [Azure security fundamentals](#).

Also see [Azure virtual machines security overview](#) and [Security best practices for IaaS workloads in Azure](#).

## Azure VMs

Azure's compute platform is based on machine virtualization. A *hypervisor* runs on the physical hardware of each Azure node or network endpoint, and creates a variable number of guest [Hyper-V virtual machines](#) (VMs) in the node. All user code executes on the VMs. For basic Azure VM deployment instructions, see [Run a Linux VM on Azure](#) or [Run a Windows VM on Azure](#). Most deployment processes are the same for the two operating systems (OSs), but OS-specific tools like disk encryption may differ.

You can use [Microsoft Defender for Cloud](#) for VM patch management and to deploy and monitor [antimalware tools](#). Alternatively, you can manage your own or third-party patching and antimalware tools, which is common when extending or migrating existing infrastructures to Azure.

Microsoft provides Basic *distributed denial of service (DDoS)* protection as part of the Azure platform. Apps that have public endpoints can use Standard [Azure DDoS Protection](#) for additional protection. However, highly sensitive workloads don't usually have public endpoints, and can only be accessed from specific locations over a *virtual private network (VPN)* or leased line.

## N-tier architectures

Many IaaS applications consist of multiple tiers, such as a web tier, business tier, and data tier, which are hosted across multiple VMs. Key aspects of deploying *n-tier* app architectures on Azure VMs include:

- **High availability (HA).** HA apps must be available more than 99.9% of the time. Placing in-tier VMs in different Azure [availability zones](#) (AZs) ensures HA, because AZs span one or more datacenters, providing resiliency through resistance to datacenter failure. Regions that don't support AZs can use [availability sets](#) (ASs), which distribute VMs across multiple isolated hardware nodes.
- **Load balancing.** [Load balancers](#) distribute traffic among VMs, to balance load and for resiliency when a VM fails. You don't need load balancers if the application manages load balancing and the individual VMs are known to the caller.
- **Virtual networks.** [Virtual networks](#) and subnets segment your network, enabling easier security management and advanced routing.
- **Domain Name System (DNS).** [Azure DNS](#) provides a highly available and secure DNS service. A [private zone](#) in Azure DNS lets you use custom domains inside your virtual networks.

## Backup and restore

To protect against human error, malicious data deletion, and ransomware, you should back up at least your data-tier VMs. [Azure Backup](#) can [back up and restore encrypted VMs](#) if it can access the encryption keys in Azure Key Vault.

For the web and business tiers, you can use [virtual machine scale set](#) autoscaling rules to automatically [destroy compromised VMs](#) and deploy fresh VM instances from a base image.

## Compute isolation

On each Azure node or network endpoint, the hypervisor and a special root OS ensure guest VMs can't access the physical host server, and user code executes only on guest VMs. This isolation prevents users from obtaining raw read, write, or execute access to the system, and mitigates the risk of sharing resources. Azure protects against all known *side-channel attacks* and *noisy neighbors* through the hypervisor and an advanced VM placement algorithm. For more information, see [Compute isolation](#).

For highly sensitive workloads, you can add additional protection against side-channel attacks with [isolated VMs](#) or [dedicated hosts](#).

## Isolated VMs

Isolated VMs are large VM sizes that are isolated to a specific hardware type and dedicated to a single customer. Using an isolated VM size guarantees that your VM is

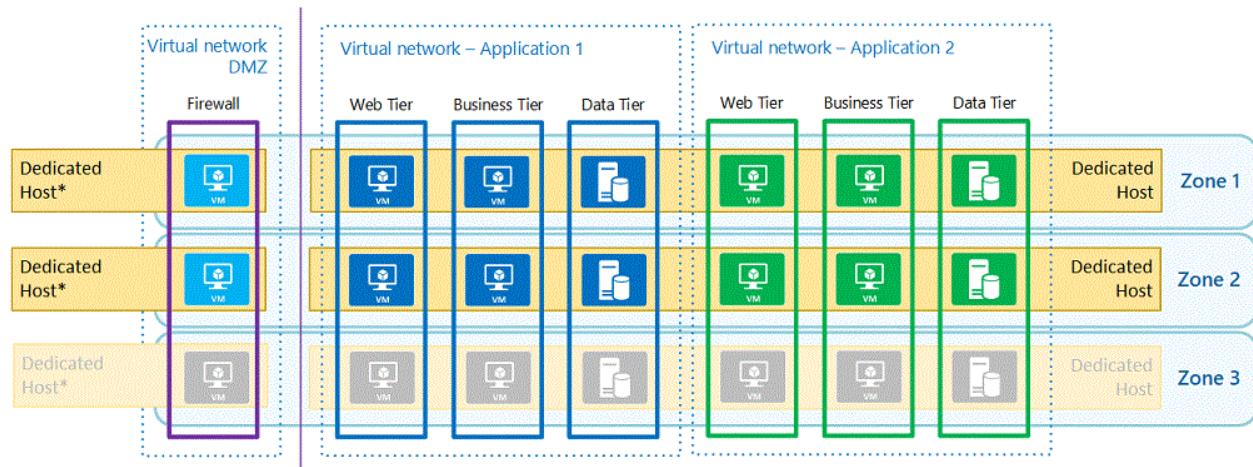
the only one running on a specific server instance. You can further subdivide the resources of isolated VMs by using [Azure support for nested virtual machines](#).

The minimum size of an isolated VM is 64 virtual CPU cores and 256 GiB of memory. These VMs are far larger than most n-tier applications require, and can create a large cost overhead. To reduce the overhead, you can run multiple app tiers on a single VM with nested virtualization, or in different processes or containers. You still need to deploy different VMs in AZs for resiliency, and run [demilitarized zone \(DMZ\) appliances](#) on separate VMs. Combining multiple apps on one infrastructure for economic reasons might also conflict with organizational app segregation policies.

As Azure region capabilities expand over time, Azure may also remove isolation guarantees from certain VM sizes, with one year's notice.

## Azure Dedicated Hosts

[Azure Dedicated Host](#) is the preferred compute isolation solution for highly sensitive workloads. A dedicated host is a physical server dedicated to one customer for hosting multiple VMs. Besides isolating VMs, dedicated hosts let you control [maintenance](#) and VM placement to avoid noisy neighbors.



Dedicated hosts have the same minimum size and many of the same sizing considerations as isolated VMs. However, a dedicated host can host VMs located in different virtual networks, to satisfy application segregation policies. You should still run [DMZ](#) VMs on a different host, to prevent any side-channel attack from a compromised VM in the DMZ.

## Encryption

Data encryption is an important component of securing workloads. Encryption encodes information so only authorized receivers can decode it by using a key or certificate.

Encryption includes *disk encryption*, for data encryption-at-rest, and *Transport Level Security (TLS)*, for encryption-in-transit over networks.

## Azure Key Vault

You can protect encryption keys and certificates by storing them in [Azure Key Vault](#), a cloud *Hardware Security Module (HSM)* solution validated for Federal Information Processing Standards (FIPS) 140-2 Level 2. For best practices to allow only authorized applications and users to access Key Vault, see [Secure access to a key vault](#).

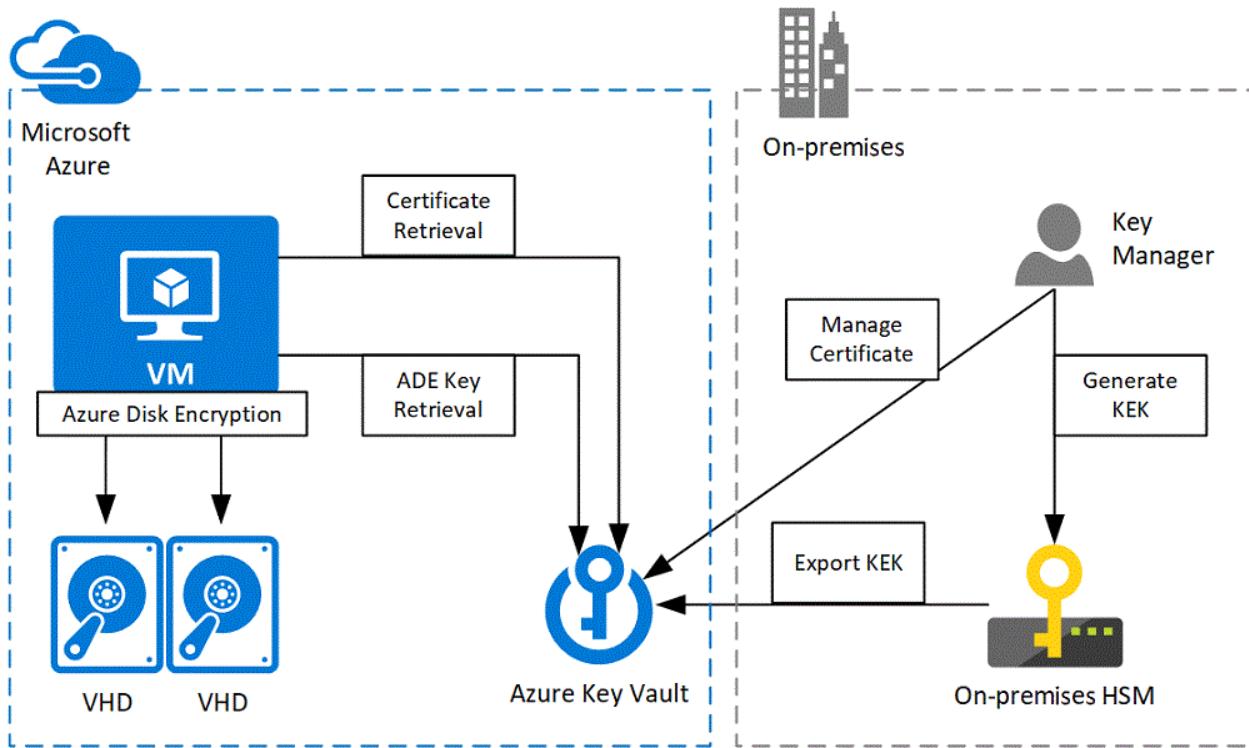
To protect keys in Key Vault, you can enable [soft delete](#), which ensures that deleted keys are recoverable. For additional protection, you can [back up individual keys](#) to an encrypted file that you can use to [restore the keys](#), potentially to another Azure region in the same geography.

When hosting SQL Server on a VM, you can use the [SQL Server Connector for Microsoft Azure Key Vault](#) to get keys for *transparent data encryption (TDE)*, *column level encryption (CLE)*, and backup encryption. For details, see [Configure Azure Key Vault integration for SQL Server on Azure virtual machines](#).

## Azure Disk Encryption

Azure Disk Encryption uses a BitLocker external key protector to provide volume encryption for the OS and data disks of Azure VMs, and can be integrated with Azure Key Vault to help you control and manage disk encryption keys and secrets. Each VM generates its own encryption keys and stores them in Azure Key Vault. To configure Azure Key Vault to enable Azure Disk Encryption, see [Create and configure a key vault for Azure Disk Encryption](#).

For highly sensitive workloads, you should also use a *key encryption key (KEK)* for an additional layer of security. When you specify a KEK, Azure Disk Encryption uses that key to wrap the encryption secrets before writing to Key Vault. You can generate a KEK in Azure Key Vault, but a more secure method is to generate a key in your on-premises HSM and import it to Azure Key Vault. This scenario is often referred to as *bring your own key*, or BYOK. Because the imported keys can't leave the HSM boundary, generating the key in your HSM ensures you're in full control of the encryption keys.



For more information about HSM-protected keys, see [How to generate and transfer HSM-protected keys for Azure Key Vault](#).

## Network traffic encryption

Network protocols like HTTPS encrypt data in transit with certificates. Client-to-application traffic usually uses a certificate from a trusted *certificate authority (CA)*. Internal apps can use a certificate from an internal CA or a public CA like DigiCert or GlobalSign. Tier-to-tier communication typically uses a certificate issued by an internal CA, or a self-signed certificate. Azure Key Vault can accommodate any of these certificate types. For more information about creating different certificate types, see [Certificate creation methods](#).

Azure Key Vault can act as a self-signed certificate CA for tier-to-tier traffic. The *Key Vault VM extension* provides monitoring and automatic refresh of specified certificates on VMs, with or without the private key depending on use case. To use the Key Vault VM extension, see [Key Vault virtual machine extension for Linux](#) or [Key Vault virtual machine extension for Windows](#).

Key Vault can also store keys for network protocols that don't use certificates. Custom workloads could require scripting a [custom script extension](#) that retrieves a key from Key Vault and stores it for applications to use. Apps can also use a VM's [managed identity](#) to retrieve secrets directly from Key Vault.

## Network security

**Network security groups** (NSGs) filter traffic between resources in Azure virtual networks. NSG security rules allow or deny network traffic to or from Azure resources based on IP addresses and ports. By default, NSGs block inbound traffic from the internet, but allow outbound connections from VMs to the internet. To prevent accidental outbound traffic, add a custom rule with the lowest possible priority, 4096, to block all inbound and outbound traffic. You can then add higher-priority rules to allow specific traffic.

NSGs create flow records for existing connections, and allow or deny communication based on the flow record's connection state. The flow record allows an NSG to be stateful. For example, if you specify an outbound security rule to any address over port 443, it's not necessary to also specify an inbound security rule for the response. You only need to specify an inbound security rule if the communication is initiated externally.

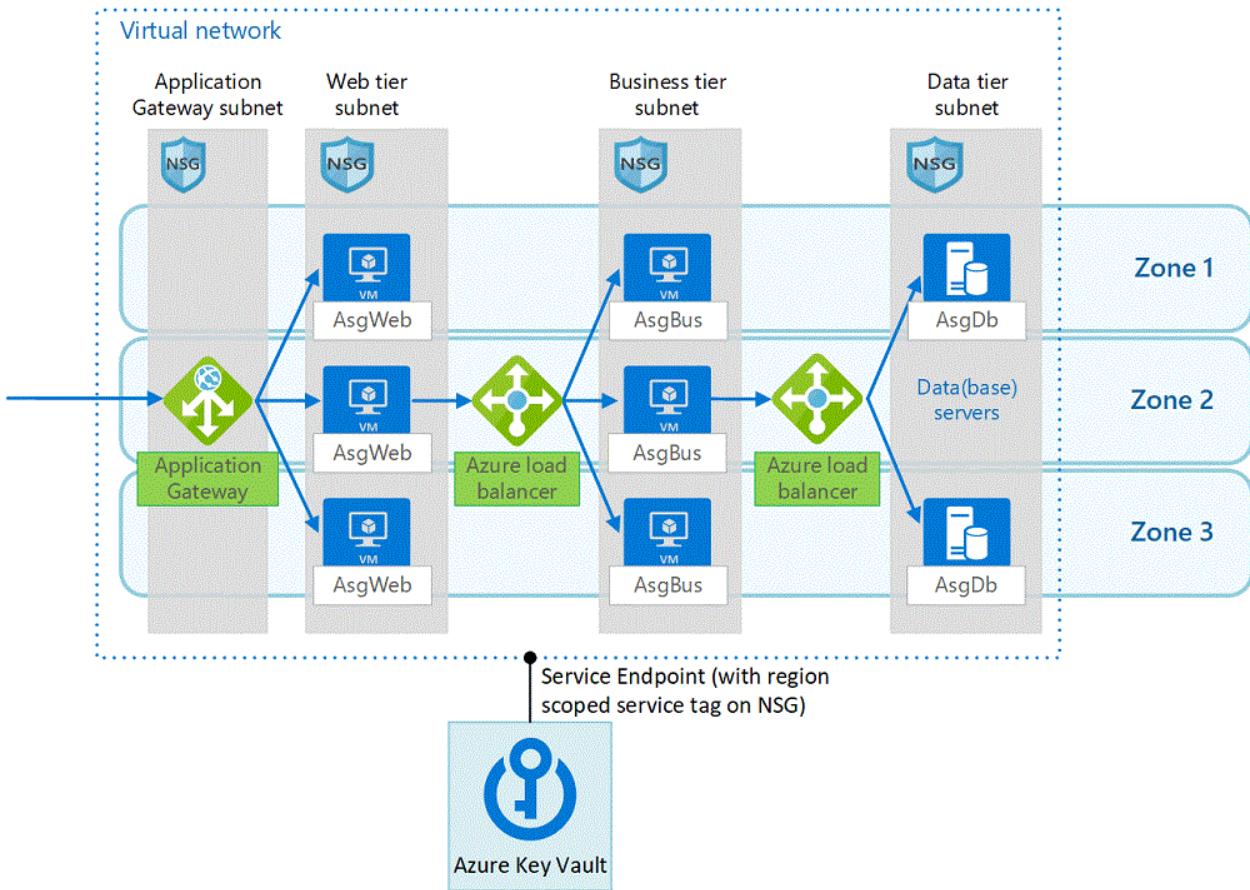
Most Azure services allow you to use a [virtual network service tag](#) instead of an NSG. A service tag represents a group of IP address prefixes from an Azure service, and helps minimize complexity from frequent network security rule updates. An Azure Key Vault service tag can allow a VM to retrieve certificates, keys, and secrets from Azure Key Vault.

Another way to control network security is through [virtual network traffic routing](#) and *forced tunneling*. Azure automatically creates system routes and assigns the routes to each subnet in a virtual network. You can't create or remove system routes, but you can override some system routes with custom routes. Custom routing lets you route traffic over a *network virtual appliance* (NVA) like a firewall or proxy, or drop unwanted traffic, which has a similar effect to blocking the traffic with an NSG.

You can use NVAs like [Azure Firewall](#) to allow, block, and inspect network traffic. Azure Firewall is a managed, highly available platform firewall service. You can also deploy third-party NVAs from the [Azure Marketplace](#). To make these NVAs highly available, see [Deploy highly available network virtual appliances](#).

## Application security groups

To filter traffic between application tiers within a virtual network, use [Application security groups](#) (ASGs). ASGs let you configure network security as an extension of an application's structure, letting you group VMs and define network security policies based on the groups. You can reuse your security policy at scale without manually maintaining explicit IP addresses.



Since ASGs are applied to a network interface instead of a subnet, they enable micro-segmentation. You can tightly control which VMs can talk to each other, even blocking traffic between VMs in the same tier, and making it easy to quarantine a VM by removing ASGs from that VM.

## Hybrid networks

Hybrid architectures connect on-premises networks with public clouds like Azure. There are several ways to connect on-premises networks to applications running in Azure:

- **Public endpoint over the internet.** You can rely on identity, transport level security (HTTPS), and the application gateway to protect the application, potentially in combination with a firewall. However, for highly sensitive workloads, exposing a public endpoint over the internet isn't recommended.
- **Azure or third-party VPN gateway.** You can connect an on-premises network to Azure by using an [Azure VPN gateway](#). Traffic still travels over the internet, but over an encrypted tunnel that uses TLS. You can also run a third-party gateway in a VM, if you have specific requirements not supported by the Azure VPN gateway.
- **ExpressRoute.** [ExpressRoute](#) connections use a private, dedicated connection through a third-party connectivity provider. The private connection extends your on-premises network into Azure, and provides scalability and a reliable service-level agreement (SLA).

- [ExpressRoute with VPN failover](#). This option uses ExpressRoute in normal conditions, but fails over to a VPN connection if there's a loss of connectivity in the ExpressRoute circuit, providing higher availability.
- [VPN over ExpressRoute](#). This option is the best for highly sensitive workloads. ExpressRoute provides a private circuit with scalability and reliability, and VPN provides an additional layer of protection that terminates the encrypted connection in a specific Azure virtual network.

For more guidance on choosing between different types of hybrid connectivity, see [Choose a solution for connecting an on-premises network to Azure](#).

## Deploy a DMZ

Connecting on-premises and Azure environments gives on-premises users access to Azure applications. A perimeter network or *demilitarized zone (DMZ)* provides additional protection for highly sensitive workloads.

An architecture like the one in [Network DMZ between Azure and an on-premises datacenter](#) deploys all DMZ and application services in the same virtual network, with NSG rules and user-defined routes to isolate the DMZ and application subnets. This architecture can make the management subnet available via public internet, to manage apps even if the on-premises gateway isn't available. However, for highly sensitive workloads, you should only allow bypassing the gateway in a [break glass scenario](#). A better solution is to use [Azure Bastion](#), which enables access directly from the Azure portal while limiting exposure of public IP addresses.

You can also use [Just-In-Time \(JIT\) VM access](#) for remote management while limiting exposure of public IP addresses. With JIT VM access, an NSG blocks remote management ports like *remote desktop protocol (RDP)* and *secure shell (SSH)* by default. Upon request, JIT VM access enables the port only for a specified time window, and potentially for a specific IP address or range. JIT access also works for VMs that have only private IP addresses. You can use Azure Bastion to block traffic to a VM until JIT VM access is enabled.

To deploy more applications, you can use a [hub-spoke network topology](#) in Azure, with the DMZ in the hub virtual network and the applications in spoke virtual networks. The hub virtual network can contain a VPN and/or ExpressRoute gateway, firewall NVA, management hosts, identity infrastructure, and other shared services. The spoke virtual networks are connected to the hub with [virtual network peering](#). An Azure virtual network doesn't allow transitive routing over the hub from one spoke to another. Spoke-to-spoke traffic is only possible via the firewall appliances in the hub. This architecture effectively isolates applications from one another.

# Multi-region deployment

Business continuity and disaster recovery might require deploying your application across multiple Azure regions, which can impact data residency and security. For a reference architecture for multi-region deployments, see [Run an N-tier application in multiple Azure regions for high availability](#).

## Regional pairs

An Azure geography is a defined area of the world that contains at least one Azure region, each with one or more datacenters. Each Azure region is paired with another region in the same geography in a *regional pair*. Regional pairs aren't both updated at the same time, and if a disaster hits both regions, one of the regions is prioritized to come back online first. For business continuity, you should deploy highly sensitive apps at least to regional pairs if you deploy in multiple regions.

For more details, see [Business continuity and disaster recovery \(BCDR\): Azure Paired Regions](#). The whitepaper [Achieve compliant data residency and security with Azure](#) ↗ discusses data residency, and what to do to meet data residency requirements.

## Replication between regions

In IaaS architectures, replicating data between regions is the responsibility of the application. The most common replication scenario uses database replication technologies built into the database server product, such as [SQL Server Always On Availability Groups](#), [Oracle Data Guard](#) ↗, or [MySQL Replication](#) ↗.

Setting up replication between IaaS database servers isn't straightforward, and you need to take business continuity requirements into account. Azure database services such as [Azure SQL Database](#), [Azure Database for MySQL](#), and [Azure Cosmos DB](#) make replication between regions easier, but may not meet security requirements for highly sensitive workloads.

For more information and guidance for multi-region SQL Server and Oracle deployments, see:

- [Configure an availability group on Azure virtual machines running SQL Server in different regions](#)
- [Disaster recovery for an Oracle Database 12c database in an Azure environment](#)

## Cross-region peering

You can enable secure communication between virtual networks in different regions by using global [virtual network peering](#). Global peering works the same as within-region peering. The traffic between regions runs over the Microsoft backbone, doesn't traverse the internet, and is isolated from other traffic. For more security, you can deploy VPN NVAs in both regions, and use *user-defined routes* to force traffic between regions over the NVAs, similar to [deploying a DMZ](#).

## Failover traffic routing

With public endpoints, you can use [Traffic Manager](#) or [Azure Front Door](#) to direct traffic to the active region or closest region in an *active-active* failover configuration. However, Traffic Manager and Azure Front Door both require public endpoints to monitor availability, and their corresponding DNS entries are public. For highly sensitive workloads, the alternative solution is to deploy DNS on-premises, and change the entries to the active region for failover.

## Management and governance

Securing your highly sensitive IaaS apps requires more than just deploying the correct architectures and implementing network security rules. Because cloud environments are easily changed, it's especially important to ensure changes can be made only with certain permissions, and within the boundaries of security policies. For example, you must prevent a malicious actor from being able to change a network security rule to allow traffic from the internet.

To deploy workloads in Azure, you need one or more *management accounts*. Securing management accounts is critical to securing your workloads. For more information, see [Secure privileged access for hybrid and cloud deployments in Microsoft Entra ID](#).

Use the resources in your management subnet to grant app tier access only to people who need to manage that tier. For example, you can use [Microsoft Identity Manager](#) with Microsoft Entra ID. However, for cloud-native scenarios, Microsoft Entra [Privileged Identity Management](#) (PIM) is preferred.

There are several other ways to control Azure roles and policies:

- [Azure role-based access control \(Azure RBAC\)](#) for Azure resources lets you assign built-in or custom roles to users, so they have only the privileges they need. You can combine Azure RBAC with PIM to implement an audited approval workflow that elevates privileges for a limited time period.
- Policies enforce corporate rules, standards, and SLAs. [Azure Policy](#) is an Azure service that creates, assigns, and manages policies, and evaluates your resources

for policy compliance.

- [Azure Blueprints](#) combine role assignments, policy assignments, and deployment templates to define a set of replicable Azure resources that implement and follow an organization's standards, patterns, and requirements. Blueprints are a declarative way to orchestrate the deployment of resource templates and other artifacts. You can create blueprints yourself, or leverage existing blueprints. For example, the [ISO 27001 Shared Services blueprint](#) deploys a shared services hub that you can modify and extend to your organization's requirements.

## Monitoring

[Microsoft Defender for Cloud](#) provides monitoring and alerts that help you maintain security of your environment. The free service automatically checks for vulnerabilities such as missing OS patches, security misconfiguration, and basic network security. The Standard paid version gives you additional features, such as [behavioral analytics](#), [Adaptive Network Hardening](#), and [JIT VM access](#). For a full list of features, see [Feature coverage for machines](#). Defender for Cloud also provides [threat protection](#) for other resources like Azure Key Vault.

You can use [Azure Monitor](#) for further monitoring and analysis. To monitor identity and access, you can [route Microsoft Entra activity logs to Azure Monitor](#). You can also monitor [VMs](#), [networks](#), and [Azure Firewall](#), and analyze imported logs with powerful [log query](#) capability. You can integrate Azure Monitor with your *Security Information and Event Manager (SIEM)*, which can be a [third-party SIEM](#) or [Microsoft Sentinel](#).

## Related resources

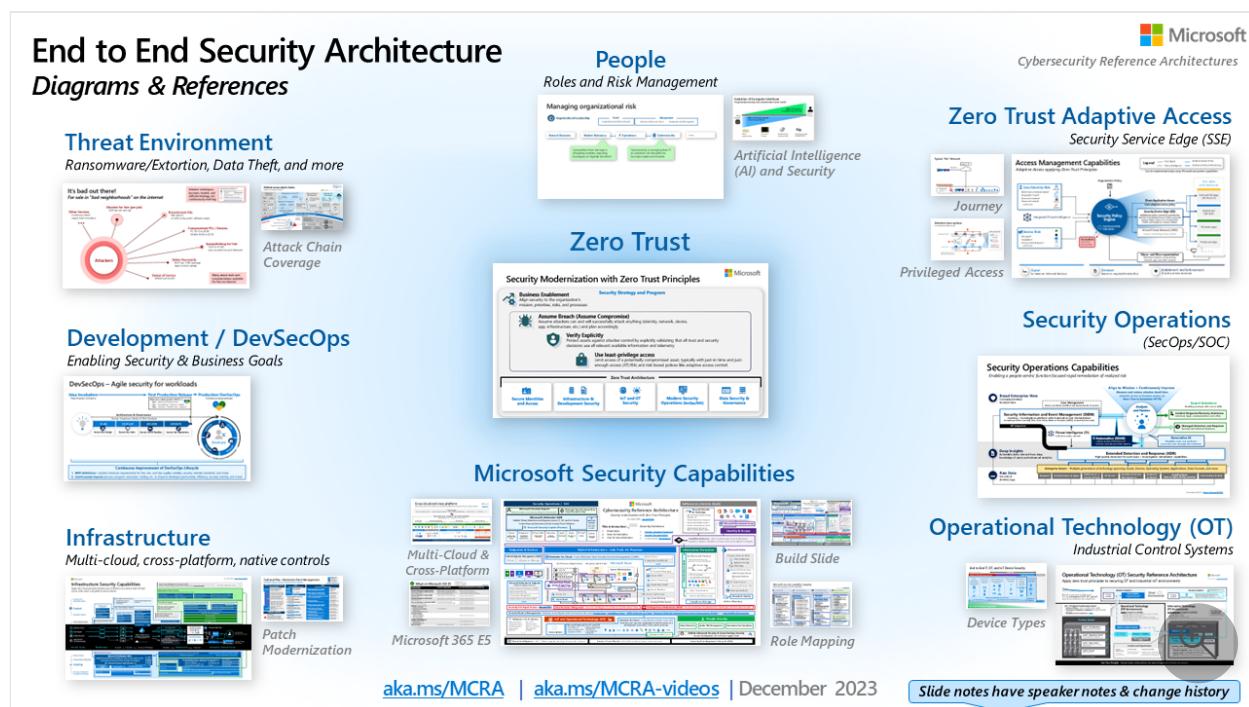
- For more information about n-tier architectures, see [Linux n-tier application in Azure with Apache Cassandra](#).
- For an end-to-end tutorial on using the Azure Key Vault virtual machine extension, see [Secure a web server on a Windows virtual machine in Azure with SSL certificates stored in Key Vault](#).
- For more information about Azure Disk Encryption, see [Azure Disk Encryption for Linux VMs](#) or [Azure Disk Encryption for Windows VMs](#).
- For more information about Azure network security, see [Azure network security overview](#).

# Microsoft Cybersecurity Reference Architectures

Article • 12/01/2023

The Microsoft Cybersecurity Reference Architectures (MCRA) are the component of [Microsoft's Security Adoption Framework \(SAF\)](#) that describe Microsoft's cybersecurity capabilities and technologies. The diagrams describe how Microsoft security capabilities integrate with Microsoft platforms and third party platforms like:

- Microsoft 365
- Microsoft Azure
- Third party apps like ServiceNow and Salesforce
- Third party platforms like Amazon Web Services (AWS) and Google Cloud Platform (GCP)
- First and third party AI capabilities



Download the updated December 2023 version of the MCRA [↗](#)

## What does the MCRA include?

The MCRA includes key information about:

- Antipatterns (common mistakes) and best practices
- Guiding rulesets for end to end architecture
- Threat trends, and attack patterns

- Mapping Microsoft capabilities to organizational roles
- Mapping Microsoft capabilities to Zero Trust standards
- Securing privileged access
- Reference plans in SAF (including example of patching modernization)
- Prioritizing using attacker return on investment (ROI)
- ...and more

The MCRA also includes detailed technical diagrams for:

- Microsoft cybersecurity capabilities
- Zero trust user access
- Security operations (SecOps/SOC)
- Operational technology (OT)
- Multicloud and cross-platform capabilities
- Attack chain coverage
- Infrastructure and Development Security
- Security organizational functions

## How to use the MCRA

We see this resource used for several purposes including

- **Starting template for a security architecture** - The most common use case we see is that organizations use the document to help define a target state for cybersecurity capabilities. Organizations find this architecture useful because it covers capabilities across the modern enterprise estate that now spans on-premises, mobile devices, multiple clouds, and IoT / Operational Technology.
- **Comparison reference for security capabilities** - Some organizations use this resource to compare Microsoft's recommendations with what they already own and have implemented. Many organizations find that they already own quite a bit of this technology already and weren't aware of it.
- **Learn about Microsoft capabilities** - We also see this resource used as a learning tool. In presentation mode, each capability has a "ScreenTip" with a short description of each capability + a link to documentation to learn more.
- **Learn about Microsoft's integration investments** - The architecture helps architects and technical teams identify how to take advantage of integration points within Microsoft capabilities and with existing security capabilities.
- **Learn about Cybersecurity** - Some folks, particularly people new to cybersecurity, use this resource as a learning tool as they prepare for their first career or a career change.

## Next Steps

- Download the updated December 2023 version of the MCRA ↗
- Watch a prerecorded version of this guidance on YouTube ↗.
- Download the December 2021 version of the MCRA ↗
- Continue your journey as part of the [Security Adoption Framework](#).

# Use Azure monitoring to integrate security components

Azure   Azure Monitor   Office 365   Microsoft Defender for Office 365

This article introduces a series of articles about how to integrate security services with your IT environment to protect its systems and resources, on site and in the cloud. Microsoft provides a wide range of security services to help your organization monitor and protect your systems and information. This series of articles describes how to integrate these services with your IT environment to improve its security posture.

Microsoft offers many documents and reference architectures about IT security. For example, you can learn about Zero Trust concepts, understand how Microsoft Defender XDR services work to protect your Office environment, and get an architectural design with various security services from Microsoft Azure Cloud. You can find a compilation of various security-oriented reference architectures on [Microsoft Cybersecurity Reference Architectures](#).

## Architectures in this series

This article is the first in a series of five articles that present a logical and organized way to understand and integrate the security solutions that are available from Microsoft Azure public cloud and from Microsoft 365 services. This first article provides an overview of the series. It briefly explains the content of the architecture and how it was built. The other articles in this series explain each part in more detail.

This series explores in depth the defense that you can build with these Microsoft cloud security services:

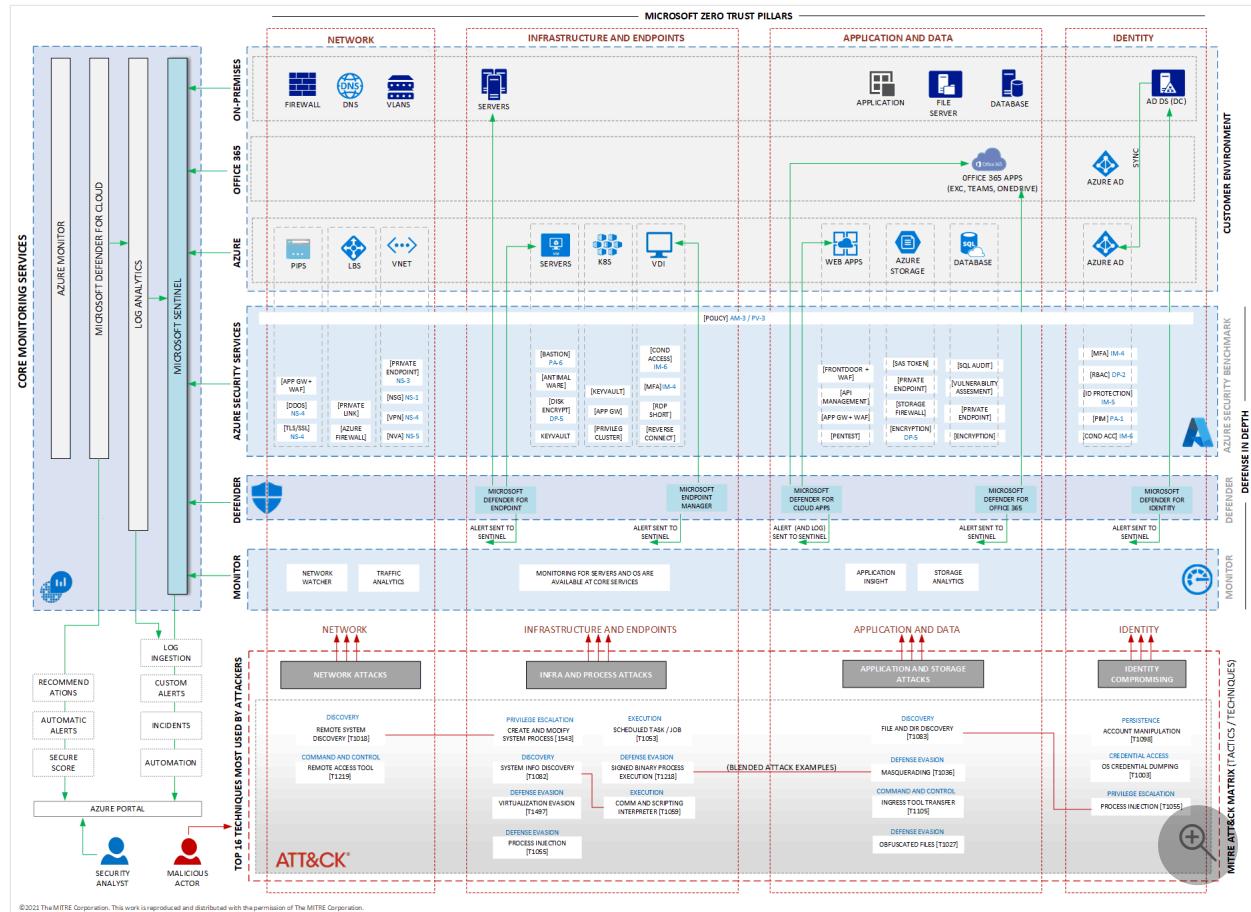
- Azure security services
- Microsoft Defender XDR Services
- Azure Monitor services

## Diagrams

This series of articles uses architectural diagrams to explain how Microsoft security services work together. The diagram in this article is the final architecture reference for this series, and it presents the whole picture.

To make the architecture more comprehensive, it was designed to be layered onto the architecture of a typical hybrid IT environment, which in many companies has three layers:

- On-premises services, such as a private datacenter
- Office 365 services that provide Microsoft Office apps
- Azure public cloud services, including servers, storage, and identity services



Download a [Visio file](#) of this architecture.

©2021 The MITRE Corporation. This work is reproduced and distributed with the permission of The MITRE Corporation.

At the bottom of the diagram is a layer that represents some of the most familiar techniques of attack according to MITRE ATT&CK matrix ([MITRE ATT&CK®](#)) and the tactics involved (in blue text). From a threat perspective, malicious actors have evolved with new technologies and scenarios, especially public and hybrid clouds.

## Articles

In addition to this introductory article, this series includes the following articles:

- Map threats to your IT environment

The second article in this series explores how you can use this architectural reference with a different set of tactics and techniques or with varying methodologies, like [the Cyber Kill Chain® ↗](#), a framework developed by Lockheed Martin.

- [Build the first layer of defense with Azure Security services](#)

The third article in this series explores in detail the security services of Microsoft's cloud services. It describes how to protect Azure services, like virtual machines, storage, network, application, database, and other Azure services.

- [Build the second layer of defense with Microsoft Defender XDR Security services](#)

The fourth article in this series explores security for Microsoft 365 services, like Office 365, Teams, and OneDrive, provided by Microsoft Defender XDR services.

- [Integrate Azure and Microsoft Defender XDR security services](#)

The fifth article in this series explains the relationship between Azure Security and Microsoft Defender XDR services and their integration. It describes how integration works and how you can accomplish it by using Microsoft Sentinel and Log Analytics, which are shown on the left side of the architecture diagram. This series calls these *core monitoring services*, because the services that are depicted in the graph can work with the comprehensive services of Azure and Microsoft 365.

## Contributors

*This article is maintained by Microsoft. It was originally written by the following contributors.*

Principal author:

- [Rudnei Oliveira ↗](#) | Senior Customer Engineer

Other contributors:

- [Gary Moore ↗](#) | Programmer/Writer
- [Andrew Nathan ↗](#) | Senior Customer Engineering Manager

## Next steps

This document refers to some services, technologies, and terminologies. You can find more information about them in the following resources:

- What are public, private, and hybrid clouds? [↗](#)
- Overview of the Azure Security Benchmark (v3)
- Embrace proactive security with Zero Trust [↗](#)
- Microsoft 365 [↗](#) subscription information
- Microsoft Defender XDR [↗](#)
- MITRE ATT&CK® [↗](#)
- The Cyber Kill Chain® [↗](#) from Lockheed Martin
- What is Microsoft Sentinel?
- Overview of Log Analytics in Azure Monitor

## Related resources

For more details about this reference architecture, see the other articles in this series:

- Part 2: Map threats to your IT environment
- Part 3: Build the first layer of defense with Azure Security services
- Part 4: Build the second layer of defense with Microsoft Defender XDR Security services
- Part 5: Integrate Azure and Microsoft Defender XDR security services

# Firewall and Application Gateway for virtual networks

Azure Application Gateway

Azure Firewall

Azure Front Door

Azure Virtual Network

Azure Web Application Firewall

To secure Azure application workloads, you should use protective measures, such as authentication and encryption, in the applications themselves. You can also add security layers to the virtual networks (VNets) that host the applications. These security layers protect the application's inbound flows from unintended utilization. They also limit outbound flows to the internet to only those endpoints your application requires. This article describes [Azure Virtual Network](#) security services like Azure DDoS Protection, Azure Firewall and Azure Application Gateway, when to use each service, and network design options that combine both.

- [Azure DDoS Protection](#), combined with application-design best practices, provides enhanced DDoS mitigation features to provide more defense against DDoS attacks. You should enable [Azure DDOS Protection](#) on any perimeter virtual network.
- [Azure Firewall](#) is a managed next-generation firewall that offers [network address translation \(NAT\)](#). Azure Firewall bases packet filtering on Internet Protocol (IP) addresses and Transmission Control Protocol and User Datagram Protocol (TCP/UDP) ports, or on application-based HTTP(S) or SQL attributes. Azure Firewall also applies Microsoft threat intelligence to identify malicious IP addresses. For more information, see the [Azure Firewall documentation](#).
- [Azure Firewall Premium](#) includes all functionality of Azure Firewall Standard and other features, such as TLS-inspection and Intrusion Detection and Protection System (IDPS).
- [Azure Application Gateway](#) is a managed web traffic load balancer and HTTP(S) full reverse proxy that can do Secure Socket Layer (SSL) encryption and decryption. Application gateway preserves the original client IP address in an X-Forwarded-For HTTP header. Application Gateway also uses Web Application Firewall to inspect web traffic and detect attacks at the HTTP layer. For more information, see the [Application Gateway documentation](#).
- [Azure Web Application Firewall \(WAF\)](#) is an optional addition to Azure Application Gateway. It provides inspection of HTTP requests, and it prevents malicious attacks at the web layer, such as SQL Injection or Cross-Site Scripting. For more information, see the [Web Application Firewall documentation](#).

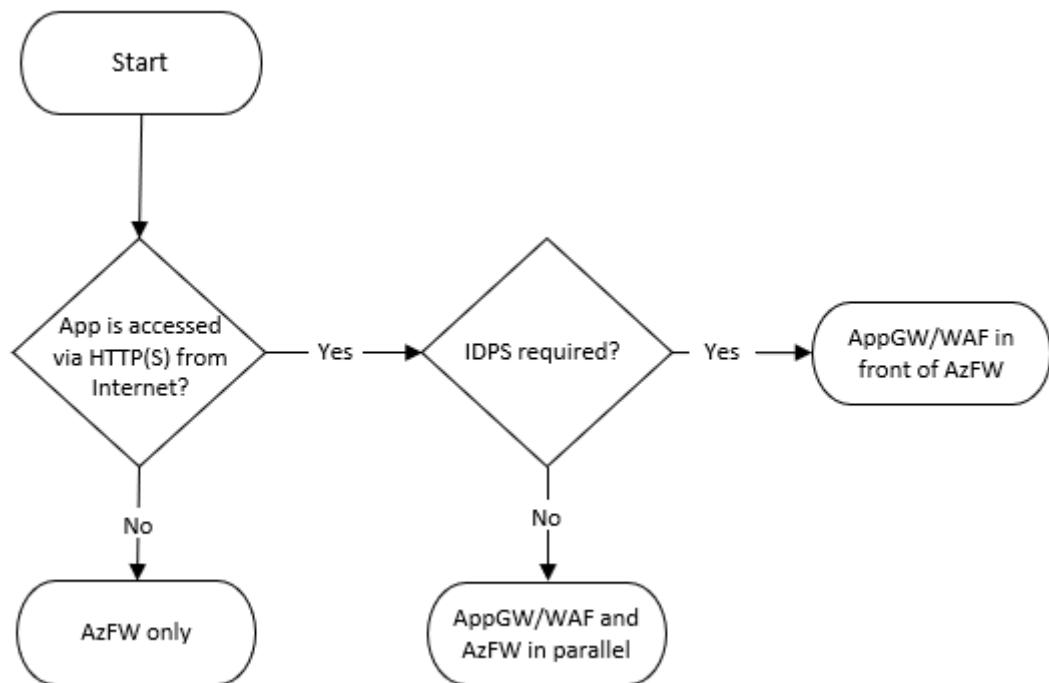
These Azure services are complementary. One or the other may be best for your workloads, or you can use them together for optimal protection at both the network and application layers. Use the following decision tree and the examples in this article to determine the best security option for your application's virtual network.

Azure Firewall and Azure Application Gateway use different technologies, and they support securitization of different flows:

[\[+\] Expand table](#)

Application Flow	Can be filtered by Azure Firewall	Can be filtered by WAF on Application Gateway
HTTP(S) traffic from on-premises/internet to Azure (inbound)	Yes	Yes
HTTP(S) traffic from Azure to on-premises/internet (outbound)	Yes	No
Non-HTTP(S) traffic, inbound/outbound	Yes	No

Depending on the network flows an application requires, the design can be different on a per-application basis. The following diagram offers a simplified decision tree that helps choosing the recommended approach for an application. The decision depends on whether the application is published via HTTP(S) or some other protocol:



This article will cover the widely recommended designs from the flow chart, and others that are applicable in less common scenarios:

- [Azure Firewall alone](#), when there are no web applications in the virtual network. It will control both inbound traffic to the applications and outbound traffic.
- [Application Gateway alone](#), when only web applications are in the virtual network, and [network security groups \(NSGs\)](#) provide sufficient output filtering. The functionalities provided by Azure Firewall can prevent many attack scenarios (such as data exfiltration, IDPS, and so on). Due to this reason, the Application Gateway alone scenario isn't typically recommended and hence not documented and is not in the flow chart above.
- [Azure Firewall and Application Gateway in parallel](#), which is one of the most common designs. Use this combination when you want Azure Application Gateway to protect HTTP(S) applications from web attacks, and Azure Firewall to protect all other workloads and filter outbound traffic.
- [Application Gateway in front of Azure Firewall](#), when you want Azure Firewall to inspect all traffic, WAF to protect web traffic, and the application to know the client's source IP address. With Azure Firewall Premium and TLS inspection, this design supports the end-to-end SSL scenario as well.
- [Azure Firewall in front of Application Gateway](#), when you want Azure Firewall to inspect and filter traffic before it reaches the Application Gateway. Because the Azure Firewall isn't going to decrypt HTTPS traffic, the functionality that it's adding to the Application Gateway is limited. This scenario isn't documented in the flow chart above.

In the last part of this article, variations of the previous fundamental designs are described. These variations include:

- On-premises application clients.
- Hub and spoke networks.
- Azure Kubernetes Service (AKS) implementations.

You can add other reverse proxy services like an [API Management](#) gateway or [Azure Front Door](#). Or you can replace the Azure resources with third-party [network virtual appliances](#).

 **Note**

In the following scenarios an Azure virtual machine is used as an example of web application workload. The scenarios are also valid to other workload types such as containers or Azure Web Apps. For setups including private endpoints please consider the recommendations in [Use Azure Firewall to inspect traffic destined to a private endpoint](#)

## Azure Firewall only

If there are no web-based workloads in the virtual network that can benefit from WAF, you can use Azure Firewall only. The design in this case is simple, but reviewing the packet flow will help understand more complex designs. In this design, all inbound traffic is sent to the Azure Firewall via user defined routes (UDRs) for connections from on-premises or other Azure VNets. It is addressed to the Azure Firewall's public IP address for connections from the public internet, as the diagram below shows. Outbound traffic from Azure VNets is sent to the Firewall via UDRs, as shown in the dialog below.

The following table summarizes the traffic flows for this scenario:

 [Expand table](#)

Flow	Goes through Application Gateway / WAF	Goes through Azure Firewall
HTTP(S) traffic from internet/onprem to Azure	N/A	Yes (see below)

Flow	Goes through Application Gateway / WAF	Goes through Azure Firewall
HTTP(S) traffic from Azure to internet/onprem	N/A	Yes
Non-HTTP(S) traffic from internet/onprem to Azure	N/A	Yes
Non-HTTP(S) traffic from Azure to internet/onprem	N/A	Yes

Azure Firewall won't inspect inbound HTTP(S) traffic. But it will be able to apply layer 3 & layer 4 rules and FQDN-based application rules. Azure Firewall will inspect outbound HTTP(S) traffic depending on the Azure Firewall tier and whether you configure TLS inspection:

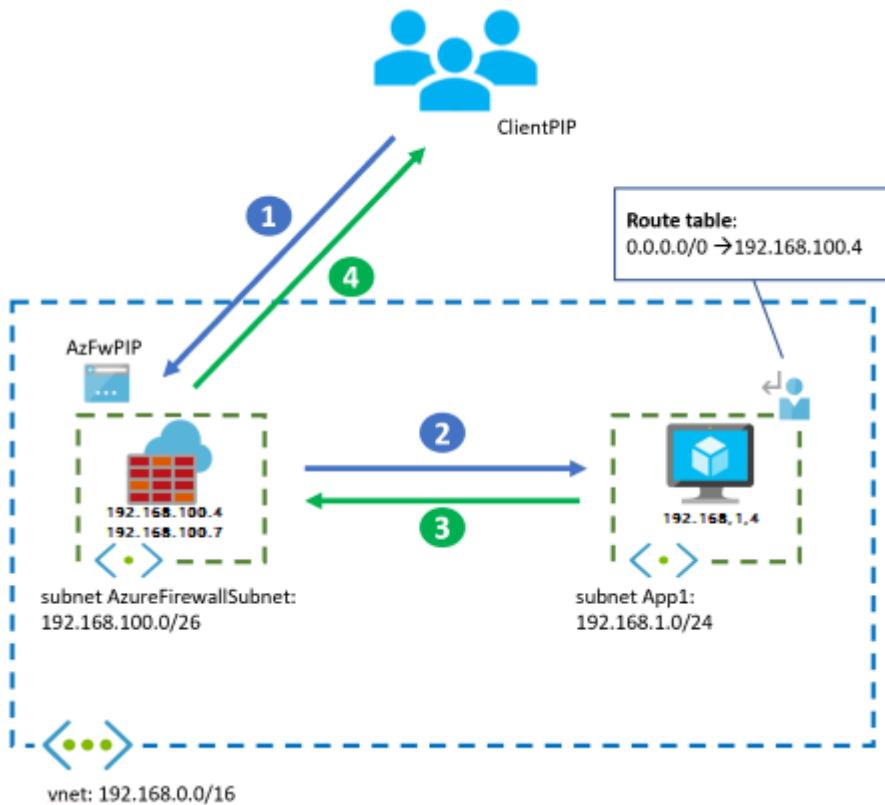
- Azure Firewall Standard will only inspect layer 3 & layer 4 attributes of the packets in network rules, and the Host HTTP header in application rules.
- Azure Firewall Premium adds capabilities such as inspecting other HTTP headers (such as the User-Agent) and enabling TLS inspection for deeper packet analysis. Azure Firewall isn't equivalent to a Web Application Firewall. If you have web workloads in your Virtual Network, using WAF is highly recommended.

The following packet walk example shows how a client accesses a VM-hosted application from the public internet. The diagram includes only one VM for simplicity. For higher availability and scalability, you'd have multiple application instances behind a load balancer. In this design, Azure Firewall inspects both incoming connections from the public internet, and outbound connections from the application subnet VM by using the UDR.

- Azure Firewall service deploys several instances under the covers, here with the front-end IP address 192.168.100.4 and internal addresses from the range 192.168.100.0/26. These individual instances are normally invisible to the Azure administrator. But noticing the difference is useful in some cases, such as when troubleshooting network issues.
- If traffic comes from an on-premises virtual private network (VPN) or [Azure ExpressRoute](#) gateway instead of the internet, the client starts the connection to the VM's IP address. It doesn't start the connection to the firewall's IP address, and the firewall will do no Source NAT per default.

# Architecture

The following diagram shows the traffic flow assuming the instance IP address is 192.168.100.7.



## Workflow

1. The client starts the connection to the public IP address of the Azure Firewall:
  - Source IP address: ClientPIP
  - Destination IP address: AzFwPIP
2. The request to the Azure Firewall public IP is distributed to a back-end instance of the firewall, in this case 192.168.100.7. The Azure Firewall [Destination NAT \(DNAT\) rule](#) translates the destination IP address to the application IP address inside the virtual network. The Azure Firewall also [Source NATs \(SNATs\)](#) the packet if it does DNAT. For more information, see [Azure Firewall known issues](#). The VM sees the following IP addresses in the incoming packet:
  - Source IP address: 192.168.100.7
  - Destination IP address: 192.168.1.4
3. The VM answers the application request, reversing source and destination IP addresses. The inbound flow doesn't require a [user-defined route \(UDR\)](#), because the source IP is Azure Firewall's IP address. The UDR in the diagram for 0.0.0.0/0 is

for outbound connections, to make sure packets to the public internet go through the Azure Firewall.

- Source IP address: 192.168.1.4
- Destination IP address: 192.168.100.7

4. Finally, Azure Firewall undoes the SNAT and DNAT operations, and delivers the response to the client:

- Source IP address: AzFwPIP
- Destination IP address: ClientPIP

## Application Gateway only

This design covers the situation where only web applications exist in the virtual network, and inspecting outbound traffic with NSGs is sufficient to protect outbound flows to the internet.

### Note

This is not a recommended design since using Azure Firewall to control outbound flows (instead of only NSGs) will prevent certain attack scenarios such as data exfiltration, where you make sure that your workloads are only sending data to an approved list of URLs. Further, NSGs only work on layer 3 and layer 4 and have no FQDN support.

The main difference from the previous design with only the Azure Firewall is that the Application Gateway doesn't act as a routing device with NAT. It behaves as a full reverse application proxy. That is, Application Gateway stops the web session from the client, and establishes a separate session with one of its backend servers. Inbound HTTP(S) connections from the Internet need to be sent to the public IP address of the Application Gateway, connections from Azure or on-premises to the gateway's private IP address. Return traffic from the Azure VMs will follow standard VNet routing back to the Application Gateway (see the packet walk further down for more details). Outbound internet flows from Azure VMs will go straight to the internet.

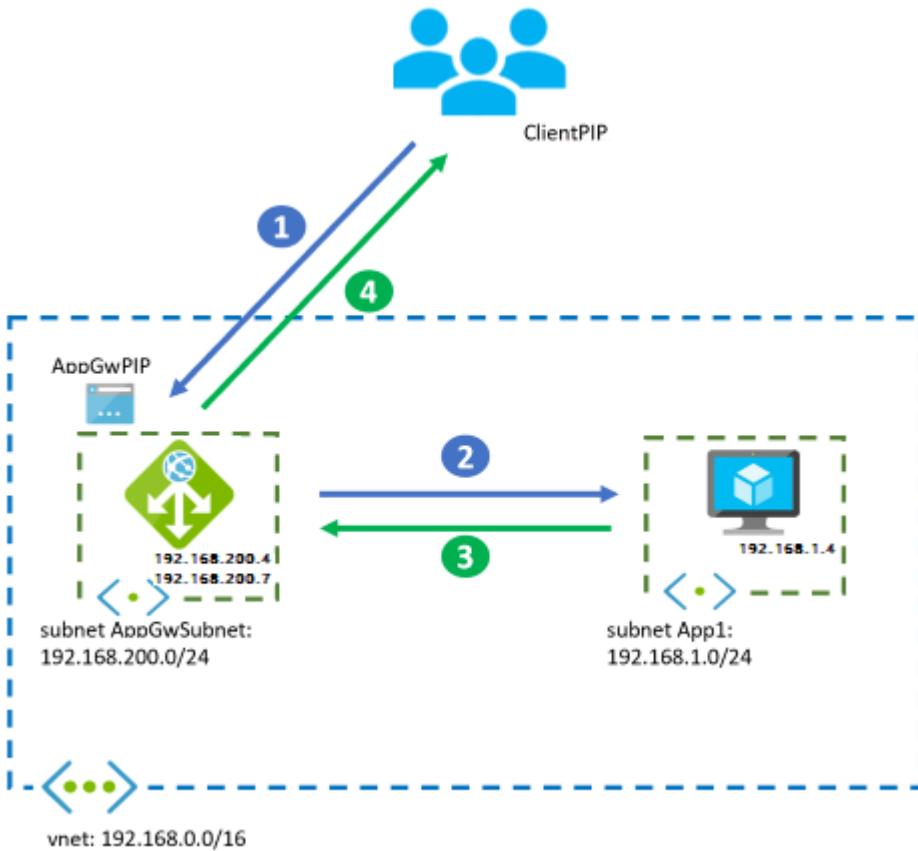
The following table summarizes traffic flows:

 Expand table

Flow	Goes through Application Gateway / WAF	Goes through Azure Firewall
HTTP(S) traffic from internet/onprem to Azure	Yes	N/A
HTTP(S) traffic from Azure to internet/onprem	No	N/A
Non-HTTP(S) traffic from internet/onprem to Azure	No	N/A
Non-HTTP(S) traffic from Azure to internet/onprem	No	N/A

## Architecture

The following packet walk example shows how a client accesses the VM-hosted application from the public internet.



## Workflow

1. The client starts the connection to the public IP address of the Azure Application Gateway:

- Source IP address: ClientPIP
- Destination IP address: AppGwPIP

2. The request to the Application Gateway public IP is distributed to a back-end instance of the gateway, in this case 192.168.200.7. The Application Gateway instance that receives the request stops the connection from the client, and establishes a new connection with one of the back ends. The back end sees the Application Gateway instance as the source IP address. The Application Gateway inserts an *X-Forwarded-For* HTTP header with the original client IP address.

- Source IP address: 192.168.200.7 (the private IP address of the Application Gateway instance)
- Destination IP address: 192.168.1.4
- X-Forwarded-For header: ClientPIP

3. The VM answers the application request, reversing source and destination IP addresses. The VM already knows how to reach the Application Gateway, so doesn't need a UDR.

- Source IP address: 192.168.1.4
- Destination IP address: 192.168.200.7

4. Finally, the Application Gateway instance answers the client:

- Source IP address: AppGwPIP
- Destination IP address: ClientPIP

Azure Application Gateway adds metadata to the packet HTTP headers, such as the *X-Forwarded-For* header containing the original client's IP address. Some application servers need the source client IP address to serve geolocation-specific content, or for logging. For more information, see [How an application gateway works](#).

- The IP address 192.168.200.7 is one of the instances the Azure Application Gateway service deploys under the covers, here with the internal, private front-end IP address 192.168.200.4. These individual instances are normally invisible to the Azure administrator. But noticing the difference is useful in some cases, such as when troubleshooting network issues.
- The flow is similar if the client comes from an on-premises network over a VPN or ExpressRoute gateway. The difference is the client accesses the private IP address of the Application Gateway instead of the public address.

## ⓘ Note

See [Preserve the original HTTP host name between a reverse proxy and its back-end web application](#) for more information on X-Forwarded-For and preserving the host name on a request.

# Firewall and Application Gateway in parallel

Because of its simplicity and flexibility, running Application Gateway and Azure Firewall in parallel is often the best scenario.

Implement this design if there's a mix of web and non-web workloads in the virtual network. Azure WAF in Azure Application Gateway protects inbound traffic to the web workloads, and the Azure Firewall inspects inbound traffic for the other applications. The Azure Firewall will cover outbound flows from both workload types.

Inbound HTTP(S) connections from the Internet should be sent to the public IP address of the Application Gateway, HTTP(S) connections from Azure or on-premises to its private IP address. Standard VNet routing will send the packets from the Application Gateway to the destination VMs, as well as from the destination VMs back to the Application Gateway (see the packet walk further down for more details). For inbound non-HTTP(S) connections, traffic should be targeting the public IP address of the Azure Firewall (if coming from the public Internet), or it will be sent through the Azure Firewall by UDRs (if coming from other Azure VNets or on-premises networks). All outbound flows from Azure VMs will be forwarded to the Azure Firewall by UDRs.

The following table summarizes the traffic flows for this scenario:

[+] Expand table

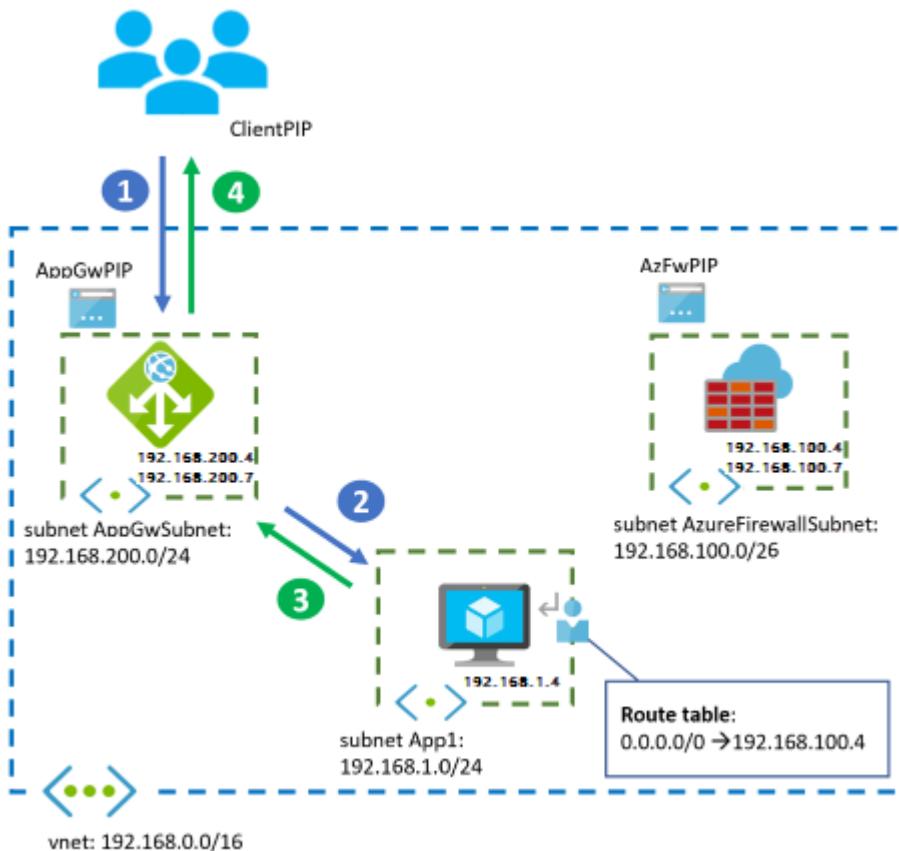
Flow	Goes through Application Gateway / WAF	Goes through Azure Firewall
HTTP(S) traffic from internet/onprem to Azure	Yes	No
HTTP(S) traffic from Azure to internet/onprem	No	Yes

Flow	Goes through Application Gateway / WAF	Goes through Azure Firewall
Non-HTTP(S) traffic from internet/onprem to Azure	No	Yes
Non-HTTP(S) traffic from Azure to internet/onprem	No	Yes

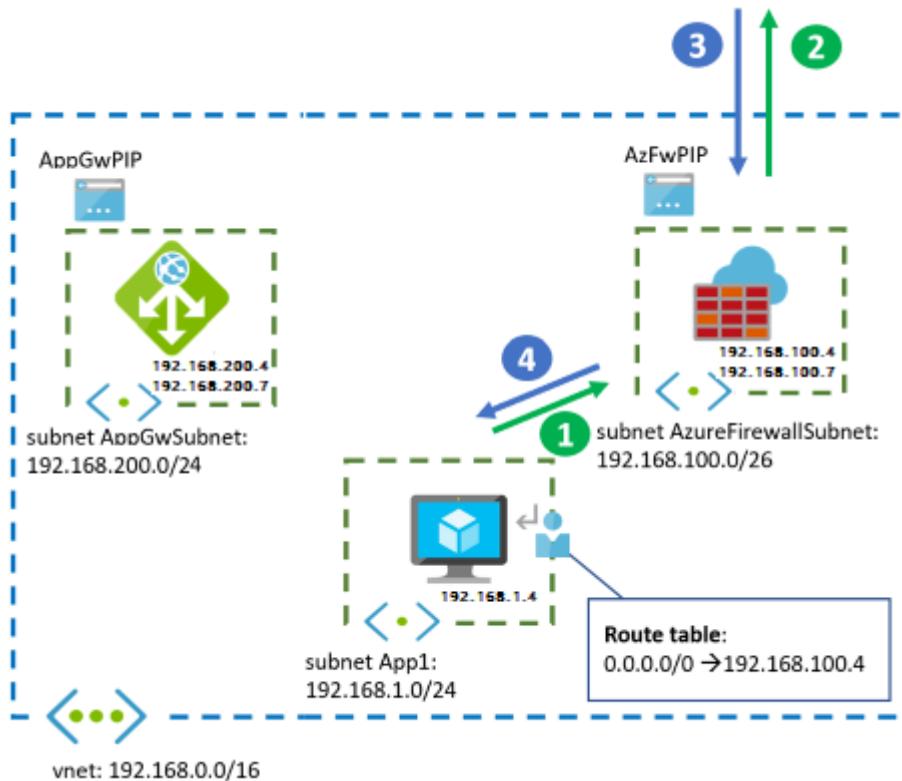
This design gives much more granular egress filtering than NSGs. For example, if applications need connectivity to a specific Azure Storage Account, you can use *fully qualified domain name (FQDN)*-based filters. With FQDN-based filters, applications aren't sending data to rogue storage accounts. That scenario couldn't be prevented just by using NSGs. This design is often used where outbound traffic requires FQDN-based filtering. One example situation is when [limiting egress traffic from an Azure Kubernetes Services cluster](#).

## Architectures

The following diagram illustrates the traffic flow for inbound HTTP(S) connections from an outside client:



The following diagram illustrates the traffic flow for outbound connections from the network VMs to the internet. One example is to connect to backend systems or get operating system updates:



The packet flow steps for each service are the same as in the previous standalone design options.

## Application Gateway before Firewall

In this option, inbound web traffic goes through both Azure Firewall and WAF. The WAF provides protection at the web application layer. Azure Firewall acts as a central logging and control point, and it inspects traffic between the Application Gateway and the backend servers. The Application Gateway and Azure Firewall aren't sitting in parallel, but one after the other.

With [Azure Firewall Premium](#), this design can support end-to-end scenarios, where the Azure Firewall applies TLS inspection to do IDPS on the encrypted traffic between the Application Gateway and the web backend.

This design is appropriate for applications that need to know incoming client source IP addresses, for example to serve geolocation-specific content or for logging. Application Gateway in front of Azure Firewall captures the incoming packet's source IP address in the *X-forwarded-for* header, so the web server can see the original IP address in this header. For more information, see [How an application gateway works](#).

Inbound HTTP(S) connections from the Internet need to be sent to the public IP address of the Application Gateway, HTTP(S) connections from Azure or on-premises to the private IP address. From the Application Gateway UDRs will make sure that the packets are routed through the Azure Firewall (see the packet walk further down for more details). For inbound non-HTTP(S) connections, traffic should be targeting the public IP address of the Azure Firewall (if coming from the public Internet), or it will be sent through the Azure Firewall by UDRs (if coming from other Azure VNets or on-premises networks). All outbound flows from Azure VMs will be forwarded to the Azure Firewall by UDRs.

The following table summarizes the traffic flows for this scenario:

[Expand table](#)

Flow	Goes through Application Gateway / WAF	Goes through Azure Firewall
HTTP(S) traffic from internet/onprem to Azure	Yes	Yes
HTTP(S) traffic from Azure to internet/onprem	No	Yes
Non-HTTP(S) traffic from internet/onprem to Azure	No	Yes
Non-HTTP(S) traffic from Azure to internet/onprem	No	Yes

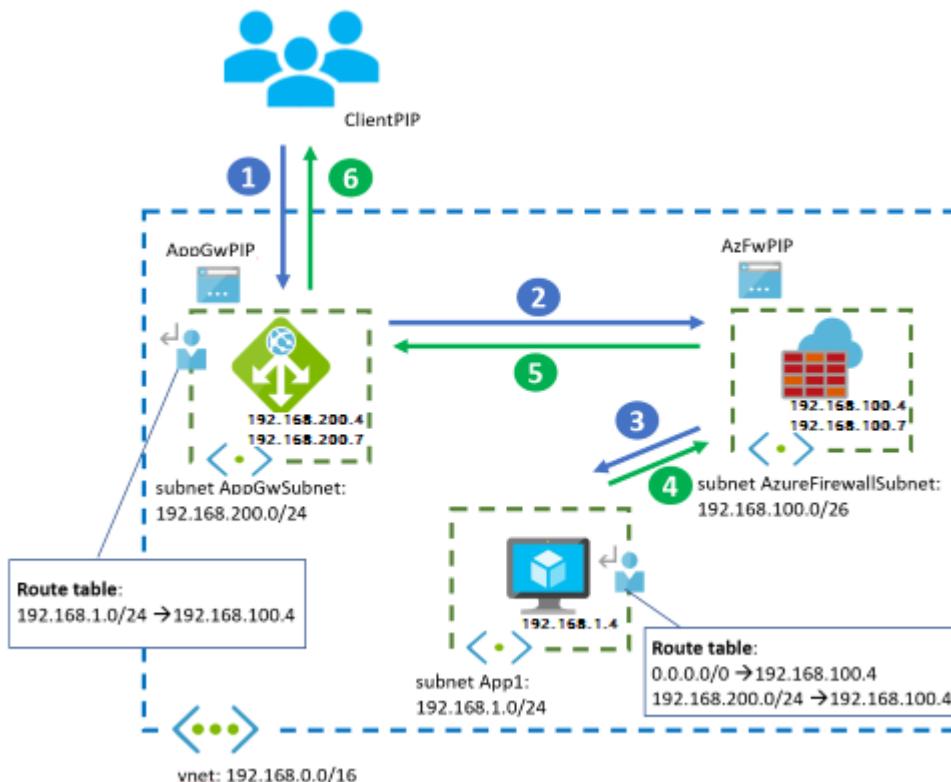
For web traffic from on-premises or internet to Azure, the Azure Firewall will inspect flows that the WAF has already allowed. Depending on whether the Application Gateway encrypts backend traffic (traffic from the Application Gateway to the application servers), you'll have different potential scenarios:

1. The Application Gateway encrypts traffic following zero-trust principles ([End-to-End TLS encryption](#)), and the Azure Firewall will receive encrypted traffic. Still, Azure Firewall Standard will be able to apply inspection rules, such as layer 3 & layer 4 filtering in network rules, or FQDN filtering in application rules using the TLS Server Name Indication (SNI) header. [Azure Firewall Premium](#) provides deeper visibility with IDPS, such as URL-based filtering.

2. If the Application Gateway is sending unencrypted traffic to the application servers, the Azure Firewall will see inbound traffic in clear text. TLS inspection isn't needed in the Azure Firewall.
3. If IDPS is enabled in the Azure Firewall, it will verify that the HTTP Host header matches the destination IP. With that purpose, it will need name resolution for the FQDN that's specified in the Host header. This name resolution can be achieved with Azure DNS Private Zones and the default Azure Firewall DNS settings using Azure DNS. It can also be achieved with custom DNS servers that need to be configured in the Azure Firewall settings. (For more information, see [Azure Firewall DNS Settings](#).) If there isn't administrative access to the Virtual Network where the Azure Firewall is deployed, the latter method is the only possibility. One example is with Azure Firewalls deployed in Virtual WAN Secured Hubs.

## Architecture

For the rest of the flows (inbound non-HTTP(S) traffic and any outbound traffic), the Azure Firewall will provide IDPS inspection and TLS inspection where appropriate. It also provides [FQDN-based filtering in network rules](#) based on DNS.



## Workflow

Network traffic from the public internet follows this flow:

1. The client starts the connection to the public IP address of the Azure Application Gateway:

- Source IP address: ClientPIP
- Destination IP address: AppGwPIP

2. The request to the Application Gateway public IP is distributed to a back-end instance of the gateway, in this case 192.168.200.7. The Application Gateway instance stops the connection from the client, and establishes a new connection with one of the back ends. The UDR to 192.168.1.0/24 in the Application Gateway subnet forwards the packet to the Azure Firewall, while preserving the destination IP to the web application:

- Source IP address: 192.168.200.7 (private IP address of the Application Gateway instance)
- Destination IP address: 192.168.1.4
- X-Forwarded-For header: ClientPIP

3. Azure Firewall doesn't SNAT the traffic, because the traffic is going to a private IP address. It forwards the traffic to the application VM if rules allow it. For more information, see [Azure Firewall SNAT](#). However, if the traffic hits an application rule in the firewall, the workload will see the source IP address of the specific firewall instance that processed the packet, since the Azure Firewall will proxy the connection:

- Source IP address if the traffic is allowed by an Azure Firewall network rule: 192.168.200.7 (the private IP address of one of the Application Gateway instances).
- Source IP address if the traffic is allowed by an Azure Firewall application rule: 192.168.100.7 (the private IP address of one of the Azure Firewall instances).
- Destination IP address: 192.168.1.4
- X-Forwarded-For header: ClientPIP

4. The VM answers the request, reversing source and destination IP addresses. The UDR to 192.168.200.0/24 captures the packet sent back to the Application Gateway and redirects it to Azure Firewall, while preserving the destination IP toward the Application Gateway.

- Source IP address: 192.168.1.4
- Destination IP address: 192.168.200.7

5. Here again the Azure Firewall doesn't SNAT the traffic, since it's going to a private IP address, and forwards the traffic to the Application Gateway.

- Source IP address: 192.168.1.4
- Destination IP address: 192.168.200.7

6. Finally, the Application Gateway instance answers the client:

- Source IP address: AppGwPIP
- Destination IP address: ClientPIP

Outbound flows from the VMs to the public internet go through Azure Firewall, as defined by the UDR to `0.0.0.0/0`.

## Application Gateway after firewall

This design lets Azure Firewall filter and discard malicious traffic before it reaches the Application Gateway. For example, it can apply features like threat intelligence-based filtering. Another benefit is that the application gets the same public IP address for both inbound and outbound traffic, regardless of protocol. However, Azure Firewall SNATs the incoming traffic, so the application will not have visibility to the original IP address of the HTTP requests. From an administration perspective, for example for troubleshooting purposes, you can obtain the actual client IP for a specific connection by correlating it with the SNAT logs of the Azure Firewall.

There's limited benefit in this scenario, because Azure Firewall will only see encrypted traffic going to the Application Gateway. There might be scenarios where this design is preferred. One case is if another WAF is earlier in the network (for example, with [Azure Front Door](#)), which could capture the original source IP in the `X-Forwarded-For` HTTP header. Or the design is preferred if many public IP addresses are required.

HTTP(S) inbound flows from the public Internet should target the public IP address of the Azure Firewall, and the Azure Firewall will DNAT (and SNAT) the packets to the private IP address of the Application Gateway. From other Azure VNets or on-premises networks, HTTP(S) traffic should be sent to the Application Gateway's private IP, and forwarded through the Azure Firewall with UDRs. Standard VNet routing will make sure that return traffic from the Azure VMs goes back to the Application Gateway, and from the Application Gateway to the Azure Firewall if DNAT rules were used. For traffic from on-premises or Azure UDRs in the Application Gateway subnet should be used (see the packet walk further down for more details). All outbound traffic from the Azure VMs to the internet will be sent through the Azure Firewall by UDRs.

The following table summarizes the traffic flows for this scenario:

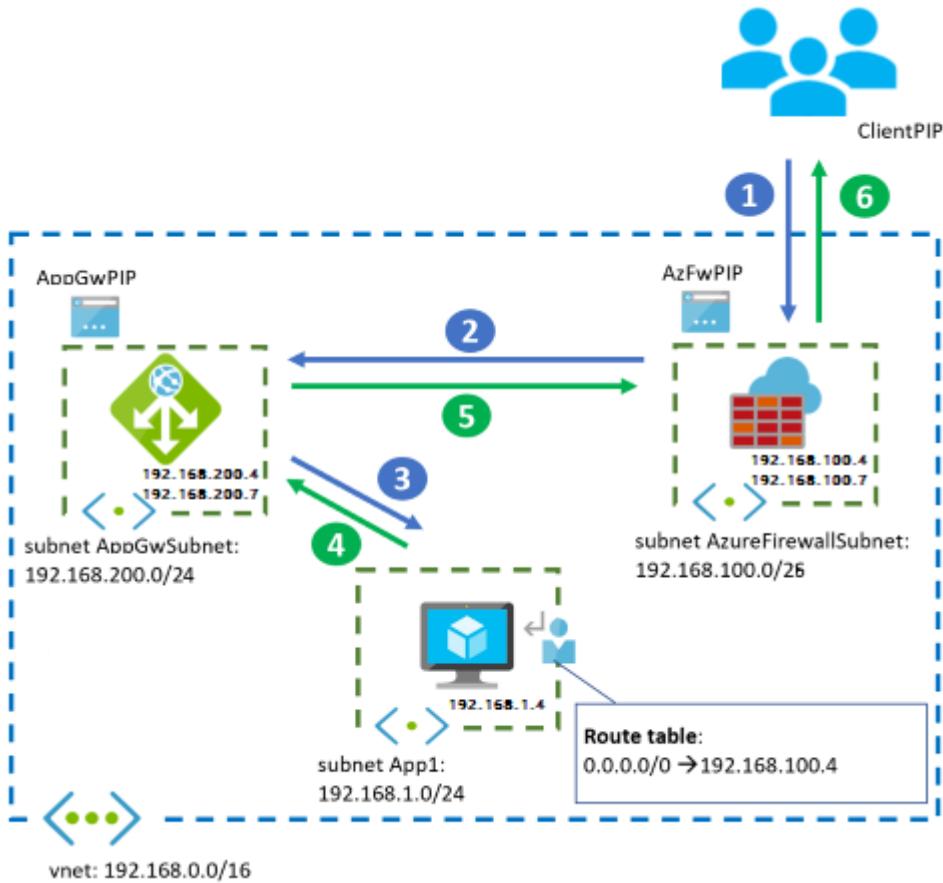
[ ] Expand table

<b>Flow</b>	<b>Goes through Application Gateway / WAF</b>	<b>Goes through Azure Firewall</b>
HTTP(S) traffic from internet/onprem to Azure	Yes	Yes (see below)
HTTP(S) traffic from Azure to internet/onprem	No	Yes
Non-HTTP(S) traffic from internet/onprem to Azure	No	Yes
Non-HTTP(S) traffic from Azure to internet/onprem	No	Yes

For inbound HTTP(S) traffic, the Azure Firewall would typically not decrypt traffic. It would instead apply IDPS policies that don't require TLS inspection, like IP-based filtering or using HTTP headers.

## Architecture

The application can't see the original source IP address of the web traffic; the Azure Firewall SNATs the packets as they come in to the virtual network. To avoid this problem, use [Azure Front Door](#) in front of the firewall. Azure Front Door injects the client's IP address as an HTTP header before it enters the Azure virtual network.



## Workflow

Network traffic from the public internet follows this flow:

1. The client starts the connection to the public IP address of the Azure Firewall:
  - Source IP address: ClientPIP
  - Destination IP address: AzFWPIP
2. The request to the Azure Firewall public IP is distributed to a back-end instance of the firewall, in this case 192.168.100.7. The Azure Firewall DNATs the web port, usually TCP 443, to the private IP address of the Application Gateway instance. Azure Firewall also SNATs when doing DNAT. For more information, see [Azure Firewall known issues](#):
  - Source IP address: 192.168.100.7 (the private IP address of the Azure Firewall instance)
  - Destination IP address: 192.168.200.4
3. The Application Gateway establishes a new session between the instance handling the connection and one of the backend servers. The original IP address of the client isn't in the packet:

- Source IP address: 192.168.200.7 (the private IP address of the Application Gateway instance)
- Destination IP address: 192.168.1.4
- X-Forwarded-For header: 192.168.100.7

4. The VM answers the Application Gateway, reversing source and destination IP addresses:

- Source IP address: 192.168.1.4
- Destination IP address: 192.168.200.7

5. The Application Gateway replies to the SNAT source IP address of the Azure Firewall instance. Even if the connection is coming from a specific Application Gateway instance like .7, Azure Firewall sees the internal IP address of the Application Gateway .4 as the source IP:

- Source IP address: 192.168.200.4
- Destination IP address: 192.168.100.7

6. Finally, Azure Firewall undoes SNAT and DNAT and answers the client:

- Source IP address: AzFwPIP
- Destination IP address: ClientPIP

Even if the Application Gateway has no listeners configured for applications, it still needs a public IP address so Microsoft can manage it.

### Note

A default route to 0.0.0.0/0 in the Application Gateway subnet pointing to the Azure Firewall is not supported, since it would break the control plane traffic required for the correct operation of the Azure Application Gateway.

## On-premises clients

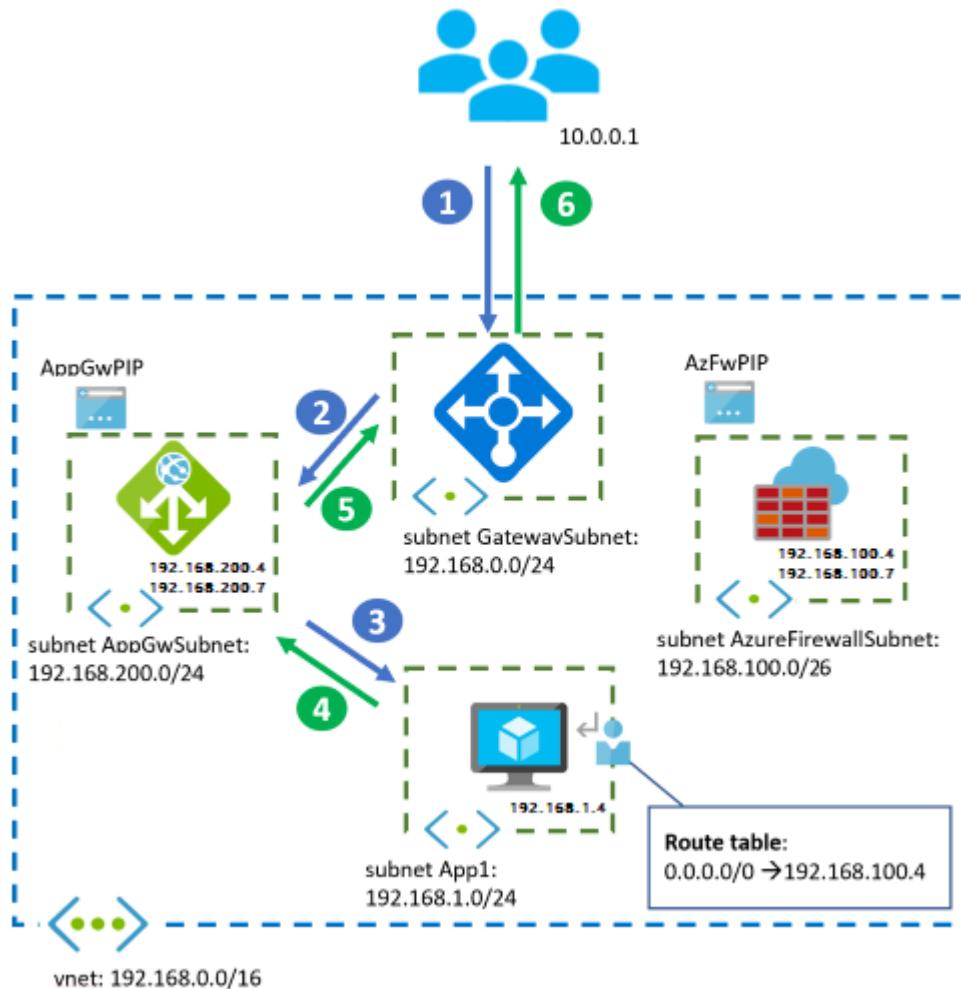
The preceding designs all show application clients coming from the public internet. On-premises networks also access applications. Most of the preceding information and traffic flows are the same as for internet clients, but there are some notable differences:

- A VPN gateway or ExpressRoute gateway sits in front of Azure Firewall or Application Gateway.
- WAF uses the private IP address of the Application Gateway.

- Azure Firewall doesn't support DNAT for private IP addresses. That's why you must use UDRs to send inbound traffic to Azure Firewall from the VPN or ExpressRoute gateways.
- Make sure to verify caveats around *forced tunneling* for the [Azure Application Gateway](#) and for the [Azure Firewall](#). Even if your workload doesn't need outbound connectivity to the public internet, you can't inject a default route like `0.0.0.0/0` for the Application Gateway that points to the on-premises network, or you'll break control traffic. For Azure Application Gateway, the default route needs to point to the public internet.

## Architecture

The following diagram shows the Azure Application Gateway and Azure Firewall parallel design. Application clients come from an on-premises network connected to Azure over VPN or ExpressRoute:

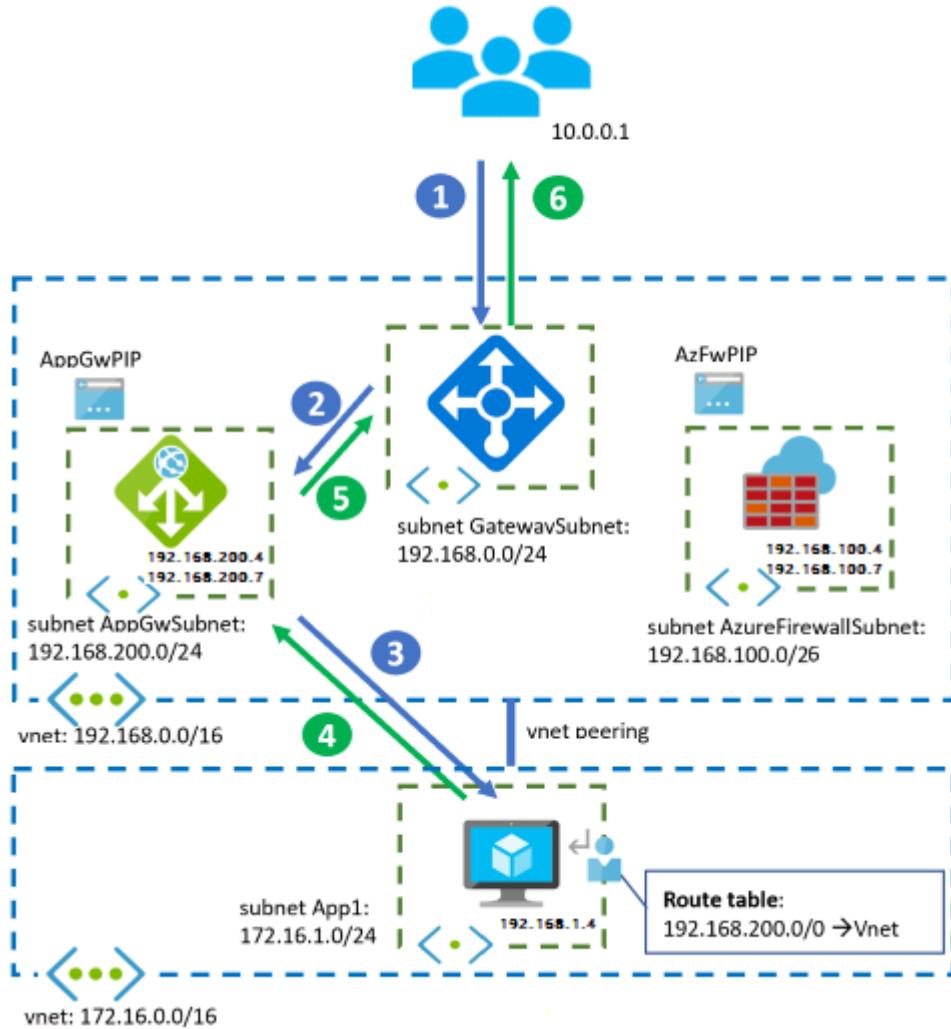


Even if all clients are located on-premises or in Azure, Azure Application Gateway and Azure Firewall both need to have public IP addresses. The public IP addresses allow Microsoft to manage the services.

# Hub and spoke topology

The designs in this article still apply in a *hub and spoke* topology. Shared resources in a central hub virtual network connect to applications in separate spoke virtual networks through virtual network peerings.

## Architecture



## Considerations

Some considerations for this topology include:

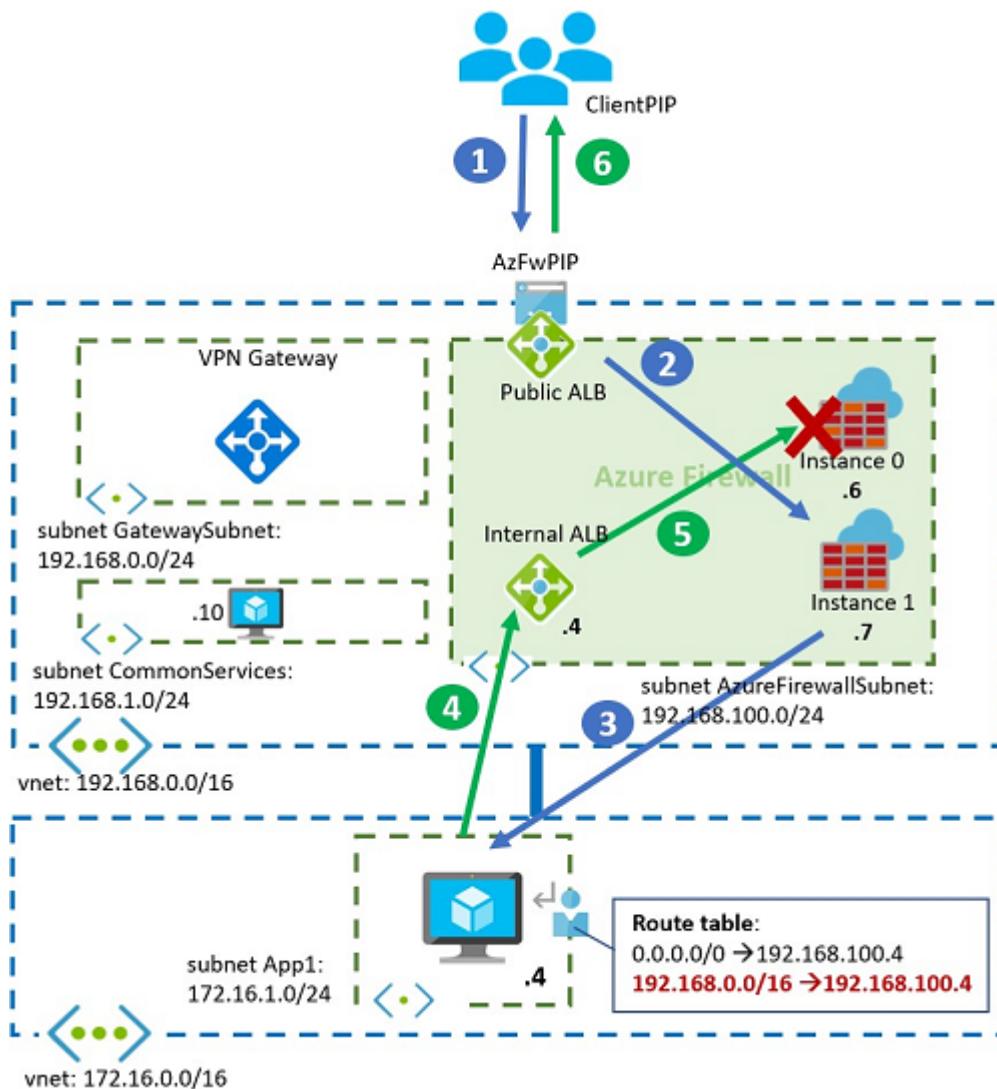
- The Azure Firewall is deployed in the central hub virtual network. The diagram above shows the practice of deploying the Application Gateway in the hub. Application teams often manage components such as Azure Application Gateways or Azure API Management gateways, though. In this case, these components are deployed in the spoke virtual networks.
- Pay special attention to UDRs in the spoke networks: When an application server in a spoke receives traffic from a specific Azure Firewall instance, like the

`192.168.100.7` address in the previous examples, it should send return traffic back to the same instance. If a UDR in the spoke sets the next hop of traffic addressed to the hub to the Azure Firewall IP address (`192.168.100.4` in the diagrams above), return packets might end up on a different Azure Firewall instance, causing asymmetric routing. Make sure that if you have UDRs in the spoke VNets to send traffic to shared services in the hub through the Azure Firewall, these UDRs don't include the prefix of the Azure Firewall subnet.

- The previous recommendation applies equally to the Application Gateway subnet and any other Network Virtual Appliances or reverse proxies that might be deployed in the hub VNet.
- You can't set the next hop for the Application Gateway or Azure Firewall subnets through static routes with a next hop type of `Virtual Network`. This next hop type is only valid in the local VNet and not across VNet peerings. For more information about user-defined routes and next hop types, see [Virtual network traffic routing](#).

## Asymmetric routing

The diagram below shows how a spoke sends back SNATted traffic back to the ALB of an Azure Firewall. This setup causes asymmetric routing:



To solve this problem, define UDRs in the spoke without the Azure Firewall subnet but with only the subnets where the shared services are located. In the example, the correct UDR in the spoke should only contain 192.168.1.0/24. It shouldn't contain the whole 192.168.0.0/16, as marked in red.

## Integration with other Azure products

You can integrate Azure Firewall and Azure Application Gateway with other Azure products and services.

## API Management Gateway

Integrate reverse proxy services like [API Management](#) gateway into the previous designs to provide functionality like API throttling or authentication proxy. Integrating an API Management gateway doesn't greatly alter the designs. The main difference is that instead of the single Application Gateway reverse proxy, there are two reverse proxies chained behind each other.

For more information, see the [Design Guide to integrate API Management and Application Gateway in a virtual network](#) and the application pattern [API Gateways for Microservices](#).

## Azure Kubernetes Service

For workloads running on an AKS cluster, you can deploy Azure Application Gateway independently of the cluster. Or you can integrate it with the AKS cluster using the [Azure Application Gateway Ingress Controller](#). When configuring certain objects at the Kubernetes levels (such as services and ingresses), the Application Gateway automatically adapts without needing extra manual steps.

Azure Firewall plays an important role in AKS cluster security. It offers the required functionality to filter egress traffic from the AKS cluster based on FQDN, not just IP address. For more information, see [Control egress traffic for AKS cluster nodes](#).

When you combine Application Gateway and Azure Firewall to protect an AKS cluster, it's best to use the parallel design option. The Application Gateway with WAF processes inbound connection requests to web applications in the cluster. Azure Firewall permits only explicitly allowed outbound connections. See [Baseline architecture for an Azure Kubernetes Service \(AKS\) cluster](#) for an example of the parallel design option.

## Azure Front Door

[Azure Front Door](#) functionality partly overlaps with Azure Application Gateway. For example, both services offer web application firewalling, SSL offloading, and URL-based routing. One main difference is that while Azure Application Gateway is inside a virtual network, Azure Front Door is a global, decentralized service.

You sometimes can simplify virtual network design by replacing Application Gateway with a decentralized Azure Front Door. Most designs described here remain valid, except for the option of placing Azure Firewall in front of Azure Front Door.

An interesting use case is using Azure Firewall in front of Application Gateway in your virtual network. As described earlier, the `X-Forwarded-For` header injected by the Application Gateway will contain the firewall instance's IP address, not the client's IP address. A workaround is to use Azure Front Door in front of the firewall to inject the client's IP address as a `X-Forwarded-For` header before the traffic enters the virtual network and hits the Azure Firewall. A second option is to [Secure your Origin with Private Link in Azure Front Door Premium](#).

For more information about the differences between the two services, or when to use each one, see [Frequently Asked Questions for Azure Front Door](#).

## Other network virtual appliances

Microsoft products aren't the only choice to implement web application firewall or next-generation firewall functionality in Azure. A wide range of Microsoft partners provide network virtual appliances (NVAs). The concepts and designs are essentially the same as in this article, but there are some important considerations:

- Partner NVAs for next-generation firewalling may offer more control and flexibility for NAT configurations unsupported by the Azure Firewall. Examples include DNAT from on-premises or DNAT from the internet without SNAT.
- Azure-managed NVAs (like Application Gateway and Azure Firewall) reduce complexity, compared to NVAs where users need to handle scalability and resiliency across many appliances.
- When using NVAs in Azure, use *active-active* and *autoscaling* setups, so these appliances aren't a bottleneck for applications running in the virtual network.

## Contributors

*This article is maintained by Microsoft. It was originally written by the following contributors.*

Principal author:

- [Jose Moreno](#) | Principal Customer Engineer

## Next steps

Learn more about the component technologies:

- [What is Azure Application Gateway?](#)
- [What is Azure Firewall?](#)
- [What is Azure Front Door?](#)
- [Azure Kubernetes Service](#)
- [What is Azure Virtual Network?](#)
- [What is Azure Web Application Firewall?](#)

## Related resources

Explore related architectures:

- [Implement a secure hybrid network](#)
- [Securely managed web applications](#)
- [Securing your Microsoft Teams channel bot and web app behind a firewall](#)
- [Consumer health portal on Azure](#)
- [Security considerations for highly sensitive IaaS apps in Azure](#)
- [Multitenant SaaS on Azure](#)
- [Enterprise deployment using App Services Environment](#)
- [High availability enterprise deployment using App Services Environment](#)
- [Baseline architecture for an Azure Kubernetes Service \(AKS\) cluster](#)

# Zero-trust network for web applications with Azure Firewall and Application Gateway

Azure Application Gateway

Azure Firewall

Azure Virtual Network

Azure Virtual WAN

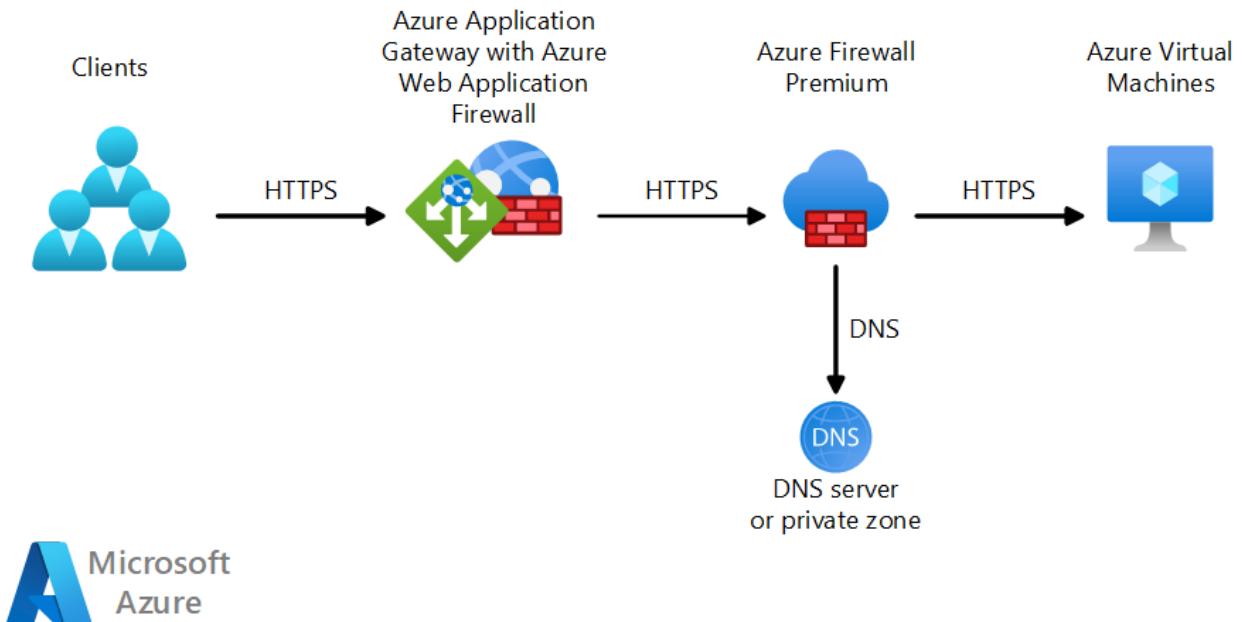
Azure Web Application Firewall

This guide outlines a strategy for implementing [zero-trust](#) security for web apps for inspection and encryption. The zero-trust paradigm includes many other concepts, such as constant verification of the identity of the actors or reducing the size of the implicit trust areas to a minimum. This article refers to the encryption and inspection component of a zero-trust architecture for traffic inbound from the public Internet. Please read other [zero-trust documents](#) for more aspects of deploying your application securely, such as authentication. For the purpose of this article, a multilayered approach works best, where network security makes up one of the layers of the zero-trust model. In this layer, network appliances inspect packets to ensure that only legitimate traffic reaches applications.

Typically, different types of network appliances inspect different aspects of network packets:

- Web application firewalls look for patterns that indicate an attack at the web application layer.
- Next-generation firewalls can also look for generic threats.

In some situations, you can combine different types of network security appliances to increase protection. A separate guide, [Firewall and Application Gateway for virtual networks](#), describes design patterns that you can use to arrange the various appliances. This document focuses on a common pattern for maximizing security, in which Azure Application Gateway acts before Azure Firewall Premium. The following diagram illustrates this pattern:



[Download a Visio file](#) of this architecture.

This architecture uses the Transport Layer Security (TLS) protocol to encrypt traffic at every step.

- A client sends packets to Application Gateway, a load balancer. It runs with the optional addition [Azure Web Application Firewall](#).
- Application Gateway decrypts the packets and searches for threats to web applications. If it doesn't find any threats, it uses zero-trust principles to encrypt the packets. Then it releases them.
- Azure Firewall Premium runs security checks:
  - [Transport layer security \(TLS\) inspection](#) decrypts and examines the packets.
  - [Intrusion detection and protection](#) features check the packets for malicious intent.
- If the packets pass the tests, Azure Firewall Premium takes these steps:
  - Encrypts the packets
  - Uses a Domain Name System (DNS) service to determine the application virtual machine (VM)
  - Forwards the packets to the application VM

Various inspection engines in this architecture ensure traffic integrity:

- Web Application Firewall uses rules to prevent attacks at the web layer. Examples of attacks include SQL code injection and cross-site scripting. For more information on rules and the Open Web Application Security Project (OWASP) Core Rule Set, see [Web Application Firewall CRS rule groups and rules](#).

- Azure Firewall Premium uses generic intrusion detection and prevention rules. These rules help identify malicious files and other threats that target web applications.

This architecture supports different types of network design, which this article discusses:

- Traditional hub and spoke networks
- Networks that use Azure Virtual WAN as a platform
- Networks that use Azure Route Server to simplify dynamic routing

## Azure Firewall Premium and name resolution

When checking for malicious traffic, Azure Firewall Premium verifies that the HTTP Host header matches the packet IP address and TCP port. For example, suppose Application Gateway sends web packets to the IP address 172.16.1.4 and TCP port 443. The value of the HTTP Host header should resolve to that IP address.

HTTP Host headers usually don't contain IP addresses. Instead, the headers contain names that match the server's digital certificate. In this case, Azure Firewall Premium uses DNS to resolve the Host header name to an IP address. The network design determines which DNS solution works best, as later sections describe.

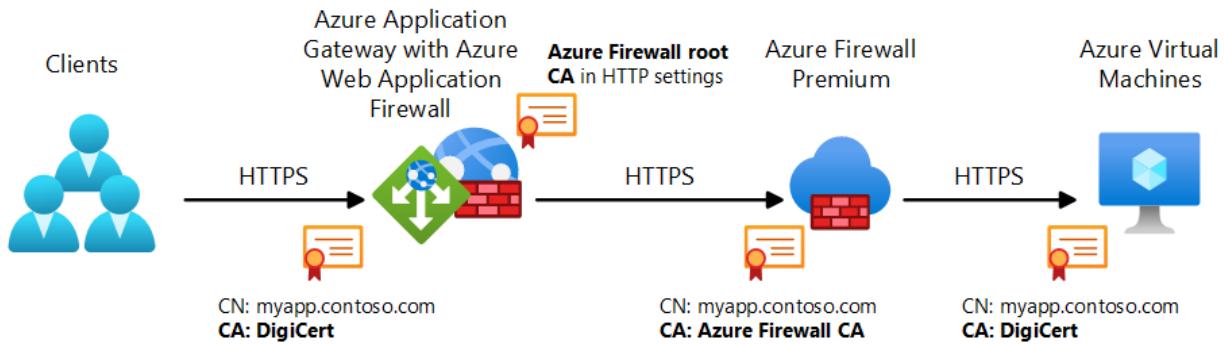
### Note

Application Gateway doesn't support port numbers in HTTP Host headers. As a result:

- Azure Firewall Premium assumes a default HTTPS TCP port of 443.
- The connection between Application Gateway and the web server only supports TCP port 443, not non-standard ports.

## Digital certificates

The following diagram shows the common names (CNs) and certificate authorities (CAs) that the architecture's TLS sessions and certificates use:



## TLS connections

This architecture contains three distinct TLS connections. Digital certificates validate each one:

### From clients to Application Gateway

In Application Gateway, you deploy the digital certificate that clients see. A well-known CA such as DigiCert or Let's Encrypt typically issues such a certificate.

### From Application Gateway to Azure Firewall Premium

To decrypt and inspect TLS traffic, Azure Firewall Premium dynamically generates certificates. Azure Firewall Premium also presents itself to Application Gateway as the web server. A private CA signs the certificates that Azure Firewall Premium generates. For more information, see [Azure Firewall Premium certificates](#). Application Gateway needs to validate those certificates. In the application's HTTP settings, you configure the root CA that Azure Firewall Premium uses.

### From Azure Firewall Premium to the web server

Azure Firewall Premium establishes a TLS session with the destination web server. Azure Firewall Premium verifies that a well-known CA signs the web server TLS packets.

## Component roles

Application Gateway and Azure Firewall Premium handle certificates differently from one another because their roles differ:

- Application Gateway is a *reverse web proxy*. It protects web servers from malicious clients by intercepting HTTP and HTTPS requests. You declare each protected server that's in the back-end pool of Application Gateway with its IP address or fully qualified domain name. Legitimate clients should be able to access each application. So you configure Application Gateway with a digital certificate that a public CA has signed. Use a CA that any TLS client will accept.
- Azure Firewall Premium is a *forward web proxy* or, simply, a web proxy. It protects clients from malicious web servers by intercepting TLS calls from the protected clients. When a protected client makes an HTTP request, the forward proxy impersonates the target web server by generating digital certificates and presenting them to the client. Azure Firewall Premium uses a private CA, which signs the dynamically generated certificates. You configure the protected clients to trust that private CA. In this architecture, Azure Firewall Premium protects requests from Application Gateway to the web server. Application Gateway trusts the private CA that Azure Firewall Premium uses.

## Hub and spoke example

Typically, a hub and spoke design deploys shared network components in the hub virtual network and application-specific components in the spokes. In most systems, Azure Firewall Premium is a shared resource. But Web Application Firewall can be a shared network device or an application-specific component. For the following reasons, it's usually best to treat Application Gateway as an application component and deploy it in a spoke virtual network:

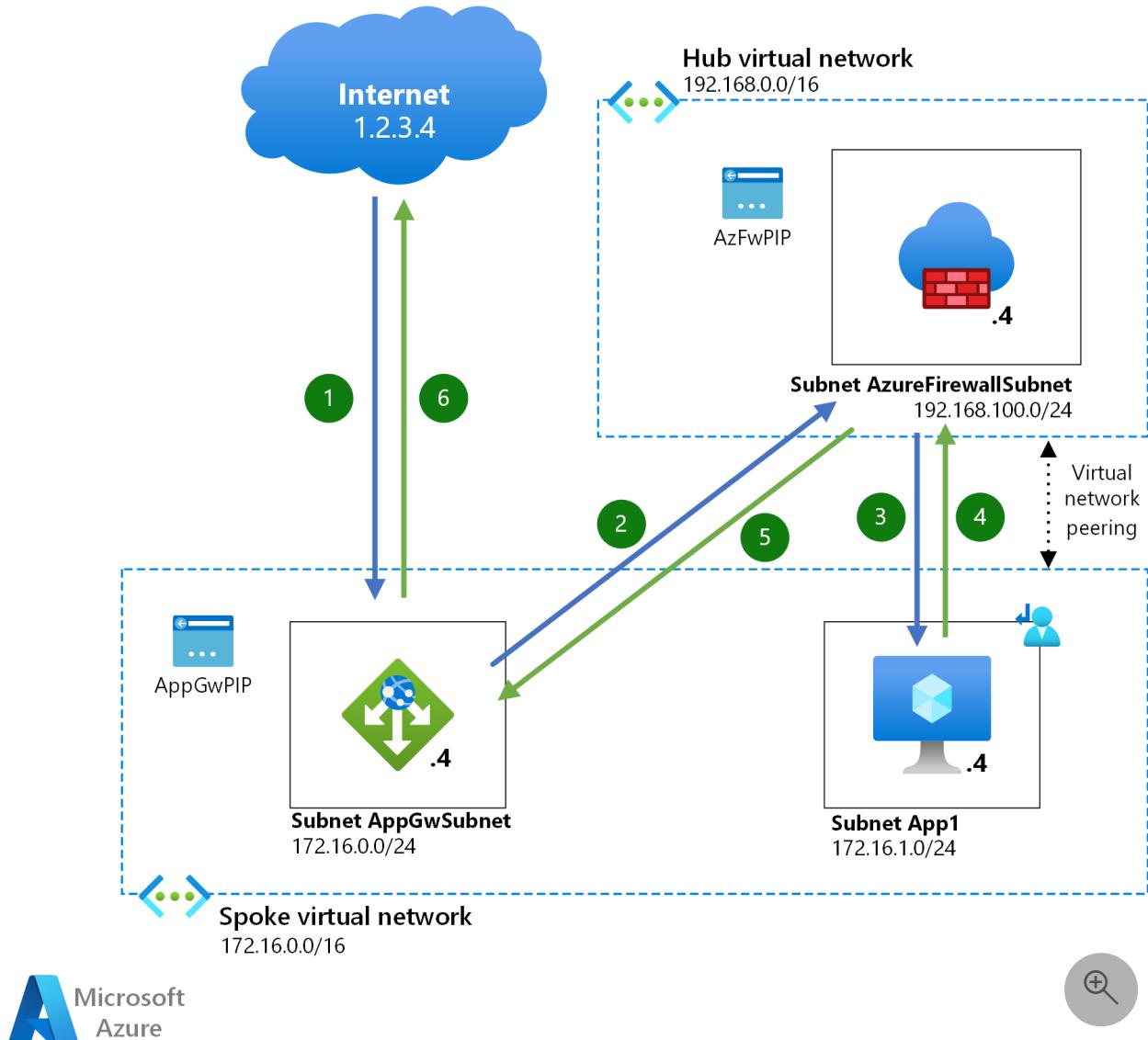
- It can be difficult to troubleshoot Web Application Firewall alerts. You generally need in-depth knowledge of the application to decide whether the messages that trigger those alarms are legitimate.
- If you treat Application Gateway as a shared resource, you might exceed [Azure Application Gateway limits](#).
- You might face role-based access control problems if you deploy Application Gateway in the hub. This situation can come up when teams manage different applications but use the same instance of Application Gateway. Each team then has access to the entire Application Gateway configuration.

With traditional hub and spoke architectures, DNS private zones provide an easy way to use DNS:

- Configure a DNS private zone.
- Link the zone to the virtual network that contains Azure Firewall Premium.

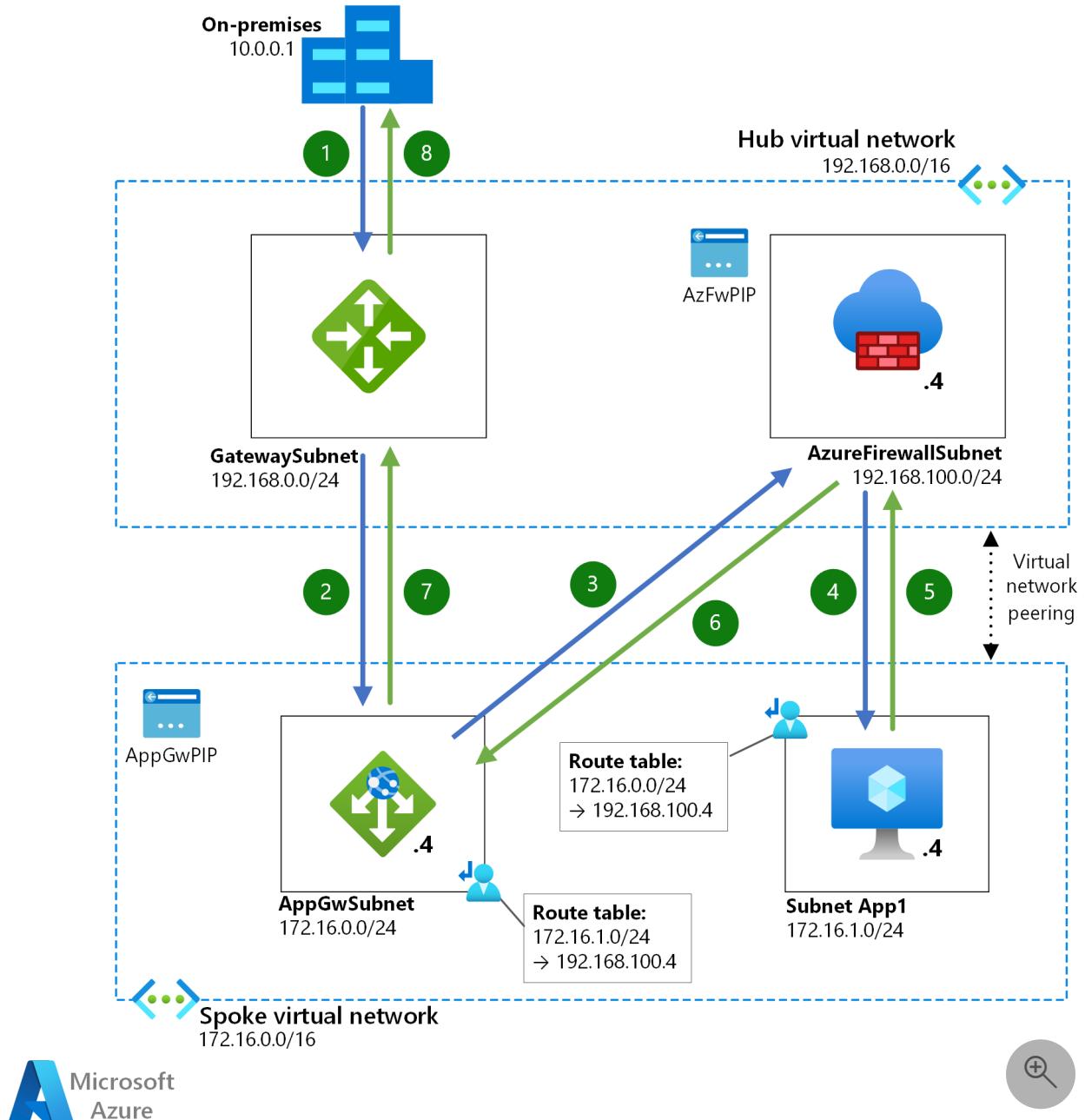
- Make sure that an A record exists for the value that Application Gateway uses for traffic and for health checks.

The following diagram shows the packet flow when Application Gateway is in a spoke virtual network. In this case, a client connects from the public internet.



1. A client submits a request to a web server.
2. Application Gateway intercepts the client packets and examines them. If the packets pass inspection, the Application Gateway would send the packet to the backend VM. When the packet hits Azure, a user-defined route (UDR) in the Application Gateway subnet forwards the packets to Azure Firewall Premium.
3. Azure Firewall Premium runs security checks on the packets. If they pass the tests, Azure Firewall Premium forwards the packets to the application VM.
4. The VM responds and sets the destination IP address to the Application Gateway. A UDR in the VM subnet redirects the packets to Azure Firewall Premium.
5. Azure Firewall Premium forwards the packets to Application Gateway.
6. Application Gateway answers the client.

Traffic can also arrive from an on-premises network instead of the public internet. The traffic flows either through a site-to-site virtual private network (VPN) or through ExpressRoute. In this scenario, the traffic first reaches a virtual network gateway in the hub. The rest of the network flow is the same as the previous case.



1. An on-premises client connects to the virtual network gateway.
2. The gateway forwards the client packets to Application Gateway.
3. Application Gateway examines the packets. If they pass inspection, a UDR in the Application Gateway subnet forwards the packets to Azure Firewall Premium.
4. Azure Firewall Premium runs security checks on the packets. If they pass the tests, Azure Firewall Premium forwards the packets to the application VM.
5. The VM responds and sets the destination IP address to Application Gateway. A UDR in the VM subnet redirects the packets to Azure Firewall Premium.
6. Azure Firewall Premium forwards the packets to Application Gateway.
7. Application Gateway sends the packets to the virtual network gateway.

8. The gateway answers the client.

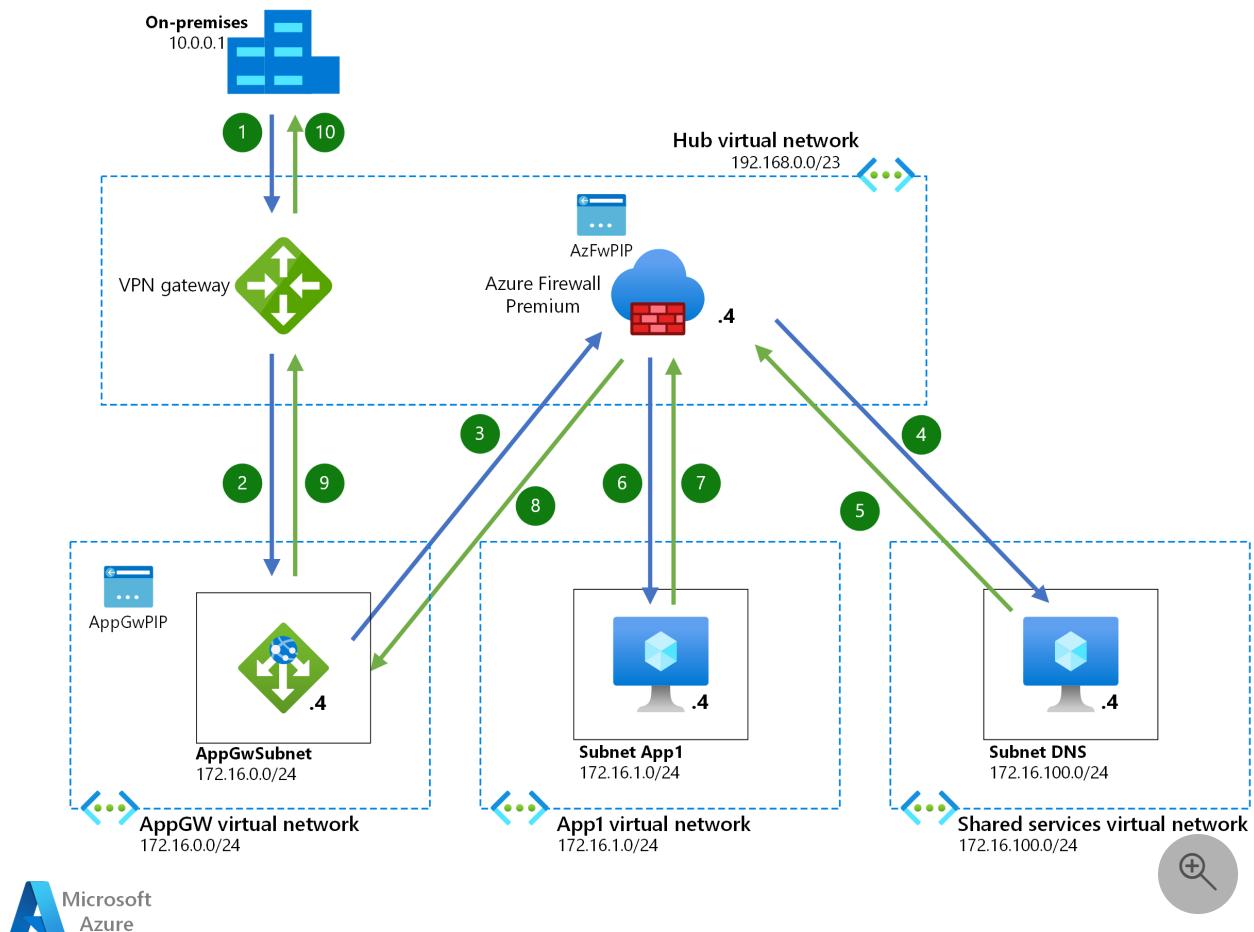
## Virtual WAN example

You can also use the networking service [Virtual WAN](#) in this architecture. This component offers many benefits. For instance, it eliminates the need for user-maintained UDRs in spoke virtual networks. You can define static routes in virtual hub route tables instead. The programming of every virtual network that you connect to the hub then contains these routes.

When you use Virtual WAN as a networking platform, two main differences result:

- You can't link DNS private zones to a virtual hub because Microsoft manages virtual hubs. As the subscription owner, you don't have permissions for linking private DNS zones. As a result, you can't associate a DNS private zone with the secure hub that contains Azure Firewall Premium. To implement DNS resolution for Azure Firewall Premium, use DNS servers instead:
  - Configure the [Azure Firewall DNS Settings](#) to use custom DNS servers.
  - Deploy the servers in a shared services virtual network that you connect to the virtual WAN.
  - Link a DNS private zone to the shared services virtual network. The DNS servers can then resolve the names that Application Gateway uses in HTTP Host headers. For more information, see [Azure Firewall DNS Settings](#).
- You can only use Virtual WAN to program routes in a spoke if the prefix is shorter (less specific) than the virtual network prefix. For example, in the diagrams above the spoke VNet has the prefix 172.16.0.0/16: in this case, Virtual WAN would not be able to inject a route that matches the VNet prefix (172.16.0.0/16) or any of the subnets (172.16.0.0/24, 172.16.1.0/24). In other words, Virtual WAN cannot attract traffic between two subnets that are in the same VNet. This limitation becomes apparent when Application Gateway and the destination web server are in the same virtual network: Virtual WAN can't force the traffic between Application Gateway and the web server to go through Azure Firewall Premium (a workaround would be manually configuring User Defined Routes in the subnets of the Application Gateway and web server).

The following diagram shows the packet flow in a case that uses Virtual WAN. In this situation, access to Application Gateway is from an on-premises network. A site-to-site VPN or ExpressRoute connects that network to Virtual WAN. Access from the internet is similar.



1. An on-premises client connects to the VPN.
2. The VPN forwards the client packets to Application Gateway.
3. Application Gateway examines the packets. If they pass inspection, the Application Gateway subnet forwards the packets to Azure Firewall Premium.
4. Azure Firewall Premium requests DNS resolution from a DNS server in the shared services virtual network.
5. The DNS server answers the resolution request.
6. Azure Firewall Premium runs security checks on the packets. If they pass the tests, Azure Firewall Premium forwards the packets to the application VM.
7. The VM responds and sets the destination IP address to Application Gateway. The Application subnet redirects the packets to Azure Firewall Premium.
8. Azure Firewall Premium forwards the packets to Application Gateway.
9. Application Gateway sends the packets to the VPN.
10. The VPN answers the client.

With this design, you might need to modify the routing that the hub advertises to the spoke virtual networks. Specifically, Application Gateway v2 only supports a 0.0.0.0/0 route that points to the internet. Routes with this address that don't point to the internet break the connectivity that Microsoft requires for managing Application Gateway. If your virtual hub advertises a 0.0.0.0/0 route, prevent that route from propagating to the Application Gateway subnet by taking one of these steps:

- Create a route table with a route for 0.0.0.0/0 and a next hop type of `Internet`. Associate that route with the subnet that you deploy Application Gateway in.
- If you deploy Application Gateway in a dedicated spoke, disable the propagation of the default route in the settings for the virtual network connection.

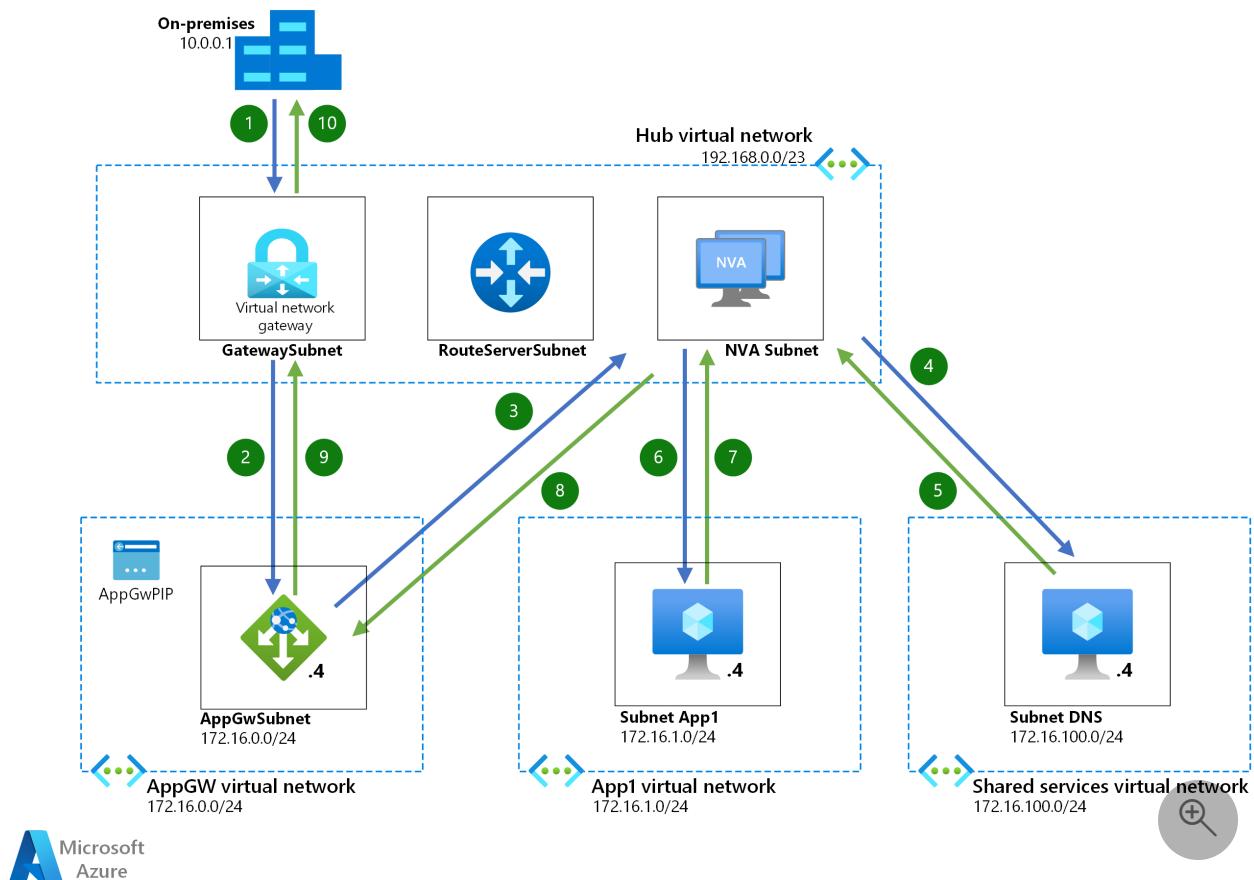
## Route Server example

[Route Server](#) offers another way to inject routes automatically in spokes. With this functionality, you avoid the administrative overhead of maintaining route tables. Route Server combines the Virtual WAN and hub and spoke variants:

- With Route Server, customers manage hub virtual networks. As a result, you can link the hub virtual network to a DNS private zone.
- Route Server has the same limitation that Virtual WAN has concerning IP address prefixes. You can only inject routes into a spoke if the prefix is shorter (less specific) than the virtual network prefix. Because of this limitation, Application Gateway and the destination web server need to be in different virtual networks.

The following diagram shows the packet flow when Route Server simplifies dynamic routing. Note these points:

- Route Server currently requires the device that injects the routes to send them over Border Gateway Protocol (BGP). Since Azure Firewall Premium doesn't support BGP, use a third-party Network Virtual Appliance (NVA) instead.
- The functionality of the NVA in the hub determines whether your implementation needs DNS.



1. An on-premises client connects to the virtual network gateway.
2. The gateway forwards the client packets to Application Gateway.
3. Application Gateway examines the packets. If they pass inspection, the Application Gateway subnet forwards the packets to a backend machine. A route in the ApplicationGateway subnet injected by the Route Server would forward the traffic to an NVA.
4. The NVA runs security checks on the packets. If they pass the tests, the NVA forwards the packets to the application VM.
5. The VM responds and sets the destination IP address to Application Gateway. A route injected in the VM subnet by the Route Server redirects the packets to the NVA.
6. The NVA forwards the packets to Application Gateway.
7. Application Gateway sends the packets to the virtual network gateway.
8. The gateway answers the client.

As with Virtual WAN, you might need to modify the routing when you use Route Server. If you advertise the 0.0.0.0/0 route, it might propagate to the Application Gateway subnet. But Application Gateway doesn't support that route. In this case, configure a route table for the Application Gateway subnet. Include a route for 0.0.0.0/0 and a next hop type of **Internet** in that table.

## Contributors

*This article is maintained by Microsoft. It was originally written by the following contributors.*

Principal author:

- [Jose Moreno](#) | Principal Customer Engineer

## Next steps

- [Secure networks with zero trust](#)
- [Virtual network traffic routing](#)
- [How an application gateway works](#)

## Related resources

- [Secure and govern workloads with network level segmentation](#)
- [Implement a secure hybrid network](#)
- [Hub-spoke network topology in Azure](#)
- [Hub-spoke network topology with Azure Virtual WAN](#)

# Design area: Azure governance

Article • 02/28/2023

Azure governance establishes the tooling needed to support cloud governance, compliance auditing, and automated guardrails.

## Design area review

**Involved roles or functions:** Azure governance is led by [cloud governance](#). The [cloud platform](#) and [cloud center of excellence](#) might be required to define and implement some technical requirements. Governance focuses on the enforcement of operations and security requirements, which might require [cloud security](#), [central IT](#) or [cloud operations](#).

**Scope:** Review decisions made during reviews of [identity](#), [network](#), [security](#), and [management](#) design areas. The team might compare review decisions from automated governance, which is part of the Azure landing zone accelerator. Review decisions might help determine what can be audited or enforced. Review decisions might evaluate what policies can be automatically deployed.

**Out of scope:** Azure governance establishes the foundation for networking. However, it doesn't address compliance-related articles such as advanced network security or automated guardrails to enforce networking decisions. These networking decisions might be addressed when reviewing compliance design areas related to [security](#) and [governance](#). Delaying the discussions might allow the cloud platform team to address initial networking requirements before addressing more complex articles.

**New (greenfield) cloud environment:** To start your cloud journey with a small set of subscriptions, see [Create your initial Azure subscriptions](#). Also, consider using Bicep deployment templates in building out your new Azure landing zones. For more information, see [Azure Landing Zones Bicep - Deployment Flow](#).

**Existing (brownfield) cloud environment:** Consider the following if you're interested in applying proven-practice Azure governance principles to existing Azure environments:

- Review our guidance for establishing a [management baseline](#) for your hybrid or multicloud environment
- Implement [Azure Cost Management + Billing](#) features like billing scopes, budgets, and alerts to ensure your Azure spend stays within prescribed bounds
- Use [Azure Policy](#) to enforce governance guardrails on Azure deployments, and trigger remediation tasks to bring existing Azure resources into a compliant state

- Consider [Azure AD entitlement management](#) to automate Azure requests, access assignments, reviews, and expiration
- Use [Azure Advisor](#) recommendations to ensure cost optimization and operational excellence in Azure, both of which are core principles of the [Microsoft Azure Well-Architected Framework](#).

The [Azure Landing Zones Bicep - Deployment Flow](#) repository contains many Bicep deployment templates that can accelerate your greenfield and brownfield Azure landing zone deployments. These templates already have Microsoft proven-practice governance guidance integrated within them.

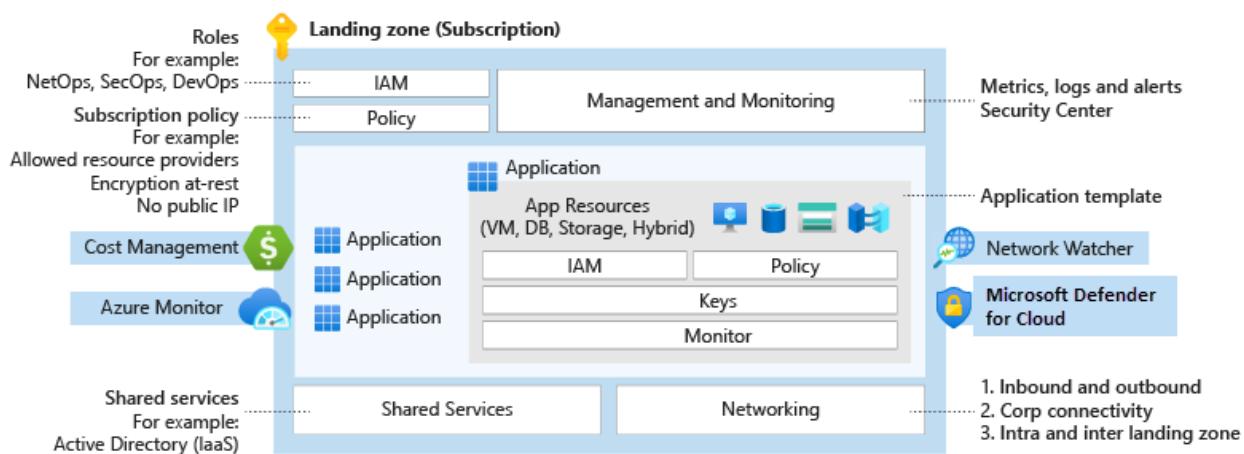
For instance, consider using the [ALZ Default Policy Assignments](#) Bicep module to get a head start on ensuring compliance for your Azure environments.

For more information on working in brownfield cloud environments, see [Brownfield environment considerations](#).

## Design area overview

An organization's cloud adoption journey starts with strong controls to govern environments.

Governance provides mechanisms and processes for maintaining control over platforms, applications, and resources in Azure.



The design area review explores the considerations and recommendations that help you make informed decisions as you plan your landing zone.

The governance design area focuses on the design decisions in the landing zone. Also, the [Govern methodology](#) of the Cloud Adoption Framework gives guidance for governance processes and tools.

The Govern methodology consists of five disciplines:

Discipline	Context
Cost management	Explore guidance to cost reporting control techniques
Security baseline	Explore further in the <a href="#">security design area</a>
Resource consistency	Explore guidance for naming and tagging resources in the environment governance
Identity baseline	Covered in depth in the <a href="#">identity and access management</a> design area
Deployment acceleration	Explore further in the <a href="#">platform automation and DevOps</a> design area

## Azure governance considerations

Azure policy ensures security and compliance for enterprise technical estates. Azure policy might enforce vital management and security conventions across Azure platform services. Azure policy supplements Azure role-based access control, which controls authorized users actions. Also, Azure Cost Management + Billing might help support your ongoing governance cost and spending in Azure, or other multicloud environments.

## Deployment acceleration considerations

Change advisory review boards might hinder an organizations innovation and business agility. Azure Policy increases workload efficiency by replacing such reviews with automated guardrails and adherence audits.

- Determine what Azure policies are needed based on your business controls or compliance regulations. Use the policies included in the Azure landing zone accelerator as a starting point.
- Use the [standards-based blueprint samples](#) to consider other policies that might align to your business requirements.
- Enforce networking, identity, management, and security conventions are often automated.
- Manage and create policy assignments by using policy definitions, which might be reused at multiple inherited assignment scopes. You can have centralized baseline policy assignments at management, subscription, and resource group scopes.
- Ensure continuous compliance with compliance reporting and auditing.
- Understand that Azure Policy has limits, such as the restriction of definitions at any particular scope: [policy limits](#).

- Understand regulatory compliance policies. The policies might include HIPAA, PCI-DSS, or SOC 2 Trust Services Criteria.

## Cost management considerations

- How is the organization's cost and recharging model structured? What are the key data points required to accurately see cloud services spend?
- Finding the structure of tags that fits your cost and recharging model might help track your cloud spend.
- Azure pricing calculator can be used to estimate the expected monthly costs for using any combination of Azure products.
- Azure Hybrid Benefit might help reduce the costs of running your workloads in the cloud. You can use your on-premises Software Assurance-enabled Windows Server and SQL Server licenses on Azure. It also applies to Red Hat and SUSE Linux subscriptions.
- Azure Reservations helps you save money by committing to one-year or three-year plans for multiple products. Committing lets you get resource discounts, which might significantly reduce your resource costs by up to 72% from pay-as-you-go prices.
- Azure savings plan for compute is our most flexible savings plan and generates savings up to 65 percent on pay-as-you-go prices. Pick a one-year or three-year commitment that will apply to compute services regardless of region, instance size, or operating system. Eligible compute services include virtual machines, dedicated hosts, container instances, Azure premium functions, and Azure app services. You can combine an Azure savings plan with Azure Reservations to optimize compute cost and flexibility. For more information, see [Azure savings plan](#).
- Azure policies can be used to allow specific regions, resource types, and resource SKUs.
- Azure Storage lifecycle management offers a rule-based policy. The policy might be used to move blob data to the appropriate access tiers, or to expire data at the end of the data lifecycle.
- Azure dev/test subscriptions give you access to select Azure services for nonproduction workloads at discounted pricing.
- Use autoscaling to save costs by dynamically allocating and de-allocating resources to match your performance needs.
- Using Azure spot virtual machines allows you to take advantage of our unused capacity at a significant cost savings. Azure spot virtual machines are great for workloads that can handle interruptions. For example, batch processing jobs, dev/test environments, large compute workloads, and more.

- Some Azure services are free for 12 months while some other services are always free. Selecting the right Azure services helps you reduce costs.
- Selecting the right compute service for your application can help with cost efficiency. Azure offers many ways to host your code.

## Resource consistency considerations

- What are the groups of resources in your environment? These groups can share configuration characteristics that might be required to help stay consistent.
- Is the application or workload subscription design the most appropriate for your operation needs?
- Are there groups of resources that should share a common lifecycle?
- Are there groups of resources that should share common access constraints (such as Role-based access control)?
- Are there standard resource configurations within your organization that might be used to ensure a consistent baseline configuration?

## Security baseline considerations

- What tools and guardrails need to be enforced across the environment as part of a security baseline?
- Who might be notified when deviations are found?
- Consider using Azure Policy to enforce tools (such as Microsoft Defender for Cloud, Microsoft Defender for Cloud).
- Consider using Azure Policy to enforce guardrails (such as the Microsoft cloud security benchmark).

## Identity management considerations

- Who might have access to audit logs for identity and access management?
- Who might be notified when suspicious sign-in events occur?
- Consider using [Azure Active Directory reports](#) to govern activity.
- Consider the logs from Azure AD, which might be sent to the central Log Analytics workspace for the platform.
- Explore the capabilities of [Azure AD access reviews](#) in your landing zone governance approach.
- Explore the capabilities of [Azure AD entitlement management](#) in your landing zone governance approach.

## Azure governance recommendations

## Deployment acceleration recommendations

- Identify required Azure tags and use the append policy mode to enforce usage.  
Use the [tagging strategy](#) article as a starting point
- Map regulatory and compliance requirements to Azure Policy definitions and Azure role assignments.
- Establish Azure Policy definitions at the top-level root management group as they might be assigned at inherited scopes.
- Manage policy assignments at the highest appropriate level with exclusions at bottom levels, if necessary.
- Use Azure Policy to control resource provider registrations at the subscription or management group levels.
- Use built-in policies to minimize operational overhead.
- Assign the built-in Resource Policy Contributor role at a particular scope to enable application-level governance.
- Limit the number of Azure Policy assignments made at the root management group scope to avoid managing through exclusions at inherited scopes.

## Cost management recommendations

- Use Azure Cost Management + Billing to implement financial oversight on resources in your environment.
- Use tags in Azure to append metadata to resources, which might enable granular analysis of spend (such as cost center or project name).

## Azure governance in the Azure landing zone accelerator

The Azure landing zone accelerator implementation includes capabilities to help organizations efficiently get mature governance controls.

For example:

- A management group hierarchy that groups resources by function or workload type might encourage best practices for resource consistency.
- A rich set of Azure policies might enable governance controls at management group level to ensure all resources are in scope.

# Connect Microsoft Sentinel to Amazon Web Services to ingest AWS service log data

Article • 12/28/2023

Use the Amazon Web Services (AWS) connectors to pull AWS service logs into Microsoft Sentinel. These connectors work by granting Microsoft Sentinel access to your AWS resource logs. Setting up the connector establishes a trust relationship between Amazon Web Services and Microsoft Sentinel. This is accomplished on AWS by creating a role that gives permission to Microsoft Sentinel to access your AWS logs.

This connector is available in two versions: the legacy connector for CloudTrail management and data logs, and the new version that can ingest logs from the following AWS services by pulling them from an S3 bucket (links are to AWS documentation):

- [Amazon Virtual Private Cloud \(VPC\)](#) - [VPC Flow Logs](#)
- [Amazon GuardDuty](#) - [Findings](#)
- [AWS CloudTrail](#) - [Management](#) and [data](#) events
- [AWS CloudWatch](#) - [CloudWatch logs](#)

## ⓘ Important

- The Amazon Web Services S3 connector is currently in **PREVIEW**. See the [Supplemental Terms of Use for Microsoft Azure Previews](#) for additional legal terms that apply to Azure features that are in beta, preview, or otherwise not yet released into general availability.

S3 connector (new)

This tab explains how to configure the AWS S3 connector. The process of setting it up has two parts: the AWS side and the Microsoft Sentinel side. Each side's process produces information used by the other side. This two-way authentication creates secure communication.

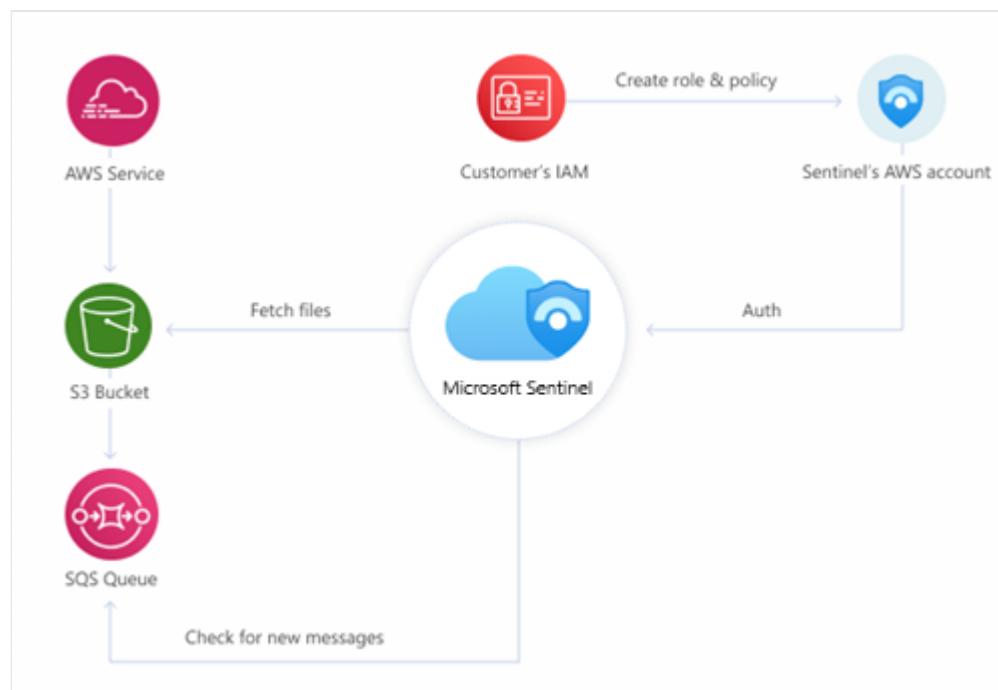
## Prerequisites

- Make sure that the logs from your selected AWS service use the format accepted by Microsoft Sentinel:

- **Amazon VPC**: .csv file in GZIP format with headers; delimiter: space.
  - **Amazon GuardDuty**: json-line and GZIP formats.
  - **AWS CloudTrail**: .json file in a GZIP format.
  - **CloudWatch**: .csv file in a GZIP format without a header. If you need to convert your logs to this format, you can use this [CloudWatch Lambda function](#).
- You must have write permission on the Microsoft Sentinel workspace.
  - Install the Amazon Web Services solution from the **Content Hub** in Microsoft Sentinel. For more information, see [Discover and manage Microsoft Sentinel out-of-the-box content](#).

## Architecture overview

This graphic and the following text show how the parts of this connector solution interact.



- AWS services are configured to send their logs to S3 (Simple Storage Service) storage buckets.
- The S3 bucket sends notification messages to the SQS (Simple Queue Service) message queue whenever it receives new logs.
- The Microsoft Sentinel AWS S3 connector polls the SQS queue at regular, frequent intervals. If there is a message in the queue, it will contain the path to the log files.

- The connector reads the message with the path, then fetches the files from the S3 bucket.
- To connect to the SQS queue and the S3 bucket, Microsoft Sentinel uses AWS credentials and connection information embedded in the AWS S3 connector's configuration. The AWS credentials are configured with a role and a permissions policy giving them access to those resources. Similarly, the Microsoft Sentinel workspace ID is embedded in the AWS configuration, so there is in effect two-way authentication.

## Connect the S3 connector

- In your AWS environment:
  - Configure your AWS service(s) to send logs to an **S3 bucket**.
  - Create a **Simple Queue Service (SQS) queue** to provide notification.
  - Create an **assumed role** to grant permissions to your Microsoft Sentinel account (external ID) to access your AWS resources.
  - Attach the appropriate **IAM permissions policies** to grant Microsoft Sentinel access to the appropriate resources (S3 bucket, SQS).

We have made available, in our GitHub repository, a script that **automates the AWS side of this process**. See the instructions for [automatic setup](#) later in this document.

- In Microsoft Sentinel:
  - Enable and configure the **AWS S3 Connector** in the Microsoft Sentinel portal. See the instructions below.

## Automatic setup

To simplify the onboarding process, Microsoft Sentinel has provided a [PowerShell script to automate the setup](#) of the AWS side of the connector - the required AWS resources, credentials, and permissions.

The script takes the following actions:

- Creates an *IAM assumed role* with the minimal necessary permissions, to grant Microsoft Sentinel access to your logs in a given S3 bucket and SQS queue.

- Enables specified AWS services to send logs to that S3 bucket, and notification messages to that SQS queue.
- If necessary, creates that S3 bucket and that SQS queue for this purpose.
- Configures any necessary IAM permissions policies and applies them to the IAM role created above.

## Prerequisites for automatic setup

- You must have PowerShell and the AWS CLI on your machine.
  - [Installation instructions for PowerShell](#)
  - [Installation instructions for the AWS CLI](#) (from AWS documentation)

## Instructions

To run the script to set up the connector, use the following steps:

1. From the Microsoft Sentinel navigation menu, select **Data connectors**.
2. Select **Amazon Web Services S3** from the data connectors gallery.

If you don't see the connector, install the Amazon Web Services solution from the **Content Hub** in Microsoft Sentinel.
3. In the details pane for the connector, select **Open connector page**.
4. In the **Configuration** section, under **1. Set up your AWS environment**, expand **Setup with PowerShell script (recommended)**.
5. Follow the on-screen instructions to download and extract the [AWS S3 Setup Script](#) (link downloads a zip file containing the main setup script and helper scripts) from the connector page.
6. Before running the script, run the `aws configure` command from your PowerShell command line, and enter the relevant information as prompted. See [AWS Command Line Interface | Configuration basics](#) (from AWS documentation) for details.
7. Now run the script. Copy the command from the connector page (under "Run script to set up the environment") and paste it in your command line.
8. The script will prompt you to enter your Workspace ID. This ID appears on the connector page. Copy it and paste it at the prompt of the script.

## Amazon Web Services S3 (Preview) ...

**Description**  
This connector allows you to ingest AWS service logs, collected in AWS S3 buckets, to Microsoft Sentinel. The currently supported data types are:

- AWS CloudTrail
- VPC Flow Logs
- AWS GuardDuty
- AWS CloudWatch

**Last data received**  
--

**Content source** (1) **Version**  
Amazon Web Services 1.0.0

**Author** Microsoft **Supported by** Microsoft Corporation | Email

**Related content**

- 0 Workbooks
- 3 Queries
- 0 Analytics rules templates

**Configuration**

- Set up your AWS environment

There are two options for setting up your AWS environment to send logs from an S3 bucket to your Log Analytics Workspace:

Setup with PowerShell script (recommended)

Download and extract the files from the following link: [AWS S3 Setup Script](#).

- Make sure that you have PowerShell on your machine:  
[Installation instructions for PowerShell](#).
- Make sure that you have the AWS CLI on your machine:  
[Installation instructions for the AWS CLI](#).

Before running the script, run the aws configure command from your PowerShell command line, and enter the relevant information as prompted. See [AWS Command Line Interface | Configuration basics](#) for details.

(7) Run script to set up the environment  
./ConfigAwsConnector.ps1

(8) External ID (Workspace ID)  
12345678-abcd-abcd-123456abcdef

9. When the script finishes running, copy the **Role ARN** and the **SQS URL** from the script's output (see example in first screenshot below) and paste them in their respective fields in the connector page under **2. Add connection** (see second screenshot below).

```
Updating the SQS policy to allow S3 notifications, and ARN to read/delete/change visibility of SQS messages and get queue url
Changes S3: SQS SendMessage permission to 'vpc-sentinel-demo' s3 bucket
Changes Role ARN: SQS ChangeMessageVisibility, DeleteMessage, ReceiveMessage and GetQueueUrl permissions to 'SentinelDemo' rule

Attaching S3 read policy to Sentinel role.

Changes Role arn: S3 Get and List permissions to 'SentinelDemo' rule

Updating the S3 policy to allow Sentinel to read the data.
Changes: S3 Get and List permissions to 'SentinelDemo' rule

Enabling S3 event notifications (for *.gz file)

Please enter the event notifications name: demo
Using event notification name: demo
Event notification prefix definition, to Limit the notifications to objects with key starting with specified characters.

The default prefix is 'AWSLogs/123412341234/vpcflowlogs/'.
Do you want to override the event notification prefix? [y/n]: n

Use the values below to configure the Amazon Web Service S3 data connector in the Azure Sentinel portal.

Role Arn: arn:aws:iam::123412341234:role/SentinelDemo
Sqs Url: https://sqs.us-east-1.amazonaws.com/123412341234/vpc-sentinel-demo
Destination Table: AWSVPCFlow
```

The screenshot shows the Microsoft Sentinel Log Analytics workspace. On the left, there's a sidebar with 'Amazon Web Services S3 (Preview)' selected. The main area has tabs for 'Configuration', 'Setup with PowerShell script (recommended)', and 'Manual Setup'. The 'Configuration' tab is active. It shows the 'Role to add' field with 'arn:awsiam::123412341234:role/MicrosoftSentinelRole' and the 'SQS URL' field with 'https://sqs.us-east-2.amazonaws.com/123412341234/Contoso\_Sentinel-queue'. A dropdown for 'Destination table' is set to 'AWSVPCFlow'. A red box highlights the 'Add connection' button at the bottom of this section.

10. Select a data type from the **Destination table** drop-down list. This tells the connector which AWS service's logs this connection is being established to collect, and into which Log Analytics table it will store the ingested data. Then select **Add connection**.

#### (!) Note

The script may take up to 30 minutes to finish running.

## Manual setup

Microsoft recommends using the automatic setup script to deploy this connector. If for whatever reason you do not want to take advantage of this convenience, follow the steps below to set up the connector manually.

- [Prepare your AWS resources](#)
- [Create an AWS assumed role and grant access to the AWS Sentinel account](#)
- [Add the AWS role and queue information to the S3 data connector](#)
- [Configure an AWS service to export logs to an S3 bucket](#)

## Prepare your AWS resources

- Create an **S3 bucket** to which you will ship the logs from your AWS services - VPC, GuardDuty, CloudTrail, or CloudWatch.

- See the [instructions to create an S3 storage bucket](#) in the AWS documentation.
- Create a standard **Simple Queue Service (SQS)** message queue to which the S3 bucket will publish notifications.
  - See the [instructions to create a standard Simple Queue Service \(SQS\) queue](#) in the AWS documentation.
- Configure your S3 bucket to send notification messages to your SQS queue.
  - See the [instructions to publish notifications to your SQS queue](#) in the AWS documentation.

## Create an AWS assumed role and grant access to the AWS Sentinel account

1. In Microsoft Sentinel, select **Data connectors** from the navigation menu.
2. Select **Amazon Web Services S3** from the data connectors gallery.

If you don't see the connector, install the Amazon Web Services solution from the **Content Hub** in Microsoft Sentinel. For more information, see [Discover and manage Microsoft Sentinel out-of-the-box content](#).
3. In the details pane for the connector, select **Open connector page**.
4. Under **Configuration**, expand **Setup with PowerShell script (recommended)**, then copy the **External ID (Workspace ID)** to your clipboard.
5. In a different browser window or tab, open the AWS console. Follow the [instructions in the AWS documentation for creating a role for an AWS account](#).
  - For the account type, instead of **This account**, choose **Another AWS account**.
  - In the **Account ID** field, enter the number **197857026523** (you can copy and paste it from here). This number is **Microsoft Sentinel's service account ID for AWS**. It tells AWS that the account using this role is a Microsoft Sentinel user.
  - In the options, select **Require external ID** (*do not* select **Require MFA**). In the **External ID** field, paste your Microsoft Sentinel **Workspace ID** that you copied in the previous step. This identifies *your specific Microsoft Sentinel account* to AWS.

- Assign the necessary permissions policies. These policies include:
  - `AmazonSQSReadOnlyAccess`
  - `AWSLambdaSQSQueueExecutionRole`
  - `AmazonS3ReadOnlyAccess`
  - `ROSAKMSProviderPolicy`
  - Additional policies for ingesting the different types of AWS service logs.

For information on these policies, see the [AWS S3 connector permissions policies page](#) in the Microsoft Sentinel GitHub repository.

- Name the role with a meaningful name that includes a reference to Microsoft Sentinel. Example: "*MicrosoftSentinelRole*".

## Add the AWS role and queue information to the S3 data connector

1. In the browser tab open to the AWS console, enter the **Identity and Access Management (IAM)** service and navigate to the list of **Roles**. Select the role you created above.
2. Copy the **ARN** to your clipboard.
3. Enter the **Simple Queue Service**, select the SQS queue you created, and copy the **URL** of the queue to your clipboard.
4. Return to your Microsoft Sentinel browser tab, which should be open to the **Amazon Web Services S3 (Preview)** data connector page. Under **2. Add connection**:
  - a. Paste the IAM role ARN you copied two steps ago into the **Role to add** field.
  - b. Paste the URL of the SQS queue you copied in the last step into the **SQS URL** field.
  - c. Select a data type from the **Destination table** drop-down list. This tells the connector which AWS service's logs this connection is being established to collect, and into which Log Analytics table it will store the ingested data.
  - d. Select **Add connection**.

The screenshot shows the Microsoft Sentinel Data connectors page. On the left, there's a sidebar with navigation links like Home, Microsoft Sentinel, Data connectors, and a search bar. The main content area is titled "Amazon Web Services S3 (Preview)". It has tabs for Disconnected Status, Amazon Provider, and Last Log Received. Below these are sections for Description, Content source (Amazon Web Services), Author (Microsoft), and Related content (Workbooks, Queries, Analytics rules templates). A chart shows Data received over time. On the right, there's a "Configuration" section with two steps: 1. Set up your AWS environment (with options for Setup with PowerShell script or Manual Setup) and 2. Add connection (with fields for Role to add, SQS URL, Destination table, and an Add Connection button). The "Add connection" section is highlighted with a red box.

## Configure an AWS service to export logs to an S3 bucket

See Amazon Web Services documentation (linked below) for the instructions for sending each type of log to your S3 bucket:

- [Publish a VPC flow log to an S3 bucket](#).

### ⚠ Note

If you choose to customize the log's format, you must include the *start* attribute, as it maps to the *TimeGenerated* field in the Log Analytics workspace. Otherwise, the *TimeGenerated* field will be populated with the event's *ingested time*, which doesn't accurately describe the log event.

- [Export your GuardDuty findings to an S3 bucket](#).

### ⚠ Note

- In AWS, findings are exported by default every 6 hours. Adjust the export frequency for updated Active findings based on your environment requirements. To expedite the process, you can modify the default setting to export findings every 15 minutes. See [Setting the frequency for exporting updated active findings](#).

- The *TimeGenerated* field is populated with the finding's *Update at* value.
- AWS CloudTrail trails are stored in S3 buckets by default.
  - [Create a trail for a single account](#).
  - [Create a trail spanning multiple accounts across an organization](#).
- [Export your CloudWatch log data to an S3 bucket](#).

## Known issues and troubleshooting

### Known issues

- Different types of logs can be stored in the same S3 bucket, but should not be stored in the same path.
- Each SQS queue should point to one type of message, so if you want to ingest GuardDuty findings *and* VPC flow logs, you should set up separate queues for each type.
- Similarly, a single SQS queue can serve only one path in an S3 bucket, so if for any reason you are storing logs in multiple paths, each path requires its own dedicated SQS queue.

### Troubleshooting

Learn how to [troubleshoot Amazon Web Services S3 connector issues](#).

## Next steps

In this document, you learned how to connect to AWS resources to ingest their logs into Microsoft Sentinel. To learn more about Microsoft Sentinel, see the following articles:

- Learn how to [get visibility into your data, and potential threats](#).
- Get started [detecting threats with Microsoft Sentinel](#).
- [Use workbooks](#) to monitor your data.

# Connect Microsoft Entra data to Microsoft Sentinel

Article • 10/12/2023

You can use Microsoft Sentinel's built-in connector to collect data from [Microsoft Entra ID](#) and stream it into Microsoft Sentinel. The connector allows you to stream the following log types:

- [Sign-in logs](#), which contain information about interactive user sign-ins where a user provides an authentication factor.

The Microsoft Entra connector now includes the following three additional categories of sign-in logs, all currently in [PREVIEW](#):

- [Non-interactive user sign-in logs](#), which contain information about sign-ins performed by a client on behalf of a user without any interaction or authentication factor from the user.
- [Service principal sign-in logs](#), which contain information about sign-ins by apps and service principals that do not involve any user. In these sign-ins, the app or service provides a credential on its own behalf to authenticate or access resources.
- [Managed Identity sign-in logs](#), which contain information about sign-ins by Azure resources that have secrets managed by Azure. For more information, see [What are managed identities for Azure resources?](#)
- [Audit logs](#), which contain information about system activity relating to user and group management, managed applications, and directory activities.
- [Provisioning logs](#) (also in [PREVIEW](#)), which contain system activity information about users, groups, and roles provisioned by the Microsoft Entra provisioning service.

## Important

Some of the available log types are currently in [PREVIEW](#). See the [Supplemental Terms of Use for Microsoft Azure Previews](#) for additional legal terms that apply to Azure features that are in beta, preview, or otherwise not yet released into general availability.

### Note

For information about feature availability in US Government clouds, see the Microsoft Sentinel tables in [Cloud feature availability for US Government customers](#).

## Prerequisites

- A Microsoft Entra ID P1 or P2 license is required to ingest sign-in logs into Microsoft Sentinel. Any Microsoft Entra ID license (Free/O365/P1 or P2) is sufficient to ingest the other log types. Additional per-gigabyte charges may apply for Azure Monitor (Log Analytics) and Microsoft Sentinel.
- Your user must be assigned the [Microsoft Sentinel Contributor](#) role on the workspace.
- Your user must be assigned the [Global Administrator](#) or [Security Administrator](#) roles on the tenant you want to stream the logs from.
- Your user must have read and write permissions to the Microsoft Entra diagnostic settings in order to be able to see the connection status.
- Install the solution for **Microsoft Entra ID** from the **Content Hub** in Microsoft Sentinel. For more information, see [Discover and manage Microsoft Sentinel out-of-the-box content](#).

## Connect to Microsoft Entra ID

1. In Microsoft Sentinel, select **Data connectors** from the navigation menu.
2. From the data connectors gallery, select **Microsoft Entra ID** and then select [Open connector page](#).
3. Mark the check boxes next to the log types you want to stream into Microsoft Sentinel (see above), and select **Connect**.

## Find your data

After a successful connection is established, the data appears in **Logs**, under the **LogManagement** section, in the following tables:

- `SigninLogs`
- `AuditLogs`
- `AADNonInteractiveUserSignInLogs`
- `AADServicePrincipalSignInLogs`
- `AADManagedIdentitySignInLogs`
- `AADProvisioningLogs`

To query the Microsoft Entra logs, enter the relevant table name at the top of the query window.

## Next steps

In this document, you learned how to connect Microsoft Entra ID to Microsoft Sentinel.

To learn more about Microsoft Sentinel, see the following articles:

- Learn how to [get visibility into your data and potential threats](#).
- Get started [detecting threats with Microsoft Sentinel](#).

# Connect Microsoft Defender for Cloud alerts to Microsoft Sentinel

Article • 05/09/2023

[Microsoft Defender for Cloud](#)'s integrated cloud workload protections allow you to detect and quickly respond to threats across hybrid and multi-cloud workloads.

This connector allows you to stream [security alerts from Defender for Cloud](#) into Microsoft Sentinel, so you can view, analyze, and respond to Defender alerts, and the incidents they generate, in a broader organizational threat context.

As [Microsoft Defender for Cloud Defender plans](#) are enabled per subscription, this data connector is also enabled or disabled separately for each subscription.

Microsoft Defender for Cloud was formerly known as Azure Security Center. Defender for Cloud's enhanced security features were formerly known collectively as Azure Defender.

## ⓘ Note

For information about feature availability in US Government clouds, see the Microsoft Sentinel tables in [Cloud feature availability for US Government customers](#).

## Alert synchronization

- When you connect Microsoft Defender for Cloud to Microsoft Sentinel, the status of security alerts that get ingested into Microsoft Sentinel is synchronized between the two services. So, for example, when an alert is closed in Defender for Cloud, that alert will display as closed in Microsoft Sentinel as well.
- Changing the status of an alert in Defender for Cloud will *not* affect the status of any Microsoft Sentinel **incidents** that contain the Microsoft Sentinel alert, only that of the alert itself.

## Bi-directional alert synchronization

Enabling **bi-directional sync** will automatically sync the status of original security alerts with that of the Microsoft Sentinel incidents that contain those alerts. So, for example,

when a Microsoft Sentinel incident containing a security alerts is closed, the corresponding original alert will be closed in Microsoft Defender for Cloud automatically.

## Prerequisites

- You must have read and write permissions on your Microsoft Sentinel workspace.
- You must have the **Contributor** or **Owner** role on the subscription you want to connect to Microsoft Sentinel.
- You will need to enable at least one plan within Microsoft Defender for Cloud for each subscription where you want to enable the connector. To enable Microsoft Defender plans on a subscription, you must have the **Security Admin** role for that subscription.
- You will need the `SecurityInsights` resource provider to be registered for each subscription where you want to enable the connector. Review the guidance on the [resource provider registration status](#) and the ways to register it.
- To enable bi-directional sync, you must have the **Contributor** or **Security Admin** role on the relevant subscription.
- Install the solution for Microsoft Defender for Cloud from the **Content Hub** in Microsoft Sentinel. For more information, see [Discover and manage Microsoft Sentinel out-of-the-box content](#).

## Connect to Microsoft Defender for Cloud

1. In Microsoft Sentinel, select **Data connectors** from the navigation menu.
2. From the data connectors gallery, select **Microsoft Defender for Cloud**, and select [Open connector page](#) in the details pane.
3. Under **Configuration**, you will see a list of the subscriptions in your tenant, and the status of their connection to Microsoft Defender for Cloud. Select the **Status** toggle next to each subscription whose alerts you want to stream into Microsoft Sentinel. If you want to connect several subscriptions at once, you can do this by marking the check boxes next to the relevant subscriptions and then selecting the **Connect** button on the bar above the list.

 **Note**

- The check boxes and **Connect** toggles will be active only on the subscriptions for which you have the required permissions.
- The **Connect** button will be active only if at least one subscription's check box has been marked.

4. To enable bi-directional sync on a subscription, locate the subscription in the list, and choose **Enabled** from the drop-down list in the **Bi-directional sync** column. To enable bi-directional sync on several subscriptions at once, mark their check boxes and select the **Enable bi-directional sync** button on the bar above the list.

 **Note**

- The check boxes and drop-down lists will be active only on the subscriptions for which you have the **required permissions**.
- The **Enable bi-directional sync** button will be active only if at least one subscription's check box has been marked.

5. In the **Microsoft Defender plans** column of the list, you can see if Microsoft Defender plans are enabled on your subscription (a prerequisite for enabling the connector). The value for each subscription in this column will either be blank (meaning no Defender plans are enabled), "All enabled," or "Some enabled." Those that say "Some enabled" will also have an **Enable all** link you can select, that will take you to your Microsoft Defender for Cloud configuration dashboard for that subscription, where you can choose Defender plans to enable. The **Enable Microsoft Defender for all subscriptions** link button on the bar above the list will take you to your Microsoft Defender for Cloud Getting Started page, where you can choose on which subscriptions to enable Microsoft Defender for Cloud altogether.

Instructions Next steps

## Configuration

Connect Microsoft Defender to Microsoft Sentinel

Mark the check box of each Microsoft Defender subscription whose alerts you want to import into Microsoft Sentinel, then select **Connect** above the list.

The connector can be enabled only on subscriptions that have at least one Microsoft Defender plan enabled on Microsoft Defender for Cloud, and only by users with Security Admin / Contributor permissions on the subscription.

Connect Disconnect | Enable bi-directional sync | Disable bi-directional sync | **Enable Azure Defender for all subscriptions >**

contoso

Subscription ↑	Status	Bi-directional sync (Preview)	Azure Defender plans
<input type="checkbox"/> Contoso Hotels	Disconnected	✓	x No permissions
<input type="checkbox"/> Contoso Hotels Tenant - Production	Disconnected	✗ Disabled ✓	All enabled
<input type="checkbox"/> Contoso IT - Retail - Prod	Connected	✗ Disabled ✓	Some enabled <a href="#">Enable all &gt;</a>
<input type="checkbox"/> Contoso demo	Connected	✗ Disabled ✓	Some enabled <a href="#">Enable all &gt;</a>
<input type="checkbox"/> Contoso New ASC Billing	Connected	✗ Disabled ✓	Some enabled <a href="#">Enable all &gt;</a>

**⚠ Microsoft Defender is not enabled on 8 of your subscriptions. Turn on Microsoft Defender to receive alerts.**

6. You can select whether you want the alerts from Microsoft Defender for Cloud to automatically generate incidents in Microsoft Sentinel. Under **Create incidents**, select **Enabled** to turn on the default analytics rule that automatically [creates incidents from alerts](#). You can then edit this rule under **Analytics**, in the **Active rules** tab.

### 💡 Tip

When configuring **custom analytics rules** for alerts from Microsoft Defender for Cloud, consider the alert severity to avoid opening incidents for informational alerts.

Informational alerts in Microsoft Defender for Cloud don't represent a security risk on their own, and are relevant only in the context of an existing, open incident. For more information, see [Security alerts and incidents in Microsoft Defender for Cloud](#).

## Find and analyze your data

### ⓘ Note

Alert synchronization *in both directions* can take a few minutes. Changes in the status of alerts might not be displayed immediately.

- Security alerts are stored in the *SecurityAlert* table in your Log Analytics workspace.

- To query security alerts in Log Analytics, copy the following into your query window as a starting point:

```
Kusto  
  
SecurityAlert  
| where ProductName == "Azure Security Center"
```

- See the **Next steps** tab in the connector page for additional useful sample queries, analytics rule templates, and recommended workbooks.

## Next steps

In this document, you learned how to connect Microsoft Defender for Cloud to Microsoft Sentinel and synchronize alerts between them. To learn more about Microsoft Sentinel, see the following articles:

- Learn how to [get visibility into your data and potential threats](#).
- Get started [detecting threats with Microsoft Sentinel](#).
- Write your own rules to [detect threats](#).

# Connect data from Microsoft Defender XDR to Microsoft Sentinel

Article • 12/07/2023

Microsoft Sentinel's [Microsoft Defender XDR](#) connector with incident integration allows you to stream all Microsoft Defender XDR incidents and alerts into Microsoft Sentinel, and keeps the incidents synchronized between both portals. Microsoft Defender XDR incidents include all their alerts, entities, and other relevant information, and they group together, and are enriched by, alerts from Microsoft Defender XDR's component services **Microsoft Defender for Endpoint**, **Microsoft Defender for Identity**, **Microsoft Defender for Office 365**, **Microsoft Defender for Cloud Apps**, and **Microsoft Defender for Cloud**, as well as alerts from other services such as **Microsoft Purview Data Loss Prevention** and **Microsoft Entra ID Protection**.

The connector also lets you stream **advanced hunting** events from *all* of the above Defender components into Microsoft Sentinel, allowing you to copy those Defender components' advanced hunting queries into Microsoft Sentinel, enrich Sentinel alerts with the Defender components' raw event data to provide additional insights, and store the logs with increased retention in Log Analytics.

For more information about incident integration and advanced hunting event collection, see [Microsoft Defender XDR integration with Microsoft Sentinel](#).

The Microsoft Defender XDR connector is now generally available.

## ⓘ Note

For information about feature availability in US Government clouds, see the Microsoft Sentinel tables in [Cloud feature availability for US Government customers](#).

## Prerequisites

- You must have a valid license for Microsoft Defender XDR, as described in [Microsoft Defender XDR prerequisites](#).
- Your user must be assigned the [Global Administrator](#) or [Security Administrator](#) roles on the tenant you want to stream the logs from.

- Your user must have read and write permissions on your Microsoft Sentinel workspace.
- To make any changes to the connector settings, your user must be a member of the same Microsoft Entra tenant with which your Microsoft Sentinel workspace is associated.
- Install the solution for Microsoft Defender XDR from the **Content Hub** in Microsoft Sentinel. For more information, see [Discover and manage Microsoft Sentinel out-of-the-box content](#).

## Prerequisites for Active Directory sync via MDI

- Your tenant must be onboarded to Microsoft Defender for Identity.
- You must have the MDI sensor installed.

## Connect to Microsoft Defender XDR

In Microsoft Sentinel, select **Data connectors**, select **Microsoft Defender XDR** from the gallery and select **Open connector** page.

The **Configuration** section has three parts:

1. **Connect incidents and alerts** enables the basic integration between Microsoft Defender XDR and Microsoft Sentinel, synchronizing incidents and their alerts between the two platforms.
2. **Connect entities** enables the integration of on-premises Active Directory user identities into Microsoft Sentinel through Microsoft Defender for Identity.
3. **Connect events** enables the collection of raw advanced hunting events from Defender components.

These are explained in greater detail below. See [Microsoft Defender XDR integration with Microsoft Sentinel](#) for more information.

## Connect incidents and alerts

To ingest and synchronize Microsoft Defender XDR incidents, with all their alerts, to your Microsoft Sentinel incidents queue:

1. Mark the check box labeled **Turn off all Microsoft incident creation rules for these products. Recommended**, to avoid duplication of incidents.  
(This check box will not appear once the Microsoft Defender XDR connector is connected.)
2. Select the **Connect incidents & alerts** button.

 **Note**

When you enable the Microsoft Defender XDR connector, all of the Microsoft Defender XDR components' connectors (the ones mentioned at the beginning of this article) are automatically connected in the background. In order to disconnect one of the components' connectors, you must first disconnect the Microsoft Defender XDR connector.

To query Microsoft Defender XDR incident data, use the following statement in the query window:

Kusto

```
SecurityIncident  
| where ProviderName == "Microsoft 365 Defender"
```

## Connect entities

Use Microsoft Defender for Identity to sync user entities from your on-premises Active Directory to Microsoft Sentinel.

Verify that you've satisfied the [prerequisites](#) for syncing on-premises Active Directory users through Microsoft Defender for Identity (MDI).

1. Select the [Go to the UEBA configuration page](#) link.
2. In the **Entity behavior configuration** page, if you haven't yet enabled UEBA, then at the top of the page, move the toggle to **On**.
3. Mark the **Active Directory (Preview)** check box and select **Apply**.

## Entity behavior configuration

X

1. Turn on the UEBA feature to sync Microsoft Sentinel with Azure Active Directory, creating profiles for the users and entities in your organization. [Learn more.](#)



On



Only a Global Administrator or a Security Administrator in your Azure Active Directory can turn this feature on or off

2. Turn on the UEBA feature to sync Microsoft Sentinel with Azure Active Directory.

This feature creates profiles for the users and entities in your organization, and also creates data stores in Microsoft Sentinel

Only tenants onboarded to Microsoft Defender for Identity can enable Active Directory syncing

<input type="checkbox"/>	Active Directory (Preview)
<input checked="" type="checkbox"/>	Azure Active Directory

**Apply**

## Connect events

If you want to collect advanced hunting events from Microsoft Defender for Endpoint or Microsoft Defender for Office 365, the following types of events can be collected from their corresponding advanced hunting tables.

1. Mark the check boxes of the tables with the event types you wish to collect:

Defender for Endpoint

Expand table

Table name	Events type
DeviceInfo	Machine information, including OS information
DeviceNetworkInfo	Network properties of devices, including physical adapters, IP and MAC addresses, as well as connected networks and domains
DeviceProcessEvents	Process creation and related events
DeviceNetworkEvents	Network connection and related events
DeviceFileEvents	File creation, modification, and other file system events
DeviceRegistryEvents	Creation and modification of registry entries
DeviceLogonEvents	Sign-ins and other authentication events on devices
DeviceImageLoadEvents	DLL loading events

Table name	Events type
DeviceEvents	Multiple event types, including events triggered by security controls such as Windows Defender Antivirus and exploit protection
DeviceFileCertificateInfo	Certificate information of signed files obtained from certificate verification events on endpoints

2. Click **Apply Changes**.

3. To query the advanced hunting tables in Log Analytics, enter the table name from the list above in the query window.

## Verify data ingestion

The data graph in the connector page indicates that you are ingesting data. You'll notice that it shows one line each for incidents, alerts, and events, and the events line is an aggregation of event volume across all enabled tables. Once you have enabled the connector, you can use the following KQL queries to generate more specific graphs.

Use the following KQL query for a graph of the incoming Microsoft Defender XDR incidents:

```
Kusto

let Now = now();
(range TimeGenerated from ago(14d) to Now-1d step 1d
| extend Count = 0
| union isfuzzy=true (
    SecurityIncident
    | where ProviderName == "Microsoft 365 Defender"
    | summarize Count = count() by bin_at(TimeGenerated, 1d, Now)
)
| summarize Count=max(Count) by bin_at(TimeGenerated, 1d, Now)
| sort by TimeGenerated
| project Value = iff(isnull(Count), 0, Count), Time = TimeGenerated, Legend = "Events")
| render timechart
```

Use the following KQL query to generate a graph of event volume for a single table (change the *DeviceEvents* table to the required table of your choosing):

```
Kusto
```

```
let Now = now();
(range TimeGenerated from ago(14d) to Now-1d step 1d
| extend Count = 0
| union isfuzzy=true (
    DeviceEvents
    | summarize Count = count() by bin_at(TimeGenerated, 1d, Now)
)
| summarize Count=max(Count) by bin_at(TimeGenerated, 1d, Now)
| sort by TimeGenerated
| project Value = iff(isnull(Count), 0, Count), Time = TimeGenerated, Legend
= "Events"
| render timechart
```

In the **Next steps** tab, you'll find some useful workbooks, sample queries, and analytics rule templates that have been included. You can run them on the spot or modify and save them.

## Next steps

In this document, you learned how to integrate Microsoft Defender XDR incidents, and advanced hunting event data from Microsoft Defender component services, into Microsoft Sentinel, using the Microsoft Defender XDR connector. To learn more about Microsoft Sentinel, see the following articles:

- Learn how to [get visibility into your data, and potential threats](#).
- Get started [detecting threats with Microsoft Sentinel](#).

# Data retention and archive in Azure Monitor Logs

Article • 08/08/2023

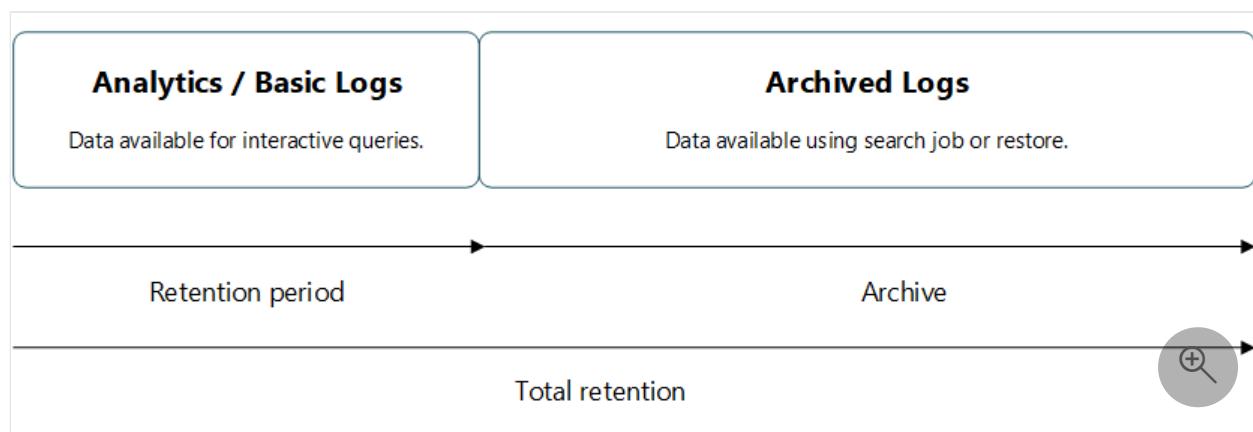
Azure Monitor Logs retains data in two states:

- **Interactive retention:** Lets you retain Analytics logs for [interactive queries](#) of up to 2 years.
- **Archive:** Lets you keep older, less used data in your workspace at a reduced cost. You can access data in the archived state by using [search jobs](#) and [restore](#). You can keep data in archived state for up to 12 years.

This article describes how to configure data retention and archiving.

## How retention and archiving work

Each workspace has a default retention setting that's applied to all tables. You can configure a different retention setting on individual tables.



During the interactive retention period, data is available for monitoring, troubleshooting, and analytics. When you no longer use the logs, but still need to keep the data for compliance or occasional investigation, archive the logs to save costs.

Archived data stays in the same table, alongside the data that's available for interactive queries. When you set a total retention period that's longer than the interactive retention period, Log Analytics automatically archives the relevant data immediately at the end of the retention period.

You can access archived data by [running a search job](#) or [restoring archived logs](#).

### ⓘ Note

The archive period can only be set at the table level, not at the workspace level.

## Adjustments to retention and archive settings

When you shorten an existing retention setting, Azure Monitor waits 30 days before removing the data, so you can revert the change and avoid data loss in the event of an error in configuration. You can [purge data](#) immediately when required.

When you increase the retention setting, the new retention period applies to all data that's already been ingested into the table and hasn't yet been purged or removed.

If you change the archive settings on a table with existing data, the relevant data in the table is also affected immediately. For example, you might have an existing table with 180 days of interactive retention and no archive period. You decide to change the retention setting to 90 days of interactive retention without changing the total retention period of 180 days. Log Analytics immediately archives any data that's older than 90 days and none of the data is deleted.

## Permissions required

[ ] Expand table

Action	Permissions required
Configure data retention and archive policies for a Log Analytics workspace	<code>Microsoft.OperationalInsights/workspaces/write</code> and <code>microsoft.operationalinsights/workspaces/tables/write</code> permissions to the Log Analytics workspace, as provided by the <a href="#">Log Analytics Contributor built-in role</a> , for example
Get the retention and archive policy by table for a Log Analytics workspace	<code>Microsoft.OperationalInsights/workspaces/tables/read</code> permissions to the Log Analytics workspace, as provided by the <a href="#">Log Analytics Reader built-in role</a> , for example
Purge data from a Log Analytics workspace	<code>Microsoft.OperationalInsights/workspaces/purge/action</code> permissions to the Log Analytics workspace, as provided by the <a href="#">Log Analytics Contributor built-in role</a> , for example
Set data retention for a classic Application Insights resource	<code>microsoft.insights/components/write</code> permissions to the classic Application Insights resource, as provided by the <a href="#">Application Insights Component Contributor built-in role</a> , for example

Action	Permissions required
Purge data from a classic Application Insights resource	<code>Microsoft.Insights/components/purge/action</code> permissions to the classic Application Insights resource, as provided by the <a href="#">Application Insights Component Contributor built-in role</a> , for example

## Configure the default workspace retention

You can set a Log Analytics workspace's default retention in the Azure portal to 30, 31, 60, 90, 120, 180, 270, 365, 550, and 730 days. You can apply a different setting to specific tables by [configuring retention and archive at the table level](#). If you're on the *free* tier, you need to upgrade to the paid tier to change the data retention period.

To set the default workspace retention:

1. From the **Log Analytics workspaces** menu in the Azure portal, select your workspace.
2. Select **Usage and estimated costs** in the left pane.
3. Select **Data Retention** at the top of the page.

The screenshot shows the Azure portal interface. On the left, the 'Usage and estimated costs' blade is open, displaying 'Pricing Tiers' (Pay-as-you-go) and 'Estimated costs' for Log data ingestion, Log data retention, and a Total. It also shows a note about being the current pricing tier. On the right, a 'Data Retention' dialog box is open, showing a chart of billable data ingestion per solution and a slider for Data Retention (Days) set to 31. A note states that 31 days are included with the pricing plan, and longer retention incurs additional charges. It also mentions that Application Insights data types default to 90 days. An 'OK' button and a magnifying glass icon are visible at the bottom right of the dialog.

4. Move the slider to increase or decrease the number of days, and then select **OK**.

## Configure retention and archive at the table level

By default, all tables in your workspace inherit the workspace's interactive retention setting and have no archive. You can modify the retention and archive settings of individual tables, except for workspaces in the legacy Free Trial pricing tier.

The Analytics log data plan includes 30 days of interactive retention. You can increase the interactive retention period to up to 730 days at an [additional cost](#). If needed, you can reduce the interactive retention period to as little as four days using the API or CLI. However, since 30 days are included in the ingestion price, lowering the retention period below 30 days doesn't reduce costs. You can set the archive period to a total retention time of up to 4,383 days (12 years).

### ⓘ Note

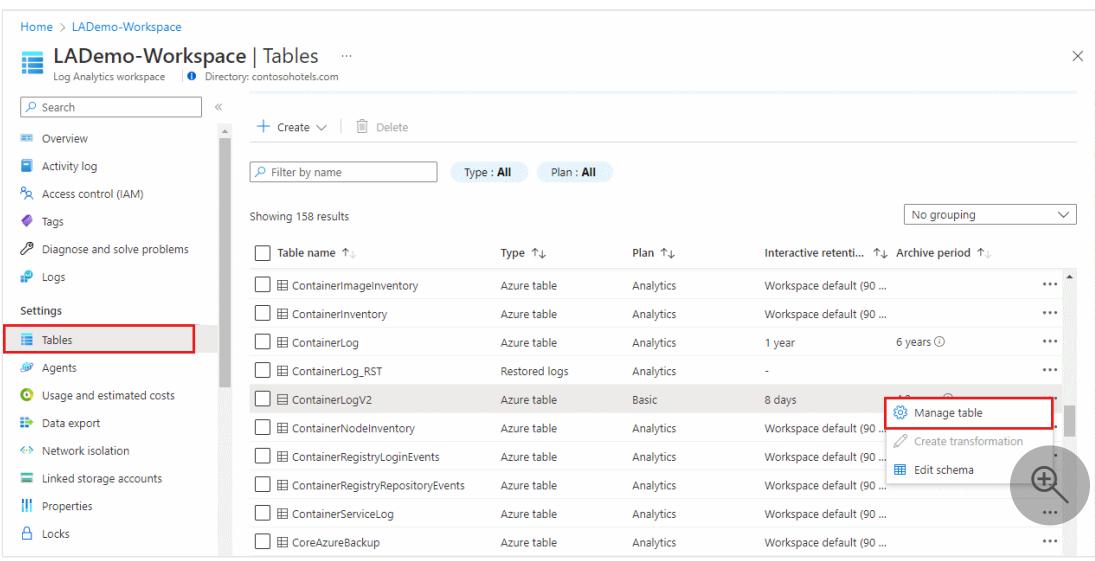
Currently, you can set total retention to up to 12 years through the Azure portal and API. CLI and PowerShell are limited to seven years; support for 12 years will follow.

Portal

To set the retention and archive duration for a table in the Azure portal:

1. From the **Log Analytics workspaces** menu, select **Tables**.

The **Tables** screen lists all the tables in the workspace.
2. Select the context menu for the table you want to configure and select **Manage table**.
3. Configure the retention and archive duration in the **Data retention settings** section of the table configuration screen.



Kubernetes Container logs in V2 schema. This is the successor for ContainerLog. This has more friendlier schema, specifically for kubernetes orchestrated containers in pods.

**Data retention settings**

- Interactive retention: 8 days
- Total retention period: 4 years
- Archive period of 1452 days (4.0 years)

## Get retention and archive settings by table

### Portal

To view the retention and archive duration for a table in the Azure portal, from the **Log Analytics workspaces** menu, select **Tables**.

The **Tables** screen shows the interactive retention and archive period for all the tables in the workspace.

Table Name	Type	Plan	Interactive retention	Archive period
ContainerImageInventory	Azure table	Analytics	Workspace default (90 d...)	...
ContainerInventory	Azure table	Analytics	Workspace default (90 d...)	...
ContainerLog	Azure table	Analytics	1 year	6 years ⓘ
ContainerLog_RST	Restored logs	Analytics	-	...
ContainerLogV2	Azure table	Basic	8 days	4.0 years ⓘ
ContainerNodeInventory	Azure table	Analytics	Workspace default (90 d...)	...
ContainerRegistryLoginEvents	Azure table	Analytics	Workspace default (90 d...)	...
ContainerRegistryRepositoryEvents	Azure table	Analytics	Workspace default (90 d...)	...
ContainerServiceLog	Azure table	Analytics	Workspace default (90 d...)	...

## Purge retained data

If you set the data retention to 30 days, you can purge older data immediately by using the `immediatePurgeDataOn30Days` parameter in Azure Resource Manager. The purge

functionality is useful when you need to remove personal data immediately. The immediate purge functionality isn't available through the Azure portal.

Workspaces with a 30-day retention might keep data for 31 days if you don't set the `immediatePurgeDataOn30Days` parameter.

You can also purge data from a workspace by using the [purge feature](#), which removes personal data. You can't purge data from archived logs.

### Important

The Log Analytics [Purge feature](#) doesn't affect your retention costs. To lower retention costs, decrease the retention period for the workspace or for specific tables.

## Tables with unique retention periods

By default, two data types, `Usage` and `AzureActivity`, keep data for at least 90 days at no charge. When you increase the workspace retention to more than 90 days, you also increase the retention of these data types. These tables are also free from data ingestion charges.

Tables related to Application Insights resources also keep data for 90 days at no charge. You can adjust the retention of each of these tables individually:

- `AppAvailabilityResults`
- `AppBrowserTimings`
- `AppDependencies`
- `AppExceptions`
- `AppEvents`
- `AppMetrics`
- `AppPageViews`
- `AppPerformanceCounters`
- `AppRequests`
- `AppSystemEvents`
- `AppTraces`

## Pricing model

The charge for maintaining archived logs is calculated based on the volume of data you archive, in GB, and the number of days for which you archive the data.

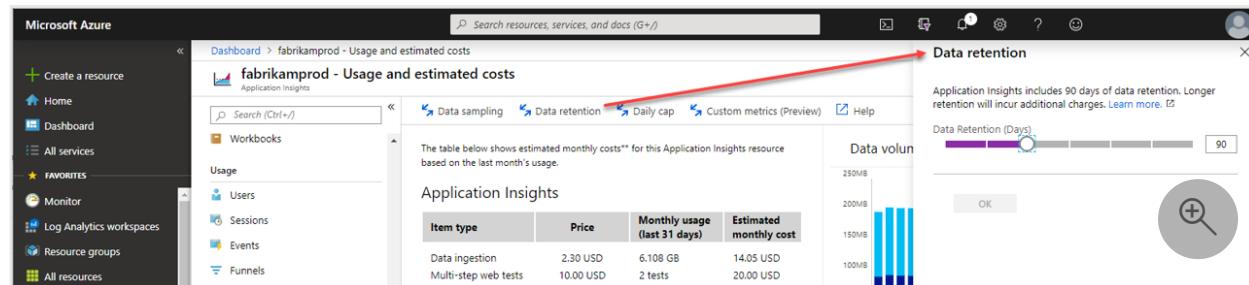
For more information, see [Azure Monitor pricing](#).

## Set data retention for classic Application Insights resources

Workspace-based Application Insights resources store data in a Log Analytics workspace, so it's included in the data retention and archive settings for the workspace. Classic Application Insights resources have separate retention settings.

The default retention for Application Insights resources is 90 days. You can select different retention periods for each Application Insights resource. The full set of available retention periods is 30, 60, 90, 120, 180, 270, 365, 550, or 730 days.

To change the retention, from your Application Insights resource, go to the **Usage and estimated costs** page and select the **Data retention** option.



A several-day grace period begins when the retention is lowered before the oldest data is removed.

The retention can also be [set programmatically with PowerShell](#) by using the `retentionInDays` parameter. If you set the data retention to 30 days, you can trigger an immediate purge of older data by using the `immediatePurgeDataOn30Days` parameter.

This approach might be useful for compliance-related scenarios. This purge functionality is only exposed via Azure Resource Manager and should be used with extreme care. The daily reset time for the data volume cap can be configured by using Azure Resource Manager to set the `dailyQuotaResetTime` parameter.

## Next steps

- Learn more about Log Analytics workspaces and data retention and archive
- Create a search job to retrieve archive data matching particular criteria

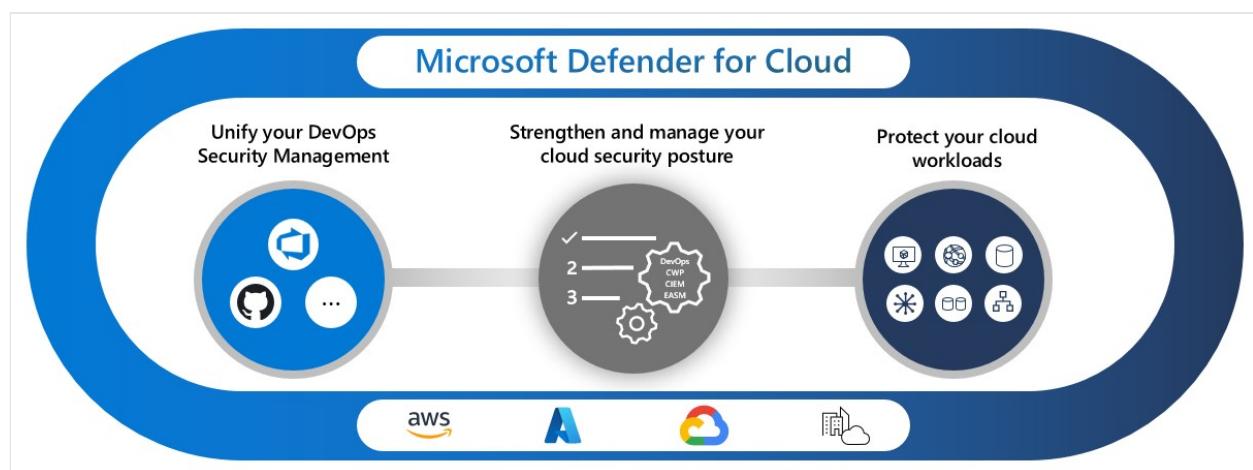
- Restore archive data within a particular time range

# What is Microsoft Defender for Cloud?

Article • 11/16/2023

Microsoft Defender for Cloud is a cloud-native application protection platform (CNAPP) that is made up of security measures and practices that are designed to protect cloud-based applications from various cyber threats and vulnerabilities. Defender for Cloud combines the capabilities of:

- A development security operations (DevSecOps) solution that unifies security management at the code level across multicloud and multiple-pipeline environments
- A cloud security posture management (CSPM) solution that surfaces actions that you can take to prevent breaches
- A cloud workload protection platform (CWPP) with specific protections for servers, containers, storage, databases, and other workloads



## ⓘ Note

For Defender for Cloud pricing information, see the [pricing page](#).

When you [enable Defender for Cloud](#), you automatically gain access to Microsoft 365 Defender.

The Microsoft 365 Defender portal helps security teams investigate attacks across cloud resources, devices, and identities. Microsoft 365 Defender provides an overview of attacks, including suspicious and malicious events that occur in cloud environments. Microsoft 365 Defender accomplishes this goal by correlating all alerts and incidents, including cloud alerts and incidents.

You can learn more about the [integration between Microsoft Defender for Cloud and Microsoft Defender XDR](#).

## Secure cloud applications

Defender for Cloud helps you to incorporate good security practices early during the software development process, or DevSecOps. You can protect your code management environments and your code pipelines, and get insights into your development environment security posture from a single location. Defender for Cloud empowers security teams to manage DevOps security across multi-pipeline environments.

Today's applications require security awareness at the code, infrastructure, and runtime levels to make sure that deployed applications are hardened against attacks.

[+] [Expand table](#)

Capability	What problem does it solve?	Get started	Defender plan
<a href="#">Code pipeline insights</a>	Empowers security teams with the ability to protect applications and resources from code to cloud across multi-pipeline environments, including GitHub, Azure DevOps, and GitLab. DevOps security findings, such as Infrastructure as Code (IaC) misconfigurations and exposed secrets, can then be correlated with other contextual cloud security insights to prioritize remediation in code.	Connect <a href="#">Azure DevOps</a> , <a href="#">GitHub</a> , and <a href="#">GitLab</a> repositories to Defender for Cloud	Foundational CSPM (Free) and Defender CSPM

## Improve your security posture

The security of your cloud and on-premises resources depends on proper configuration and deployment. Defender for Cloud recommendations identifies the steps that you can take to secure your environment.

Defender for Cloud includes Foundational CSPM capabilities for free. You can also enable advanced CSPM capabilities by enabling the Defender CSPM plan.

[+] [Expand table](#)

Capability	What problem does it solve?	Get started	Defender plan
<a href="#">Centralized policy</a>	Define the security conditions that you want to maintain across your environment.	<a href="#">Customize a security policy</a>	Foundational CSPM (Free)

Management capability	What problem does it solve?	Get started	Defender plan
	The policy translates to recommendations that identify resource configurations that violate your security policy. The <a href="#">Microsoft cloud security benchmark</a> is a built-in standard that applies security principles with detailed technical implementation guidance for Azure and other cloud providers (such as AWS and GCP).		
Secure score	Summarize your security posture based on the security recommendations. As you remediate recommendations, your secure score improves.	<a href="#">Track your secure score</a>	Foundational CSPM (Free)
Multicloud coverage	Connect to your multicloud environments with agentless methods for CSPM insight and CWP protection.	Connect your <a href="#">Amazon AWS</a> and <a href="#">Google GCP</a> cloud resources to Defender for Cloud	Foundational CSPM (Free)
Cloud Security Posture Management (CSPM)	Use the dashboard to see weaknesses in your security posture.	<a href="#">Enable CSPM tools</a>	Foundational CSPM (Free)
Advanced Cloud Security Posture Management	Get advanced tools to identify weaknesses in your security posture, including: - Governance to drive actions to improve your security posture - Regulatory compliance to verify compliance with security standards - Cloud security explorer to build a comprehensive view of your environment	<a href="#">Enable CSPM tools</a>	Defender CSPM
Data-aware Security Posture	Data-aware security posture automatically discovers datastores containing sensitive data, and helps reduce risk of data breaches.	<a href="#">Enable data-aware security posture</a>	Defender CSPM or Defender for Storage
Attack path analysis	Model traffic on your network to identify potential risks before you implement changes to your environment.	<a href="#">Build queries to analyze paths</a>	Defender CSPM
Cloud Security Explorer	A map of your cloud environment that lets you build queries to find security risks.	<a href="#">Build queries to find security risks</a>	Defender CSPM
Security governance	Drive security improvements through your organization by assigning tasks to resource	<a href="#">Define governance</a>	Defender CSPM

Capability	What problem does it solve?	Get started	Defender plan
Microsoft Entra Permissions Management	Provide comprehensive visibility and control over permissions for any identity and any resource in Azure, AWS, and GCP.	<a href="#">Review your Permission Creep Index (CPI)</a>	Defender CSPM

## Protect cloud workloads

Proactive security principles require that you implement security practices that protect your workloads from threats. Cloud workload protections (CWP) surface workload-specific recommendations that lead you to the right security controls to protect your workloads.

When your environment is threatened, security alerts right away indicate the nature and severity of the threat so you can plan your response. After you identify a threat in your environment, you need to quickly respond to limit the risk to your resources.

[\[+\] Expand table](#)

Capability	What problem does it solve?	Get started	Defender plan
Protect cloud servers	Provide server protections through Microsoft Defender for Endpoint or extended protection with just-in-time network access, file integrity monitoring, vulnerability assessment, and more.	<a href="#">Secure your multicloud and on-premises servers</a>	Defender for Servers
Identify threats to your storage resources	Detect unusual and potentially harmful attempts to access or exploit your storage accounts using advanced threat detection capabilities and Microsoft Threat Intelligence data to provide contextual security alerts.	<a href="#">Protect your cloud storage resources</a>	Defender for Storage
Protect cloud databases	Protect your entire database estate with attack detection and threat response for the most popular database types in Azure to protect the database engines and data types, according to their attack surface and security risks.	<a href="#">Deploy specialized protections for cloud and on-premises databases</a>	- Defender for Azure SQL Databases - Defender for SQL servers on machines - Defender for Open-source relational

Capability	What problem does it solve?	Get started	Defender plan
			databases - Defender for Azure Cosmos DB
Protect containers	Secure your containers so you can improve, monitor, and maintain the security of your clusters, containers, and their applications with environment hardening, vulnerability assessments, and run-time protection.	<a href="#">Find security risks in your containers</a>	Defender for Containers
Infrastructure service insights	Diagnose weaknesses in your application infrastructure that can leave your environment susceptible to attack.	<ul style="list-style-type: none"> <li>- <a href="#">Identify attacks targeting applications running over App Service</a></li> <li>- <a href="#">Detect attempts to exploit Key Vault accounts</a></li> <li>- <a href="#">Get alerted on suspicious Resource Manager operations</a></li> <li>- <a href="#">Expose anomalous DNS activities</a></li> </ul>	<ul style="list-style-type: none"> <li>- Defender for App Service</li> <li>- Defender for Key Vault</li> <li>- Defender for Resource Manager</li> <li>- Defender for DNS</li> </ul>
Security alerts	Get informed of real-time events that threaten the security of your environment. Alerts are categorized and assigned severity levels to indicate proper responses.	<a href="#">Manage security alerts</a>	Any workload protection Defender plan
Security incidents	Correlate alerts to identify attack patterns and integrate with Security Information and Event Management (SIEM), Security Orchestration Automated Response (SOAR), and IT Service Management (ITSM) solutions to respond to threats and limit the risk to your resources.	<a href="#">Export alerts to SIEM, SOAR, or ITSM systems</a>	Any workload protection Defender plan

 **Important**

As of August 1, customers with an existing subscription to Defender for DNS can continue to use the service, but new subscribers will receive alerts about suspicious DNS activity as part of Defender for Servers P2.

## Learn More

For more information about Defender for Cloud and how it works, check out:

- A [step-by-step walkthrough](#) of Defender for Cloud
- An interview about Defender for Cloud with an expert in cybersecurity in [Lessons Learned from the Field](#)
- [Microsoft Defender for Cloud - Use cases](#)
- [Microsoft Defender for Cloud PoC Series - Microsoft Defender for Containers](#)

## Next steps

[Enable Microsoft Defender plans](#)

# Security recommendations - a reference guide

Article • 10/05/2023

This article lists the recommendations you might see in Microsoft Defender for Cloud. The recommendations shown in your environment depend on the resources you're protecting and your customized configuration.

Recommendations in Defender for Cloud are based on the [Microsoft cloud security benchmark](#). the Microsoft cloud security benchmark is the Microsoft-authored set of guidelines for security and compliance best practices based on common compliance frameworks. This widely respected benchmark builds on the controls from the [Center for Internet Security \(CIS\)](#) and the [National Institute of Standards and Technology \(NIST\)](#) with a focus on cloud-centric security.

To learn about how to respond to these recommendations, see [Remediate recommendations in Defender for Cloud](#).

Your secure score is based on the number of security recommendations you've completed. To decide which recommendations to resolve first, look at the severity of each one and its potential impact on your secure score.

## 💡 Tip

If a recommendation's description says "No related policy", it's usually because that recommendation is dependent on a different recommendation and *its* policy. For example, the recommendation "Endpoint protection health failures should be remediated...", relies on the recommendation that checks whether an endpoint protection solution is even *installed* ("Endpoint protection solution should be installed..."). The underlying recommendation *does* have a policy. Limiting the policies to only the foundational recommendation simplifies policy management.

## AppServices recommendations

There are 26 recommendations in this category.

 Expand table

Recommendation	Description	Severity
<a href="#">API App should only be accessible over HTTPS ↗</a>	<p>Use of HTTPS ensures server/service authentication and protects data in transit from network layer eavesdropping attacks.</p> <p>(Related policy: <a href="#">API App should only be accessible over HTTPS ↗</a>)</p>	Medium
<a href="#">CORS should not allow every resource to access API Apps ↗</a>	<p>Cross-Origin Resource Sharing (CORS) should not allow all domains to access your API app. Allow only required domains to interact with your API app.</p> <p>(Related policy: <a href="#">CORS should not allow every resource to access your API App ↗</a>)</p>	Low
<a href="#">CORS should not allow every resource to access Function Apps ↗</a>	<p>Cross-Origin Resource Sharing (CORS) should not allow all domains to access your Function app. Allow only required domains to interact with your Function app.</p> <p>(Related policy: <a href="#">CORS should not allow every resource to access your Function Apps ↗</a>)</p>	Low
<a href="#">CORS should not allow every resource to access Web Applications ↗</a>	<p>Cross-Origin Resource Sharing (CORS) should not allow all domains to access your web application. Allow only required domains to interact with your web app.</p> <p>(Related policy: <a href="#">CORS should not allow every resource to access your Web Applications ↗</a>)</p>	Low
<a href="#">Diagnostic logs in App Service should be enabled ↗</a>	<p>Audit enabling of diagnostic logs on the app. This enables you to recreate activity trails for investigation purposes if a security incident occurs or your network is compromised</p> <p>(No related policy)</p>	Medium
<a href="#">Ensure API app has Client Certificates Incoming client certificates set to On ↗</a>	<p>Client certificates allow for the app to request a certificate for incoming requests. Only clients that have a valid certificate will be able to reach the app.</p> <p>(Related policy: <a href="#">Ensure API app has 'Client Certificates (Incoming client certificates)' set to 'On' ↗</a>)</p>	Medium
<a href="#">FTPS should be required in API apps ↗</a>	<p>Enable FTPS enforcement for enhanced security</p> <p>(Related policy: <a href="#">FTPS only should be required in your API App ↗</a>)</p>	High
<a href="#">FTPS should be required in function apps ↗</a>	<p>Enable FTPS enforcement for enhanced security</p> <p>(Related policy: <a href="#">FTPS only should be required in your Function App ↗</a>)</p>	High
<a href="#">FTPS should be required in web apps ↗</a>	<p>Enable FTPS enforcement for enhanced security</p> <p>(Related policy: <a href="#">FTPS should be required in your Web</a></p>	High

Recommendation	Description	Severity
	<p><a href="#">App ↗</a></p> <p><b>Function App should only be accessible over HTTPS ↗</b></p> <p>Use of HTTPS ensures server/service authentication and protects data in transit from network layer eavesdropping attacks.</p> <p>(Related policy: <a href="#">Function App should only be accessible over HTTPS ↗</a>)</p>	Medium
<p><b>Function apps should have Client Certificates (Incoming client certificates) enabled ↗</b></p>	<p>Client certificates allow for the app to request a certificate for incoming requests. Only clients with valid certificates will be able to reach the app.</p> <p>(Related policy: <a href="#">Function apps should have 'Client Certificates (Incoming client certificates)' enabled ↗</a>)</p>	Medium
<p><b>Java should be updated to the latest version for API apps ↗</b></p>	<p>Periodically, newer versions are released for Java either due to security flaws or to include additional functionality.</p> <p>Using the latest Python version for API apps is recommended to benefit from security fixes, if any, and/or new functionalities of the latest version.</p> <p>(Related policy: <a href="#">Ensure that 'Java version' is the latest, if used as a part of the API app ↗</a>)</p>	Medium
<p><b>Managed identity should be used in API apps ↗</b></p>	<p>For enhanced authentication security, use a managed identity.</p> <p>On Azure, managed identities eliminate the need for developers to have to manage credentials by providing an identity for the Azure resource in Azure AD and using it to obtain Azure Active Directory (Azure AD) tokens.</p> <p>(Related policy: <a href="#">Managed identity should be used in your API App ↗</a>)</p>	Medium
<p><b>Managed identity should be used in function apps ↗</b></p>	<p>For enhanced authentication security, use a managed identity.</p> <p>On Azure, managed identities eliminate the need for developers to have to manage credentials by providing an identity for the Azure resource in Azure AD and using it to obtain Azure Active Directory (Azure AD) tokens.</p> <p>(Related policy: <a href="#">Managed identity should be used in your Function App ↗</a>)</p>	Medium
<p><b>Managed identity should be used in web apps ↗</b></p>	<p>For enhanced authentication security, use a managed identity.</p> <p>On Azure, managed identities eliminate the need for developers to have to manage credentials by providing an identity for the Azure resource in Azure AD and using it to obtain Azure Active Directory</p>	Medium

Recommendation	Description	Severity
<a href="#">Microsoft Defender for App Service should be enabled ↗</a>	<p>(Azure AD) tokens.</p> <p>(Related policy: <a href="#">Managed identity should be used in your Web App ↗</a>)</p>	
<a href="#">PHP should be updated to the latest version for API apps ↗</a>	<p>Microsoft Defender for App Service leverages the scale of the cloud, and the visibility that Azure has as a cloud provider, to monitor for common web app attacks.</p> <p>Microsoft Defender for App Service can discover attacks on your applications and identify emerging attacks.</p> <p>Important: Remediating this recommendation will result in charges for protecting your App Service plans. If you don't have any App Service plans in this subscription, no charges will be incurred.</p> <p>If you create any App Service plans on this subscription in the future, they will automatically be protected and charges will begin at that time.</p> <p>Learn more in <a href="#">Protect your web apps and APIs</a>.</p> <p>(Related policy: <a href="#">Azure Defender for App Service should be enabled ↗</a>)</p>	High
<a href="#">Python should be updated to the latest version for API apps ↗</a>	<p>Periodically, newer versions are released for PHP software either due to security flaws or to include additional functionality.</p> <p>Using the latest PHP version for API apps is recommended to benefit from security fixes, if any, and/or new functionalities of the latest version.</p> <p>(Related policy: <a href="#">Ensure that 'PHP version' is the latest, if used as a part of the API app ↗</a>)</p>	Medium
<a href="#">Remote debugging should be turned off for API App ↗</a>	<p>Periodically, newer versions are released for Python software either due to security flaws or to include additional functionality.</p> <p>Using the latest Python version for API apps is recommended to benefit from security fixes, if any, and/or new functionalities of the latest version.</p> <p>(Related policy: <a href="#">Ensure that 'Python version' is the latest, if used as a part of the API app ↗</a>)</p>	Medium
<a href="#">Remote debugging should be turned off for Function App ↗</a>	<p>Remote debugging requires inbound ports to be opened on an API app. Remote debugging should be turned off.</p> <p>(Related policy: <a href="#">Remote debugging should be turned off for API Apps ↗</a>)</p>	Low

Recommendation	Description	Severity
	<p>Debugging should be turned off.</p> <p>(Related policy: <a href="#">Remote debugging should be turned off for Function Apps</a>)</p>	
Remote debugging should be turned off for Web Applications	<p>Remote debugging requires inbound ports to be opened on a web application. Remote debugging is currently enabled. If you no longer need to use remote debugging, it should be turned off.</p> <p>(Related policy: <a href="#">Remote debugging should be turned off for Web Applications</a>)</p>	Low
TLS should be updated to the latest version for API apps	<p>Upgrade to the latest TLS version</p> <p>(Related policy: <a href="#">Latest TLS version should be used in your API App</a>)</p>	High
TLS should be updated to the latest version for function apps	<p>Upgrade to the latest TLS version</p> <p>(Related policy: <a href="#">Latest TLS version should be used in your Function App</a>)</p>	High
TLS should be updated to the latest version for web apps	<p>Upgrade to the latest TLS version</p> <p>(Related policy: <a href="#">Latest TLS version should be used in your Web App</a>)</p>	High
Web Application should only be accessible over HTTPS	<p>Use of HTTPS ensures server/service authentication and protects data in transit from network layer eavesdropping attacks.</p> <p>(Related policy: <a href="#">Web Application should only be accessible over HTTPS</a>)</p>	Medium
Web apps should request an SSL certificate for all incoming requests	<p>Client certificates allow for the app to request a certificate for incoming requests.</p> <p>Only clients that have a valid certificate will be able to reach the app.</p> <p>(Related policy: <a href="#">Ensure WEB app has 'Client Certificates (Incoming client certificates)' set to 'On'</a>)</p>	Medium

## Compute recommendations

There are 58 recommendations in this category.

[Expand table](#)

Recommendation	Description	Severity
Adaptive application controls for defining	Enable application controls to define the list of known-safe applications running on your machines, and alert you when	High

Recommendation	Description	Severity
<a href="#">safe applications should be enabled on your machines ↗</a>	other applications run. This helps harden your machines against malware. To simplify the process of configuring and maintaining your rules, Defender for Cloud uses machine learning to analyze the applications running on each machine and suggest the list of known-safe applications. (Related policy: <a href="#">Adaptive application controls for defining safe applications should be enabled on your machines ↗</a> )	
<a href="#">Allowlist rules in your adaptive application control policy should be updated ↗</a>	Monitor for changes in behavior on groups of machines configured for auditing by Defender for Cloud's adaptive application controls. Defender for Cloud uses machine learning to analyze the running processes on your machines and suggest a list of known-safe applications. These are presented as recommended apps to allow in adaptive application control policies. (Related policy: <a href="#">Allowlist rules in your adaptive application control policy should be updated ↗</a> )	High
<a href="#">Authentication to Linux machines should require SSH keys ↗</a>	Although SSH itself provides an encrypted connection, using passwords with SSH still leaves the VM vulnerable to brute-force attacks. The most secure option for authenticating to an Azure Linux virtual machine over SSH is with a public-private key pair, also known as SSH keys. Learn more in <a href="#">Detailed steps: Create and manage SSH keys for authentication to a Linux VM in Azure</a> . (Related policy: <a href="#">Audit Linux machines that are not using SSH key for authentication ↗</a> )	Medium
<a href="#">Automation account variables should be encrypted ↗</a>	It is important to enable encryption of Automation account variable assets when storing sensitive data. (Related policy: <a href="#">Automation account variables should be encrypted ↗</a> )	High
<a href="#">Azure Backup should be enabled for virtual machines ↗</a>	Protect the data on your Azure virtual machines with Azure Backup. Azure Backup is an Azure-native, cost-effective, data protection solution. It creates recovery points that are stored in geo-redundant recovery vaults. When you restore from a recovery point, you can restore the whole VM or specific files. (Related policy: <a href="#">Azure Backup should be enabled for Virtual Machines ↗</a> )	Low
<a href="#">Container hosts should be configured securely ↗</a>	Remediate vulnerabilities in security configuration on machines with Docker installed to protect them from attacks.	High

Recommendation	Description	Severity
	(Related policy: <a href="#">Vulnerabilities in container security configurations should be remediated ↗</a> )	
<a href="#">Diagnostic logs in Azure Stream Analytics should be enabled ↗</a>	Enable logs and retain them for up to a year. This enables you to recreate activity trails for investigation purposes when a security incident occurs or your network is compromised. (Related policy: <a href="#">Diagnostic logs in Azure Stream Analytics should be enabled ↗</a> )	Low
<a href="#">Diagnostic logs in Batch accounts should be enabled ↗</a>	Enable logs and retain them for up to a year. This enables you to recreate activity trails for investigation purposes when a security incident occurs or your network is compromised. (Related policy: <a href="#">Diagnostic logs in Batch accounts should be enabled ↗</a> )	Low
<a href="#">Diagnostic logs in Event Hubs should be enabled ↗</a>	Enable logs and retain them for up to a year. This enables you to recreate activity trails for investigation purposes when a security incident occurs or your network is compromised. (Related policy: <a href="#">Diagnostic logs in Event Hubs should be enabled ↗</a> )	Low
<a href="#">Diagnostic logs in Logic Apps should be enabled ↗</a>	To ensure you can recreate activity trails for investigation purposes when a security incident occurs or your network is compromised, enable logging. If your diagnostic logs aren't being sent to a Log Analytics workspace, Azure Storage account, or Azure Event Hubs, ensure you've configured diagnostic settings to send platform metrics and platform logs to the relevant destinations. Learn more in <a href="#">Create diagnostic settings to send platform logs and metrics to different destinations</a> . (Related policy: <a href="#">Diagnostic logs in Logic Apps should be enabled ↗</a> )	Low
<a href="#">Diagnostic logs in Search services should be enabled ↗</a>	Enable logs and retain them for up to a year. This enables you to recreate activity trails for investigation purposes when a security incident occurs or your network is compromised. (Related policy: <a href="#">Diagnostic logs in Search services should be enabled ↗</a> )	Low
<a href="#">Diagnostic logs in Service Bus should be enabled ↗</a>	Enable logs and retain them for up to a year. This enables you to recreate activity trails for investigation purposes when a security incident occurs or your network is compromised.	Low

Recommendation	Description	Severity
	(Related policy: <a href="#">Diagnostic logs in Service Bus should be enabled</a> )	
<a href="#">Diagnostic logs in Virtual Machine Scale Sets should be enabled</a>	Enable logs and retain them for up to a year. This enables you to recreate activity trails for investigation purposes when a security incident occurs or your network is compromised.  (Related policy: <a href="#">Diagnostic logs in Virtual Machine Scale Sets should be enabled</a> )	Low
<a href="#">Endpoint protection health issues on machines should be resolved</a>	Resolve endpoint protection health issues on your virtual machines to protect them from latest threats and vulnerabilities. See the documentation for the <a href="#">endpoint protection solutions supported by Defender for Cloud</a> and the <a href="#">endpoint protection assessments</a> .  (No related policy)	Medium
<a href="#">Endpoint protection health issues on virtual machine scale sets should be resolved</a>	Remediate endpoint protection health failures on your virtual machine scale sets to protect them from threats and vulnerabilities.  (Related policy: <a href="#">Endpoint protection solution should be installed on virtual machine scale sets</a> )	Low
<a href="#">Endpoint protection should be installed on machines</a>	To protect machines from threats and vulnerabilities, install a supported endpoint protection solution.  Learn more about how endpoint protection for machines is evaluated in <a href="#">Endpoint protection assessment and recommendations in Microsoft Defender for Cloud</a> .  (No related policy)	High
<a href="#">Endpoint protection should be installed on virtual machine scale sets</a>	Install an endpoint protection solution on your virtual machines scale sets, to protect them from threats and vulnerabilities.  (Related policy: <a href="#">Endpoint protection solution should be installed on virtual machine scale sets</a> )	High
<a href="#">File integrity monitoring should be enabled on machines</a>	Defender for Cloud has identified machines that are missing a file integrity monitoring solution. To monitor changes to critical files, registry keys, and more on your servers, enable file integrity monitoring.  When the file integrity monitoring solution is enabled, create data collection rules to define the files to be monitored. To define rules, or see the files changed on machines with existing rules, go to the <a href="#">file integrity monitoring management page</a>  (No related policy)	High
<a href="#">Guest Attestation extension should be</a>	Install Guest Attestation extension on supported Linux virtual machine scale sets to allow Microsoft Defender for	Low

Recommendation	Description	Severity
<a href="#">installed on supported Linux virtual machine scale sets ↗</a>	<p>Cloud to proactively attest and monitor the boot integrity. Once installed, boot integrity will be attested via Remote Attestation. This assessment only applies to trusted launch enabled Linux virtual machine scale sets.</p>	
	<p><b>Important:</b> Trusted launch requires the creation of new virtual machines. You can't enable trusted launch on existing virtual machines that were initially created without it. Learn more about <a href="#">Trusted launch for Azure virtual machines</a>. (No related policy)</p>	
<a href="#">Guest Attestation extension should be installed on supported Linux virtual machines ↗</a>	<p>Install Guest Attestation extension on supported Linux virtual machines to allow Microsoft Defender for Cloud to proactively attest and monitor the boot integrity. Once installed, boot integrity will be attested via Remote Attestation. This assessment only applies to trusted launch enabled Linux virtual machines.</p>	Low
	<p><b>Important:</b> Trusted launch requires the creation of new virtual machines. You can't enable trusted launch on existing virtual machines that were initially created without it. Learn more about <a href="#">Trusted launch for Azure virtual machines</a>. (No related policy)</p>	
<a href="#">Guest Attestation extension should be installed on supported Windows virtual machine scale sets ↗</a>	<p>Install Guest Attestation extension on supported virtual machine scale sets to allow Microsoft Defender for Cloud to proactively attest and monitor the boot integrity. Once installed, boot integrity will be attested via Remote Attestation. This assessment only applies to trusted launch enabled virtual machine scale sets.</p>	Low
	<p><b>Important:</b> Trusted launch requires the creation of new virtual machines. You can't enable trusted launch on existing virtual machines that were initially created without it. Learn more about <a href="#">Trusted launch for Azure virtual machines</a>. (No related policy)</p>	
<a href="#">Guest Attestation extension should be</a>	<p>Install Guest Attestation extension on supported virtual machines to allow Microsoft Defender for Cloud to</p>	Low

Recommendation	Description	Severity
<a href="#">installed on supported Windows virtual machines ↗</a>	<p>proactively attest and monitor the boot integrity. Once installed, boot integrity will be attested via Remote Attestation. This assessment only applies to trusted launch enabled virtual machines.</p>	
	<p><b>Important:</b> Trusted launch requires the creation of new virtual machines. You can't enable trusted launch on existing virtual machines that were initially created without it. Learn more about <a href="#">Trusted launch for Azure virtual machines</a> (No related policy)</p>	
<a href="#">Guest Configuration extension should be installed on machines ↗</a>	<p>To ensure secure configurations of in-guest settings of your machine, install the Guest Configuration extension. In-guest settings that the extension monitors include the configuration of the operating system, application configuration or presence, and environment settings. Once installed, in-guest policies will be available such as 'Windows Exploit guard should be enabled'. <a href="#">Learn more ↗</a>. (Related policy: <a href="#">Virtual machines should have the Guest Configuration extension ↗</a>)</p>	Medium
<a href="#">Install endpoint protection solution on virtual machines ↗</a>	<p>Install an endpoint protection solution on your virtual machines, to protect them from threats and vulnerabilities. (Related policy: <a href="#">Monitor missing Endpoint Protection in Azure Security Center ↗</a>)</p>	High
<a href="#">Linux virtual machines should enforce kernel module signature validation ↗</a>	<p>To help mitigate against the execution of malicious or unauthorized code in kernel mode, enforce kernel module signature validation on supported Linux virtual machines. Kernel module signature validation ensures that only trusted kernel modules will be allowed to run. This assessment only applies to Linux virtual machines that have the Azure Monitor Agent installed. (No related policy)</p>	Low
<a href="#">Linux virtual machines should use only signed and trusted boot components ↗</a>	<p>With Secure Boot enabled, all OS boot components (boot loader, kernel, kernel drivers) must be signed by trusted publishers. Defender for Cloud has identified untrusted OS boot components on one or more of your Linux machines. To protect your machines from potentially malicious components, add them to your allowlist or remove the identified components. (No related policy)</p>	Low
<a href="#">Linux virtual machines should use Secure Boot ↗</a>	<p>To protect against the installation of malware-based rootkits and boot kits, enable Secure Boot on supported</p>	Low

Recommendation	Description	Severity
	<p>Secure Boot ensures that only signed operating systems and drivers will be allowed to run. This assessment only applies to Linux virtual machines that have the Azure Monitor Agent installed.</p> <p>(No related policy)</p>	
<a href="#">Log Analytics agent should be installed on Linux-based Azure Arc-enabled machines ↗</a>	<p>Defender for Cloud uses the Log Analytics agent (also known as OMS) to collect security events from your Azure Arc machines. To deploy the agent on all your Azure Arc machines, follow the remediation steps.</p> <p>(No related policy)</p>	High
<a href="#">Log Analytics agent should be installed on virtual machine scale sets ↗</a>	<p>Defender for Cloud collects data from your Azure virtual machines (VMs) to monitor for security vulnerabilities and threats. Data is collected using the <a href="#">Log Analytics agent</a>, formerly known as the Microsoft Monitoring Agent (MMA), which reads various security-related configurations and event logs from the machine and copies the data to your workspace for analysis. You'll also need to follow that procedure if your VMs are used by an Azure managed service such as Azure Kubernetes Service or Azure Service Fabric. You cannot configure auto-provisioning of the agent for Azure virtual machine scale sets. To deploy the agent on virtual machine scale sets (including those used by Azure managed services such as Azure Kubernetes Service and Azure Service Fabric), follow the procedure in the remediation steps.</p> <p>(Related policy: <a href="#">Log Analytics agent should be installed on your virtual machine scale sets for Azure Security Center monitoring ↗</a>)</p>	High
<a href="#">Log Analytics agent should be installed on virtual machines ↗</a>	<p>Defender for Cloud collects data from your Azure virtual machines (VMs) to monitor for security vulnerabilities and threats. Data is collected using the <a href="#">Log Analytics agent</a>, formerly known as the Microsoft Monitoring Agent (MMA), which reads various security-related configurations and event logs from the machine and copies the data to your Log Analytics workspace for analysis. This agent is also required if your VMs are used by an Azure managed service such as Azure Kubernetes Service or Azure Service Fabric. We recommend configuring <a href="#">auto-provisioning</a> to automatically deploy the agent. If you choose not to use auto-provisioning, manually deploy the agent to your VMs using the instructions in the remediation steps.</p> <p>(Related policy: <a href="#">Log Analytics agent should be installed on your virtual machine for Azure Security Center monitoring ↗</a>)</p>	High
<a href="#">Log Analytics agent should be installed on</a>	<p>Defender for Cloud uses the Log Analytics agent (also known as MMA) to collect security events from your Azure</p>	High

Recommendation	Description	Severity
<a href="#">Windows-based Azure Arc-enabled machines</a> ↗	Arc machines. To deploy the agent on all your Azure Arc machines, follow the remediation steps. (No related policy)	
<a href="#">Machines should be configured securely</a> ↗	Remediate vulnerabilities in security configuration on your machines to protect them from attacks. (Related policy: <a href="#">Vulnerabilities in security configuration on your machines should be remediated</a> ↗)	Low
<a href="#">Machines should be restarted to apply security configuration updates</a> ↗	To apply security configuration updates and protect against vulnerabilities, restart your machines. This assessment only applies to Linux virtual machines that have the Azure Monitor Agent installed. (No related policy)	Low
<a href="#">Machines should have a vulnerability assessment solution</a> ↗	Defender for Cloud regularly checks your connected machines to ensure they're running vulnerability assessment tools. Use this recommendation to deploy a vulnerability assessment solution. (Related policy: <a href="#">A vulnerability assessment solution should be enabled on your virtual machines</a> ↗)	Medium
<a href="#">Machines should have vulnerability findings resolved</a> ↗	Resolve the findings from the vulnerability assessment solutions on your virtual machines. (Related policy: <a href="#">A vulnerability assessment solution should be enabled on your virtual machines</a> ↗)	Low
<a href="#">Management ports of virtual machines should be protected with just-in-time network access control</a> ↗	Defender for Cloud has identified some overly-permissive inbound rules for management ports in your Network Security Group. Enable just-in-time access control to protect your VM from internet-based brute-force attacks. Learn more in <a href="#">Understanding just-in-time (JIT) VM access</a> . (Related policy: <a href="#">Management ports of virtual machines should be protected with just-in-time network access control</a> ↗)	High
<a href="#">Microsoft Defender for servers should be enabled</a> ↗	Microsoft Defender for servers provides real-time threat protection for your server workloads and generates hardening recommendations as well as alerts about suspicious activities. You can use this information to quickly remediate security issues and improve the security of your servers.	High
	Important: Remediating this recommendation will result in charges for protecting your servers. If you don't have any servers in this subscription, no charges will be incurred. If you create any servers on this subscription in the future, they will automatically be protected and charges will begin at that time.	

Recommendation	Description	Severity
	<p>Learn more in <a href="#">Introduction to Microsoft Defender for servers</a>.</p> <p>(Related policy: <a href="#">Azure Defender for servers should be enabled ↗</a>)</p>	
<a href="#">Microsoft Defender for servers should be enabled on workspaces ↗</a>	<p>Microsoft Defender for servers brings threat detection and advanced defenses for your Windows and Linux machines.</p> <p>With this Defender plan enabled on your subscriptions but not on your workspaces, you're paying for the full capability of Microsoft Defender for servers but missing out on some of the benefits.</p> <p>When you enable Microsoft Defender for servers on a workspace, all machines reporting to that workspace will be billed for Microsoft Defender for servers - even if they're in subscriptions without Defender plans enabled. Unless you also enable Microsoft Defender for servers on the subscription, those machines won't be able to take advantage of just-in-time VM access, adaptive application controls, and network detections for Azure resources.</p> <p>Learn more in <a href="#">Introduction to Microsoft Defender for servers</a>.</p> <p>(No related policy)</p>	Medium
<a href="#">Secure Boot should be enabled on supported Windows virtual machines ↗</a>	<p>Enable Secure Boot on supported Windows virtual machines to mitigate against malicious and unauthorized changes to the boot chain. Once enabled, only trusted bootloaders, kernel and kernel drivers will be allowed to run. This assessment only applies to trusted launch enabled Windows virtual machines.</p> <p><b>Important:</b> Trusted launch requires the creation of new virtual machines. You can't enable trusted launch on existing virtual machines that were initially created without it.</p> <p>Learn more about <a href="#">Trusted launch for Azure virtual machines..</a></p> <p>(No related policy)</p>	Low
<a href="#">Service Fabric clusters should have the ClusterProtectionLevel property set to EncryptAndSign ↗</a>	<p>Service Fabric provides three levels of protection (None, Sign and EncryptAndSign) for node-to-node communication using a primary cluster certificate. Set the protection level to ensure that all node-to-node messages are encrypted and digitally signed.</p> <p>(Related policy: <a href="#">Service Fabric clusters should have the ClusterProtectionLevel property set to EncryptAndSign ↗</a>)</p>	High

Recommendation	Description	Severity
<a href="#">Service Fabric clusters should only use Azure Active Directory for client authentication ↗</a>	<p>Perform Client authentication only via Azure Active Directory in Service Fabric</p> <p>(Related policy: <a href="#">Service Fabric clusters should only use Azure Active Directory for client authentication ↗</a>)</p>	High
<a href="#">System updates on virtual machine scale sets should be installed ↗</a>	<p>Install missing system security and critical updates to secure your Windows and Linux virtual machine scale sets.</p> <p>(Related policy: <a href="#">System updates on virtual machine scale sets should be installed ↗</a>)</p>	High
<a href="#">System updates should be installed on your machines ↗</a>	<p>Install missing system security and critical updates to secure your Windows and Linux virtual machines and computers</p> <p>(Related policy: <a href="#">System updates should be installed on your machines ↗</a>)</p>	High
<a href="#">System updates should be installed on your machines (powered by Update Center) ↗</a>	<p>Your machines are missing system, security, and critical updates. Software updates often include critical patches to security holes. Such holes are frequently exploited in malware attacks so it's vital to keep your software updated.</p> <p>To install all outstanding patches and secure your machines, follow the remediation steps.</p> <p>(No related policy)</p>	High
<a href="#">Virtual machine scale sets should be configured securely ↗</a>	<p>Remediate vulnerabilities in security configuration on your virtual machine scale sets to protect them from attacks.</p> <p>(Related policy: <a href="#">Vulnerabilities in security configuration on your virtual machine scale sets should be remediated ↗</a>)</p>	High
<a href="#">Virtual machines guest attestation status should be healthy ↗</a>	<p>Guest attestation is performed by sending a trusted log (TCGLog) to an attestation server. The server uses these logs to determine whether boot components are trustworthy. This assessment is intended to detect compromises of the boot chain which might be the result of a bootkit or rootkit infection.</p> <p>This assessment only applies to Trusted Launch enabled virtual machines that have the Guest Attestation extension installed.</p> <p>(No related policy)</p>	Medium
<a href="#">Virtual machines' Guest Configuration extension should be deployed with system-assigned managed identity ↗</a>	<p>The Guest Configuration extension requires a system assigned managed identity. Azure virtual machines in the scope of this policy will be non-compliant when they have the Guest Configuration extension installed but do not have a system assigned managed identity. <a href="#">Learn more ↗</a></p> <p>(Related policy: <a href="#">Guest Configuration extension should be deployed to Azure virtual machines with system assigned managed identity ↗</a>)</p>	Medium

Recommendation	Description	Severity
<a href="#">Virtual machines should be migrated to new Azure Resource Manager resources ↗</a>	<p>Virtual Machines (classic) was deprecated and these VMs should be migrated to Azure Resource Manager. Because Azure Resource Manager now has full IaaS capabilities and other advancements, we deprecated the management of IaaS virtual machines (VMs) through Azure Service Manager (ASM) on February 28, 2020. This functionality will be fully retired on March 1, 2023.</p> <p>To view all affected classic VMs make sure to select all your Azure subscriptions under 'directories + subscriptions' tab.</p> <p>Available resources and information about this tool &amp; migration:</p> <ul style="list-style-type: none"> <li><a href="#">Overview of Virtual machines (classic) deprecation, step by step process for migration &amp; available Microsoft resources.</a></li> <li><a href="#">Details about Migrate to Azure Resource Manager migration tool.</a></li> <li><a href="#">Migrate to Azure Resource Manager migration tool using PowerShell.</a></li> </ul> <p>(Related policy: <a href="#">Virtual machines should be migrated to new Azure Resource Manager resources ↗</a>)</p>	High
<a href="#">Virtual machines should encrypt temp disks, caches, and data flows between Compute and Storage resources ↗</a>	<p>By default, a virtual machine's OS and data disks are encrypted-at-rest using platform-managed keys; temp disks and data caches aren't encrypted, and data isn't encrypted when flowing between compute and storage resources.</p> <p>For a comparison of different disk encryption technologies in Azure, see <a href="https://aka.ms/diskencryptioncomparison">https://aka.ms/diskencryptioncomparison ↗</a>.</p> <p>Use Azure Disk Encryption to encrypt all this data.</p> <p>Disregard this recommendation if:</p> <ol style="list-style-type: none"> <li>1. You're using the encryption-at-host feature, or 2. Server-side encryption on Managed Disks meets your security requirements.</li> </ol> <p>Learn more in <a href="#">Server-side encryption of Azure Disk Storage ↗</a>.</p> <p>(Related policy: <a href="#">Disk encryption should be applied on virtual machines ↗</a>)</p>	High
<a href="#">vTPM should be enabled on supported virtual machines ↗</a>	<p>Enable virtual TPM device on supported virtual machines to facilitate Measured Boot and other OS security features that require a TPM. Once enabled, vTPM can be used to attest boot integrity. This assessment only applies to trusted launch enabled virtual machines.</p> <p><b>Important:</b> Trusted launch requires the creation of new virtual machines.</p>	Low

Recommendation	Description	Severity
	<p>You can't enable trusted launch on existing virtual machines that were initially created without it.</p> <p>Learn more about <a href="#">Trusted launch for Azure virtual machines</a>.</p> <p>(No related policy)</p>	
<a href="#">Vulnerabilities in security configuration on your Linux machines should be remediated (powered by Guest Configuration)</a>	<p>Remediate vulnerabilities in security configuration on your Linux machines to protect them from attacks.</p> <p>(Related policy: <a href="#">Linux machines should meet requirements for the Azure security baseline</a>)</p>	Low
<a href="#">Vulnerabilities in security configuration on your Windows machines should be remediated (powered by Guest Configuration)</a>	<p>Remediate vulnerabilities in security configuration on your Windows machines to protect them from attacks.</p> <p>(No related policy)</p>	Low
<a href="#">Windows Defender Exploit Guard should be enabled on machines</a>	<p>Windows Defender Exploit Guard uses the Azure Policy Guest Configuration agent. Exploit Guard has four components that are designed to lock down devices against a wide variety of attack vectors and block behaviors commonly used in malware attacks while enabling enterprises to balance their security risk and productivity requirements (Windows only).</p> <p>(Related policy: <a href="#">Audit Windows machines on which Windows Defender Exploit Guard is not enabled</a>)</p>	Medium
<a href="#">Windows web servers should be configured to use secure communication protocols</a>	<p>To protect the privacy of information communicated over the Internet, your web servers should use the latest version of the industry-standard cryptographic protocol, Transport Layer Security (TLS). TLS secures communications over a network by using security certificates to encrypt a connection between machines.</p> <p>(Related policy: <a href="#">Audit Windows web servers that are not using secure communication protocols</a>)</p>	High
<a href="#">[Preview]: Linux virtual machines should enable Azure Disk Encryption or EncryptionAtHost</a>	<p>By default, a virtual machine's OS and data disks are encrypted-at-rest using platform-managed keys; temp disks and data caches aren't encrypted, and data isn't encrypted when flowing between compute and storage resources. Use Azure Disk Encryption or EncryptionAtHost to encrypt all this data. Visit <a href="https://aka.ms/diskencryptioncomparison">https://aka.ms/diskencryptioncomparison</a> to compare encryption offerings. This policy requires two prerequisites to be deployed to the policy assignment scope. For details, visit <a href="https://aka.ms/gcpol">https://aka.ms/gcpol</a>.</p>	High

Recommendation	Description	Severity
	(Related policy: [Preview]: Linux virtual machines should enable Azure Disk Encryption or EncryptionAtHost <a href="#">↗</a> )	
<a href="#">[Preview]: Windows virtual machines should enable Azure Disk Encryption or EncryptionAtHost <a href="#">↗</a></a>	<p>By default, a virtual machine's OS and data disks are encrypted-at-rest using platform-managed keys; temp disks and data caches aren't encrypted, and data isn't encrypted when flowing between compute and storage resources. Use Azure Disk Encryption or EncryptionAtHost to encrypt all this data. Visit <a href="https://aka.ms/diskencryptioncomparison">https://aka.ms/diskencryptioncomparison</a> <a href="#">↗</a> to compare encryption offerings. This policy requires two prerequisites to be deployed to the policy assignment scope. For details, visit <a href="https://aka.ms/gcpol">https://aka.ms/gcpol</a> <a href="#">↗</a>.</p> <p>(Related policy: [Preview]: Windows virtual machines should enable Azure Disk Encryption or EncryptionAtHost <a href="#">↗</a>)</p>	High
<a href="#">Virtual machines and virtual machine scale sets should have encryption at host enabled <a href="#">↗</a></a>	<p>Use encryption at host to get end-to-end encryption for your virtual machine and virtual machine scale set data. Encryption at host enables encryption at rest for your temporary disk and OS/data disk caches. Temporary and ephemeral OS disks are encrypted with platform-managed keys when encryption at host is enabled. OS/data disk caches are encrypted at rest with either customer-managed or platform-managed key, depending on the encryption type selected on the disk. Learn more at <a href="https://aka.ms/vm-hbe">https://aka.ms/vm-hbe</a> <a href="#">↗</a>. (Related policy: Virtual machines and virtual machine scale sets should have encryption at host enabled <a href="#">↗</a>)</p>	Medium

## Container recommendations

There are 32 recommendations in this category.

[\[+\] Expand table](#)

Recommendation	Description	Severity	Type
<a href="#">[Enable if required] Container registries should be encrypted with a customer-managed key (CMK) <a href="#">↗</a></a>	<p>Recommendations to use customer-managed keys for encryption of data at rest are not assessed by default, but are available to enable for applicable scenarios. Data is encrypted automatically using platform-managed keys, so the use of customer-managed keys should only be applied when obligated by compliance or restrictive policy requirements.</p> <p>To enable this recommendation, navigate to</p>	Low	Control plane

Recommendation	Description	Severity	Type
	<p>your Security Policy for the applicable scope, and update the <i>Effect</i> parameter for the corresponding policy to audit or enforce the use of customer-managed keys. Learn more in <a href="#">Manage security policies</a>.</p> <p>Use customer-managed keys to manage the encryption at rest of the contents of your registries. By default, the data is encrypted at rest with service-managed keys, but customer-managed keys (CMK) are commonly required to meet regulatory compliance standards.</p> <p>CMKs enable the data to be encrypted with an Azure Key Vault key created and owned by you. You have full control and responsibility for the key lifecycle, including rotation and management. Learn more about CMK encryption at <a href="https://aka.ms/acr/CMK">https://aka.ms/acr/CMK</a>.</p> <p>(Related policy: <a href="#">Container registries should be encrypted with a customer-managed key (CMK)</a>)</p>		
<a href="#">Azure Arc-enabled Kubernetes clusters should have the Azure Policy extension installed</a>	<p>Azure Policy extension for Kubernetes extends <a href="#">Gatekeeper</a> v3, an admission controller webhook for <a href="#">Open Policy Agent</a> (OPA), to apply at-scale enforcements and safeguards on your clusters in a centralized, consistent manner.</p> <p>(No related policy)</p>	High	Control plane
<a href="#">Azure Arc-enabled Kubernetes clusters should have the Defender extension installed</a>	<p>Defender's extension for Azure Arc provides threat protection for your Arc-enabled Kubernetes clusters. The extension collects data from all control plane (master) nodes in the cluster and sends it to the Microsoft Defender for Kubernetes backend in the cloud for further analysis. <a href="#">Learn more</a>.</p> <p>(No related policy)</p>	High	Control plane
<a href="#">Azure Kubernetes Service clusters should have Defender profile enabled</a>	<p>Microsoft Defender for Containers provides cloud-native Kubernetes security capabilities including environment hardening, workload protection, and run-time protection.</p> <p>When you enable the <code>SecurityProfile.AzureDefender</code> profile on your Azure Kubernetes Service cluster, an agent is deployed to your cluster to collect security event data.</p> <p>Learn more in <a href="#">Introduction to Microsoft</a></p>	High	Control plane

Recommendation	Description	Severity	Type
	<p>Defender for Containers. (No related policy)</p>		
<a href="#">Azure Kubernetes Service clusters should have the Azure Policy add-on for Kubernetes installed ↗</a>	<p>Azure Policy add-on for Kubernetes extends <a href="#">Gatekeeper</a> v3, an admission controller webhook for <a href="#">Open Policy Agent</a> (OPA), to apply at-scale enforcements and safeguards on your clusters in a centralized, consistent manner.</p> <p>Defender for Cloud requires the Add-on to audit and enforce security capabilities and compliance inside your clusters. <a href="#">Learn more</a>.</p> <p>Requires Kubernetes v1.14.0 or later.</p> <p>(Related policy: <a href="#">Azure Policy Add-on for Kubernetes service (AKS) should be installed and enabled on your clusters ↗</a>)</p>	High	Control plane
<a href="#">Container registries should not allow unrestricted network access ↗</a>	<p>Azure container registries by default accept connections over the internet from hosts on any network. To protect your registries from potential threats, allow access from only specific public IP addresses or address ranges. If your registry doesn't have an IP/firewall rule or a configured virtual network, it will appear in the unhealthy resources. Learn more about Container Registry network rules here: <a href="https://aka.ms/acr/portal/public-network">https://aka.ms/acr/portal/public-network</a> ↗ and here <a href="https://aka.ms/acr/vnet">https://aka.ms/acr/vnet</a> ↗.</p> <p>(Related policy: <a href="#">Container registries should not allow unrestricted network access ↗</a>)</p>	Medium	Control plane
<a href="#">Container registries should use private link ↗</a>	<p>Azure Private Link lets you connect your virtual network to Azure services without a public IP address at the source or destination. The private link platform handles the connectivity between the consumer and services over the Azure backbone network. By mapping private endpoints to your container registries instead of the entire service, you'll also be protected against data leakage risks. Learn more at: <a href="https://aka.ms/acr/private-link">https://aka.ms/acr/private-link</a> ↗.</p> <p>(Related policy: <a href="#">Container registries should use private link ↗</a>)</p>	Medium	Control plane
<a href="#">Diagnostic logs in Kubernetes services</a>	<p>Enable diagnostic logs in your Kubernetes services and retain them up to a year. This</p>	Low	Control plane

Recommendation	Description	Severity	Type
<a href="#">Should be enabled ↴</a>	enables you to recreate activity trails for investigation purposes when a security incident occurs. (No related policy)		
<a href="#">Kubernetes API server should be configured with restricted access ↴</a>	To ensure that only applications from allowed networks, machines, or subnets can access your cluster, restrict access to your Kubernetes API server. You can restrict access by defining authorized IP ranges, or by setting up your API servers as private clusters as explained in <a href="#">Create a private Azure Kubernetes Service cluster</a> . (Related policy: <a href="#">Authorized IP ranges should be defined on Kubernetes Services ↴</a> )	High	Control plane
<a href="#">Role-Based Access Control should be used on Kubernetes Services ↴</a>	To provide granular filtering on the actions that users can perform, use Role-Based Access Control (RBAC) to manage permissions in Kubernetes Service Clusters and configure relevant authorization policies. For more information, see <a href="#">Azure role-based access control</a> . (Related policy: <a href="#">Role-Based Access Control (RBAC) should be used on Kubernetes Services ↴</a> )	High	Control plane
<a href="#">Microsoft Defender for Containers should be enabled ↴</a>	<p>Microsoft Defender for Containers provides hardening, vulnerability assessment and runtime protections for your Azure, hybrid, and multicloud Kubernetes environments.</p> <p>You can use this information to quickly remediate security issues and improve the security of your containers.</p> <p>Important: Remediating this recommendation will result in charges for protecting your Kubernetes clusters. If you don't have any Kubernetes clusters in this subscription, no charges will be incurred.</p> <p>If you create any Kubernetes clusters on this subscription in the future, they'll automatically be protected and charges will begin at that time.</p> <p>Learn more in <a href="#">Introduction to Microsoft Defender for Containers</a>.</p> <p>(No related policy)</p>	High	Control plane
<a href="#">Container CPU and memory limits should</a>	Enforcing CPU and memory limits prevents resource exhaustion attacks (a form of denial	Medium	Kubernetes Data plane

Recommendation	Description	Severity	Type
<a href="#">be enforced ↗</a>	<p>of service attack).</p> <p>We recommend setting limits for containers to ensure the runtime prevents the container from using more than the configured resource limit.</p> <p>(Related policy: <a href="#">Ensure container CPU and memory resource limits do not exceed the specified limits in Kubernetes cluster ↗</a>)</p>		
<a href="#">Container images should be deployed from trusted registries only ↗</a>	<p>Images running on your Kubernetes cluster should come from known and monitored container image registries. Trusted registries reduce your cluster's exposure risk by limiting the potential for the introduction of unknown vulnerabilities, security issues and malicious images.</p> <p>(Related policy: <a href="#">Ensure only allowed container images in Kubernetes cluster ↗</a>)</p>	High	Kubernetes Data plane
<a href="#">Container with privilege escalation should be avoided ↗</a>	<p>Containers shouldn't run with privilege escalation to root in your Kubernetes cluster. The AllowPrivilegeEscalation attribute controls whether a process can gain more privileges than its parent process.</p> <p>(Related policy: <a href="#">Kubernetes clusters should not allow container privilege escalation ↗</a>)</p>	Medium	Kubernetes data plane
<a href="#">Containers sharing sensitive host namespaces should be avoided ↗</a>	<p>To protect against privilege escalation outside the container, avoid pod access to sensitive host namespaces (host process ID and host IPC) in a Kubernetes cluster.</p> <p>(Related policy: <a href="#">Kubernetes cluster containers should not share host process ID or host IPC namespace ↗</a>)</p>	Medium	Kubernetes data plane
<a href="#">Containers should only use allowed AppArmor profiles ↗</a>	<p>Containers running on Kubernetes clusters should be limited to allowed AppArmor profiles only.</p> <p>;AppArmor (Application Armor) is a Linux security module that protects an operating system and its applications from security threats. To use it, a system administrator associates an AppArmor security profile with each program.</p> <p>(Related policy: <a href="#">Kubernetes cluster containers should only use allowed AppArmor profiles ↗</a>)</p>	High	Kubernetes data plane

Recommendation	Description	Severity	Type
<a href="#">Immutable (read-only) root filesystem should be enforced for containers ↗</a>	<p>Containers should run with a read only root file system in your Kubernetes cluster.</p> <p>Immutable filesystem protects containers from changes at run-time with malicious binaries being added to PATH.</p> <p>(Related policy: <a href="#">Kubernetes cluster containers should run with a read only root file system ↗</a>)</p>	Medium	Kubernetes data plane
<a href="#">Kubernetes clusters should be accessible only over HTTPS ↗</a>	<p>Use of HTTPS ensures authentication and protects data in transit from network layer eavesdropping attacks. This capability is currently generally available for Kubernetes Service (AKS), and in preview for AKS Engine and Azure Arc-enabled Kubernetes. For more info, visit <a href="https://aka.ms/kubepolicydoc">https://aka.ms/kubepolicydoc ↗</a></p> <p>(Related policy: <a href="#">Enforce HTTPS ingress in Kubernetes cluster ↗</a>)</p>	High	Kubernetes Data plane
<a href="#">Kubernetes clusters should disable automounting API credentials ↗</a>	<p>Disable automounting API credentials to prevent a potentially compromised Pod resource to run API commands against Kubernetes clusters. For more information, see <a href="https://aka.ms/kubepolicydoc">https://aka.ms/kubepolicydoc ↗</a>.</p> <p>(Related policy: <a href="#">Kubernetes clusters should disable automounting API credentials ↗</a>)</p>	High	Kubernetes Data plane
<a href="#">Kubernetes clusters should not grant CAPSYSADMIN security capabilities ↗</a>	<p>To reduce the attack surface of your containers, restrict CAP_SYS_ADMIN Linux capabilities. For more information, see <a href="https://aka.ms/kubepolicydoc">https://aka.ms/kubepolicydoc ↗</a>.</p> <p>(No related policy)</p>	High	Kubernetes data plane
<a href="#">Kubernetes clusters should not use the default namespace ↗</a>	<p>Prevent usage of the default namespace in Kubernetes clusters to protect against unauthorized access for ConfigMap, Pod, Secret, Service, and ServiceAccount resource types. For more information, see <a href="https://aka.ms/kubepolicydoc">https://aka.ms/kubepolicydoc ↗</a>.</p> <p>(Related policy: <a href="#">Kubernetes clusters should not use the default namespace ↗</a>)</p>	Low	Kubernetes data plane
<a href="#">Least privileged Linux capabilities should be enforced for containers ↗</a>	<p>To reduce attack surface of your container, restrict Linux capabilities and grant specific privileges to containers without granting all the privileges of the root user. We recommend dropping all capabilities, then adding those that are required</p>	Medium	Kubernetes data plane

Recommendation	Description	Severity	Type
	(Related policy: <a href="#">Kubernetes cluster containers should only use allowed capabilities ↗</a> )		
<a href="#">Privileged containers should be avoided ↗</a>	<p>To prevent unrestricted host access, avoid privileged containers whenever possible. Privileged containers have all of the root capabilities of a host machine. They can be used as entry points for attacks and to spread malicious code or malware to compromised applications, hosts and networks.</p>	Medium	Kubernetes data plane
	(Related policy: <a href="#">Do not allow privileged containers in Kubernetes cluster ↗</a> )		
<a href="#">Running containers as root user should be avoided ↗</a>	<p>Containers shouldn't run as root users in your Kubernetes cluster. Running a process as the root user inside a container runs it as root on the host. If there's a compromise, an attacker has root in the container, and any misconfigurations become easier to exploit.</p> <p>(Related policy: <a href="#">Kubernetes cluster pods and containers should only run with approved user and group IDs ↗</a>)</p>	High	Kubernetes Data plane
<a href="#">Services should listen on allowed ports only ↗</a>	<p>To reduce the attack surface of your Kubernetes cluster, restrict access to the cluster by limiting services access to the configured ports.</p> <p>(Related policy: <a href="#">Ensure services listen only on allowed ports in Kubernetes cluster ↗</a>)</p>	Medium	Kubernetes data plane
<a href="#">Usage of host networking and ports should be restricted ↗</a>	<p>Restrict pod access to the host network and the allowable host port range in a Kubernetes cluster. Pods created with the hostNetwork attribute enabled will share the node's network space. To avoid compromised container from sniffing network traffic, we recommend not putting your pods on the host network. If you need to expose a container port on the node's network, and using a Kubernetes Service node port does not meet your needs, another possibility is to specify a hostPort for the container in the pod spec.</p> <p>(Related policy: <a href="#">Kubernetes cluster pods should only use approved host network and port range ↗</a>)</p>	Medium	Kubernetes data plane
<a href="#">Usage of pod HostPath volume</a>	We recommend limiting pod HostPath volume mounts in your Kubernetes cluster to the	Medium	Kubernetes Data plane

Recommendation	Description	Severity	Type
<a href="#">Azure registry container images should have vulnerabilities resolved (powered by Qualys)</a>	<p>Defines allowed host paths. If there's a compromise, the container node access from the containers should be restricted.</p> <p>(Related policy: <a href="#">Kubernetes cluster pod hostPath volumes should only use allowed host paths</a>)</p>	High	Vulnerability Assessment
<a href="#">Azure registry container images should have vulnerabilities resolved (powered by Microsoft Defender Vulnerability Management)</a>	<p>Container image vulnerability assessment scans your registry for security vulnerabilities and exposes detailed findings for each image. Resolving the vulnerabilities can greatly improve your containers' security posture and protect them from attacks.</p> <p>(Related policy: <a href="#">Vulnerabilities in Azure Container Registry images should be remediated</a>)</p>	High	Vulnerability Assessment
<a href="#">Azure running container images should have vulnerabilities resolved - (powered by Qualys)</a>	<p>Container image vulnerability assessment scans container images running on your Kubernetes clusters for security vulnerabilities and exposes detailed findings for each image. Resolving the vulnerabilities can greatly improve your containers' security posture and protect them from attacks.</p> <p>(No related policy)</p>	High	Vulnerability Assessment
<a href="#">Azure running container images should have vulnerabilities resolved (powered by Microsoft Defender Vulnerability Management)</a>	<p>Container image vulnerability assessment scans your registry for commonly known vulnerabilities (CVEs) and provides a detailed vulnerability report for each image. This recommendation provides visibility to vulnerable images currently running in your Kubernetes clusters. Remediating vulnerabilities in container images that are currently running is key to improving your security posture, significantly reducing the attack surface for your containerized workloads.</p>	High	Vulnerability Assessment

Recommendation	Description	Severity	Type
<a href="#">AWS registry container images should have vulnerabilities resolved - (powered by Trivy) ↗</a>	Container image vulnerability assessment scans your registry for security vulnerabilities and exposes detailed findings for each image. Resolving the vulnerabilities can greatly improve your containers' security posture and protect them from attacks.	High	Vulnerability Assessment

## Data recommendations

There are **78** recommendations in this category.

[Expand table](#)

Recommendation	Description	Severity
<a href="#">[Enable if required] Azure Cosmos DB accounts should use customer-managed keys to encrypt data at rest ↗</a>	<p>Recommendations to use customer-managed keys for encryption of data at rest are not assessed by default, but are available to enable for applicable scenarios. Data is encrypted automatically using platform-managed keys, so the use of customer-managed keys should only be applied when obligated by compliance or restrictive policy requirements.</p> <p>To enable this recommendation, navigate to your Security Policy for the applicable scope, and update the <i>Effect</i> parameter for the corresponding policy to audit or enforce the use of customer-managed keys. Learn more in <a href="#">Manage security policies</a>.</p> <p>Use customer-managed keys to manage the encryption at rest of your Azure Cosmos DB. By default, the data is encrypted at rest with service-managed keys, but customer-managed keys (CMK) are commonly required to meet regulatory compliance standards. CMKs enable the data to be encrypted with an Azure Key Vault key created and owned by you. You have full control and responsibility for the key lifecycle, including rotation and management. Learn more about CMK encryption at <a href="https://aka.ms/cosmosdb-cmk">https://aka.ms/cosmosdb-cmk</a> ↗.</p> <p>(Related policy: <a href="#">Azure Cosmos DB accounts should use customer-managed keys to encrypt data at rest</a> ↗)</p>	Low
<a href="#">[Enable if required] Azure Machine Learning workspaces should be encrypted with a customer-managed key (CMK) ↗</a>	<p>Recommendations to use customer-managed keys for encryption of data at rest are not assessed by default, but are available to enable for applicable scenarios. Data is encrypted automatically using platform-managed keys, so the use of customer-managed keys should only be applied when obligated by compliance or restrictive policy requirements.</p> <p>To enable this recommendation, navigate to your Security</p>	Low

Recommendation	Description	Severity
	<p>Policy for the applicable scope, and update the <i>Effect</i> parameter for the corresponding policy to audit or enforce the use of customer-managed keys. Learn more in <a href="#">Manage security policies</a>.</p> <p>Manage encryption at rest of your Azure Machine Learning workspace data with customer-managed keys (CMK). By default, customer data is encrypted with service-managed keys, but CMKs are commonly required to meet regulatory compliance standards. CMKs enable the data to be encrypted with an Azure Key Vault key created and owned by you. You have full control and responsibility for the key lifecycle, including rotation and management. Learn more about CMK encryption at <a href="https://aka.ms/azureml-workspaces-cmk">https://aka.ms/azureml-workspaces-cmk</a>.</p> <p>(Related policy: <a href="#">Azure Machine Learning workspaces should be encrypted with a customer-managed key (CMK)</a>)</p>	
<a href="#">[Enable if required] Cognitive Services accounts should enable data encryption with a customer-managed key (CMK)</a>	<p>Recommendations to use customer-managed keys for encryption of data at rest are not assessed by default, but are available to enable for applicable scenarios. Data is encrypted automatically using platform-managed keys, so the use of customer-managed keys should only be applied when obligated by compliance or restrictive policy requirements.</p> <p>To enable this recommendation, navigate to your Security Policy for the applicable scope, and update the <i>Effect</i> parameter for the corresponding policy to audit or enforce the use of customer-managed keys. Learn more in <a href="#">Manage security policies</a>.</p> <p>Customer-managed keys (CMK) are commonly required to meet regulatory compliance standards. CMKs enable the data stored in Cognitive Services to be encrypted with an Azure Key Vault key created and owned by you. You have full control and responsibility for the key lifecycle, including rotation and management. Learn more about CMK encryption at <a href="https://aka.ms/cosmosdb-cmk">https://aka.ms/cosmosdb-cmk</a>.</p> <p>(Related policy: <a href="#">Cognitive Services accounts should enable data encryption with a customer-managed key?(CMK)</a>)</p>	Low
<a href="#">[Enable if required] MySQL servers should use customer-managed keys to encrypt data at rest</a>	<p>Recommendations to use customer-managed keys for encryption of data at rest are not assessed by default, but are available to enable for applicable scenarios. Data is encrypted automatically using platform-managed keys, so the use of customer-managed keys should only be applied when obligated by compliance or restrictive policy requirements.</p> <p>To enable this recommendation, navigate to your Security Policy for the applicable scope, and update the <i>Effect</i> parameter for the corresponding policy to audit or enforce the use of customer-managed keys. Learn more in <a href="#">Manage security policies</a>.</p>	Low

Recommendation	Description	Severity
	<p>Use customer-managed keys to manage the encryption at rest of your MySQL servers. By default, the data is encrypted at rest with service-managed keys, but customer-managed keys (CMK) are commonly required to meet regulatory compliance standards. CMKs enable the data to be encrypted with an Azure Key Vault key created and owned by you. You have full control and responsibility for the key lifecycle, including rotation and management.</p> <p>(Related policy: <a href="#">Bring your own key data protection should be enabled for MySQL servers ↗</a>)</p>	
<a href="#">[Enable if required] PostgreSQL servers should use customer-managed keys to encrypt data at rest ↗</a>	<p>Recommendations to use customer-managed keys for encryption of data at rest are not assessed by default, but are available to enable for applicable scenarios. Data is encrypted automatically using platform-managed keys, so the use of customer-managed keys should only be applied when obligated by compliance or restrictive policy requirements.</p> <p>To enable this recommendation, navigate to your Security Policy for the applicable scope, and update the <i>Effect</i> parameter for the corresponding policy to audit or enforce the use of customer-managed keys. Learn more in <a href="#">Manage security policies</a>.</p> <p>Use customer-managed keys to manage the encryption at rest of your PostgreSQL servers. By default, the data is encrypted at rest with service-managed keys, but customer-managed keys (CMK) are commonly required to meet regulatory compliance standards. CMKs enable the data to be encrypted with an Azure Key Vault key created and owned by you. You have full control and responsibility for the key lifecycle, including rotation and management.</p> <p>(Related policy: <a href="#">Bring your own key data protection should be enabled for PostgreSQL servers ↗</a>)</p>	Low
<a href="#">[Enable if required] SQL managed instances should use customer-managed keys to encrypt data at rest ↗</a>	<p>Recommendations to use customer-managed keys for encryption of data at rest are not assessed by default, but are available to enable for applicable scenarios. Data is encrypted automatically using platform-managed keys, so the use of customer-managed keys should only be applied when obligated by compliance or restrictive policy requirements.</p> <p>To enable this recommendation, navigate to your Security Policy for the applicable scope, and update the <i>Effect</i> parameter for the corresponding policy to audit or enforce the use of customer-managed keys. Learn more in <a href="#">Manage security policies</a>.</p> <p>Implementing Transparent Data Encryption (TDE) with your own key provides you with increased transparency and control over the TDE Protector, increased security with an HSM-backed external service, and promotion of separation of duties.</p>	Low

Recommendation	Description	Severity
	<p>This recommendation applies to organizations with a related compliance requirement.</p> <p>(Related policy: <a href="#">SQL managed instances should use customer-managed keys to encrypt data at rest</a>)</p>	
<a href="#">[Enable if required] SQL servers should use customer-managed keys to encrypt data at rest</a>	<p>Recommendations to use customer-managed keys for encryption of data at rest are not assessed by default, but are available to enable for applicable scenarios. Data is encrypted automatically using platform-managed keys, so the use of customer-managed keys should only be applied when obligated by compliance or restrictive policy requirements.</p> <p>To enable this recommendation, navigate to your Security Policy for the applicable scope, and update the <i>Effect</i> parameter for the corresponding policy to audit or enforce the use of customer-managed keys. Learn more in <a href="#">Manage security policies</a>.</p> <p>Implementing Transparent Data Encryption (TDE) with your own key provides increased transparency and control over the TDE Protector, increased security with an HSM-backed external service, and promotion of separation of duties. This recommendation applies to organizations with a related compliance requirement.</p> <p>(Related policy: <a href="#">SQL servers should use customer-managed keys to encrypt data at rest</a>)</p>	Low
<a href="#">[Enable if required] Storage accounts should use customer-managed key (CMK) for encryption</a>	<p>Recommendations to use customer-managed keys for encryption of data at rest are not assessed by default, but are available to enable for applicable scenarios. Data is encrypted automatically using platform-managed keys, so the use of customer-managed keys should only be applied when obligated by compliance or restrictive policy requirements.</p> <p>To enable this recommendation, navigate to your Security Policy for the applicable scope, and update the <i>Effect</i> parameter for the corresponding policy to audit or enforce the use of customer-managed keys. Learn more in <a href="#">Manage security policies</a>.</p> <p>Secure your storage account with greater flexibility using customer-managed keys (CMKs). When you specify a CMK, that key is used to protect and control access to the key that encrypts your data. Using CMKs provides additional capabilities to control rotation of the key encryption key or cryptographically erase data.</p> <p>(Related policy: <a href="#">Storage accounts should use customer-managed key (CMK) for encryption</a>)</p>	Low
<a href="#">All advanced threat protection types should be enabled in</a>	<p>It is recommended to enable all advanced threat protection types on your SQL managed instances. Enabling all types protects against SQL injection, database vulnerabilities, and</p>	Medium

Recommendation	Description	Severity
<a href="#">SQL managed instance advanced data security settings ↗</a>	any other anomalous activities. (No related policy)	
<a href="#">All advanced threat protection types should be enabled in SQL server advanced data security settings ↗</a>	It is recommended to enable all advanced threat protection types on your SQL servers. Enabling all types protects against SQL injection, database vulnerabilities, and any other anomalous activities. (No related policy)	Medium
<a href="#">API Management services should use a virtual network ↗</a>	Azure Virtual Network deployment provides enhanced security, isolation and allows you to place your API Management service in a non-internet routable network that you control access to. These networks can then be connected to your on-premises networks using various VPN technologies, which enables access to your backend services within the network and/or on-premises. The developer portal and API gateway, can be configured to be accessible either from the Internet or only within the virtual network. (Related policy: <a href="#">API Management services should use a virtual network ↗</a> )	Medium
<a href="#">App Configuration should use private link ↗</a>	Azure Private Link lets you connect your virtual network to Azure services without a public IP address at the source or destination. The private link platform handles the connectivity between the consumer and services over the Azure backbone network. By mapping private endpoints to your app configuration instances instead of the entire service, you'll also be protected against data leakage risks. Learn more at: <a href="https://aka.ms/appconfig/private-endpoint">https://aka.ms/appconfig/private-endpoint ↗</a> . (Related policy: <a href="#">App Configuration should use private link ↗</a> )	Medium
<a href="#">Audit retention for SQL servers should be set to at least 90 days ↗</a>	Audit SQL servers configured with an auditing retention period of less than 90 days. (Related policy: <a href="#">SQL servers should be configured with 90 days auditing retention or higher. ↗</a> )	Low
<a href="#">Auditing on SQL server should be enabled ↗</a>	Enable auditing on your SQL Server to track database activities across all databases on the server and save them in an audit log. (Related policy: <a href="#">Auditing on SQL server should be enabled ↗</a> )	Low
<a href="#">Auto provisioning of the Log Analytics agent should be enabled on subscriptions ↗</a>	To monitor for security vulnerabilities and threats, Microsoft Defender for Cloud collects data from your Azure virtual machines. Data is collected by the Log Analytics agent, formerly known as the Microsoft Monitoring Agent (MMA), which reads various security-related configurations and event logs from the machine and copies the data to your Log Analytics workspace for analysis. We recommend enabling	Low

Recommendation	Description	Severity
	<p>auto provisioning to automatically deploy the agent to all supported Azure VMs and any new ones that are created.</p> <p>(Related policy: <a href="#">Auto provisioning of the Log Analytics agent should be enabled on your subscription</a>)</p>	
<a href="#">Azure Cache for Redis should reside within a virtual network</a>	<p>Azure Virtual Network (VNet) deployment provides enhanced security and isolation for your Azure Cache for Redis, as well as subnets, access control policies, and other features to further restrict access. When an Azure Cache for Redis instance is configured with a VNet, it is not publicly addressable and can only be accessed from virtual machines and applications within the VNet.</p> <p>(Related policy: <a href="#">Azure Cache for Redis should reside within a virtual network</a>)</p>	Medium
<a href="#">Azure Database for MySQL should have an Azure Active Directory administrator provisioned</a>	<p>Provision an Azure AD administrator for your Azure Database for MySQL to enable Azure AD authentication. Azure AD authentication enables simplified permission management and centralized identity management of database users and other Microsoft services</p> <p>(Related policy: <a href="#">An Azure Active Directory administrator should be provisioned for MySQL servers</a>)</p>	Medium
<a href="#">Azure Database for PostgreSQL should have an Azure Active Directory administrator provisioned</a>	<p>Provision an Azure AD administrator for your Azure Database for PostgreSQL to enable Azure AD authentication. Azure AD authentication enables simplified permission management and centralized identity management of database users and other Microsoft services</p> <p>(Related policy: <a href="#">An Azure Active Directory administrator should be provisioned for PostgreSQL servers</a>)</p>	Medium
<a href="#">Azure Cosmos DB accounts should have firewall rules</a>	<p>Firewall rules should be defined on your Azure Cosmos DB accounts to prevent traffic from unauthorized sources.</p> <p>Accounts that have at least one IP rule defined with the virtual network filter enabled are deemed compliant. Accounts disabling public access are also deemed compliant.</p> <p>(Related policy: <a href="#">Azure Cosmos DB accounts should have firewall rules</a>)</p>	Medium
<a href="#">Azure Event Grid domains should use private link</a>	<p>Azure Private Link lets you connect your virtual network to Azure services without a public IP address at the source or destination. The private link platform handles the connectivity between the consumer and services over the Azure backbone network. By mapping private endpoints to your Event Grid domains instead of the entire service, you'll also be protected against data leakage risks. Learn more at: <a href="https://aka.ms/privateendpoints">https://aka.ms/privateendpoints</a>.</p>	Medium

Recommendation	Description	Severity
	(Related policy: <a href="#">Azure Event Grid domains should use private link</a> )	
<a href="#">Azure Event Grid topics should use private link</a>	Azure Private Link lets you connect your virtual network to Azure services without a public IP address at the source or destination. The private link platform handles the connectivity between the consumer and services over the Azure backbone network. By mapping private endpoints to your topics instead of the entire service, you'll also be protected against data leakage risks. Learn more at: <a href="https://aka.ms/privateendpoints">https://aka.ms/privateendpoints</a> . (Related policy: <a href="#">Azure Event Grid topics should use private link</a> )	Medium
<a href="#">Azure Machine Learning workspaces should use private link</a>	Azure Private Link lets you connect your virtual network to Azure services without a public IP address at the source or destination. The private link platform handles the connectivity between the consumer and services over the Azure backbone network. By mapping private endpoints to your Azure Machine Learning workspaces instead of the entire service, you'll also be protected against data leakage risks. Learn more at: <a href="https://aka.ms/azureml-workspaces-privatelink">https://aka.ms/azureml-workspaces-privatelink</a> . (Related policy: <a href="#">Azure Machine Learning workspaces should use private link</a> )	Medium
<a href="#">Azure SignalR Service should use private link</a>	Azure Private Link lets you connect your virtual network to Azure services without a public IP address at the source or destination. The private link platform handles the connectivity between the consumer and services over the Azure backbone network. By mapping private endpoints to your SignalR resources instead of the entire service, you'll also be protected against data leakage risks. Learn more at: <a href="https://aka.ms/asrs/privatelink">https://aka.ms/asrs/privatelink</a> . (Related policy: <a href="#">Azure SignalR Service should use private link</a> )	Medium
<a href="#">Azure Spring Cloud should use network injection</a>	Azure Spring Cloud instances should use virtual network injection for the following purposes: 1. Isolate Azure Spring Cloud from Internet. 2. Enable Azure Spring Cloud to interact with systems in either on premises data centers or Azure service in other virtual networks. 3. Empower customers to control inbound and outbound network communications for Azure Spring Cloud. (Related policy: <a href="#">Azure Spring Cloud should use network injection</a> )	Medium
<a href="#">Azure SQL Managed Instance authentication mode</a>	Disabling local authentication methods and allowing only Azure Active Directory Authentication improves security by ensuring that Azure SQL Managed Instances can exclusively be	Medium

Recommendation	Description	Severity
<a href="#">should be Azure Active Directory Only</a> <small>↗</small>	accessed by Azure Active Directory identities. (Related policy: <a href="#">Azure SQL Managed Instance should have Azure Active Directory Only Authentication enabled</a> <small>↗</small> )	
<a href="#">Azure Synapse Workspace authentication mode should be Azure Active Directory Only</a> <small>↗</small>	Azure Synapse Workspace authentication mode should be Azure Active Directory Only Azure Active Directory only authentication methods improves security by ensuring that Synapse Workspaces exclusively require Azure AD identities for authentication. <a href="#">Learn more</a> <small>↗</small> . (Related policy: <a href="#">Synapse Workspaces should use only Azure Active Directory identities for authentication</a> <small>↗</small> )	Medium
<a href="#">Code repositories should have code scanning findings resolved</a> <small>↗</small>	Defender for DevOps has found vulnerabilities in code repositories. To improve the security posture of the repositories, it is highly recommended to remediate these vulnerabilities. (No related policy)	Medium
<a href="#">Code repositories should have Dependabot scanning findings resolved</a> <small>↗</small>	Defender for DevOps has found vulnerabilities in code repositories. To improve the security posture of the repositories, it is highly recommended to remediate these vulnerabilities. (No related policy)	Medium
<a href="#">Code repositories should have infrastructure as code scanning findings resolved</a> <small>↗</small>	Defender for DevOps has found infrastructure as code security configuration issues in repositories. The issues shown below have been detected in template files. To improve the security posture of the related cloud resources, it is highly recommended to remediate these issues. (No related policy)	Medium
<a href="#">Code repositories should have secret scanning findings resolved</a> <small>↗</small>	Defender for DevOps has found a secret in code repositories. This should be remediated immediately to prevent a security breach. Secrets found in repositories can be leaked or discovered by adversaries, leading to compromise of an application or service. For Azure DevOps, the Microsoft Security DevOps CredScan tool only scans builds on which it has been configured to run. Therefore, results may not reflect the complete status of secrets in your repositories. (No related policy)	High
<a href="#">Cognitive Services accounts should enable data encryption</a> <small>↗</small>	This policy audits any Cognitive Services account not using data encryption. For each Cognitive Services account with storage, should enable data encryption with either customer managed or Microsoft managed key. (Related policy: <a href="#">Cognitive Services accounts should enable data encryption</a> <small>↗</small> )	Low

Recommendation	Description	Severity
<a href="#">Cognitive Services accounts should restrict network access ↗</a>	<p>Network access to Cognitive Services accounts should be restricted. Configure network rules so only applications from allowed networks can access the Cognitive Services account. To allow connections from specific internet or on-premises clients, access can be granted to traffic from specific Azure virtual networks or to public internet IP address ranges.</p> <p>(Related policy: <a href="#">Cognitive Services accounts should restrict network access ↗</a>)</p>	Medium
<a href="#">Cognitive Services accounts should use customer owned storage or enable data encryption ↗</a>	<p>This policy audits any Cognitive Services account not using customer owned storage nor data encryption. For each Cognitive Services account with storage, use either customer owned storage or enable data encryption.</p> <p>(Related policy: <a href="#">Cognitive Services accounts should use customer owned storage or enable data encryption. ↗</a>)</p>	Low
<a href="#">Diagnostic logs in Azure Data Lake Store should be enabled ↗</a>	<p>Enable logs and retain them for up to a year. This enables you to recreate activity trails for investigation purposes when a security incident occurs or your network is compromised.</p> <p>(Related policy: <a href="#">Diagnostic logs in Azure Data Lake Store should be enabled ↗</a>)</p>	Low
<a href="#">Diagnostic logs in Data Lake Analytics should be enabled ↗</a>	<p>Enable logs and retain them for up to a year. This enables you to recreate activity trails for investigation purposes when a security incident occurs or your network is compromised.</p> <p>(Related policy: <a href="#">Diagnostic logs in Data Lake Analytics should be enabled ↗</a>)</p>	Low
<a href="#">Email notification for high severity alerts should be enabled ↗</a>	<p>To ensure the relevant people in your organization are notified when there is a potential security breach in one of your subscriptions, enable email notifications for high severity alerts in Defender for Cloud.</p> <p>(Related policy: <a href="#">Email notification for high severity alerts should be enabled ↗</a>)</p>	Low
<a href="#">Email notification to subscription owner for high severity alerts should be enabled ↗</a>	<p>To ensure your subscription owners are notified when there is a potential security breach in their subscription, set email notifications to subscription owners for high severity alerts in Defender for Cloud.</p> <p>(Related policy: <a href="#">Email notification to subscription owner for high severity alerts should be enabled ↗</a>)</p>	Medium
<a href="#">Enforce SSL connection should be enabled for MySQL database servers ↗</a>	<p>Azure Database for MySQL supports connecting your Azure Database for MySQL server to client applications using Secure Sockets Layer (SSL).</p> <p>Enforcing SSL connections between your database server and your client applications helps protect against 'man in the middle' attacks by encrypting the data stream between the</p>	Medium

Recommendation	Description	Severity
	<p>server and your application.</p> <p>This configuration enforces that SSL is always enabled for accessing your database server.</p> <p>(Related policy: <a href="#">Enforce SSL connection should be enabled for MySQL database servers ↗</a>)</p>	
<a href="#">Enforce SSL connection should be enabled for PostgreSQL database servers ↗</a>	<p>Azure Database for PostgreSQL supports connecting your Azure Database for PostgreSQL server to client applications using Secure Sockets Layer (SSL).</p> <p>Enforcing SSL connections between your database server and your client applications helps protect against 'man in the middle' attacks by encrypting the data stream between the server and your application.</p> <p>This configuration enforces that SSL is always enabled for accessing your database server.</p> <p>(Related policy: <a href="#">Enforce SSL connection should be enabled for PostgreSQL database servers ↗</a>)</p>	Medium
<a href="#">Function apps should have vulnerability findings resolved ↗</a>	<p>Runtime vulnerability scanning for functions scans your function apps for security vulnerabilities and exposes detailed findings. Resolving the vulnerabilities can greatly improve your serverless applications security posture and protect them from attacks.</p> <p>(No related policy)</p>	High
<a href="#">Geo-redundant backup should be enabled for Azure Database for MariaDB ↗</a>	<p>Azure Database for MariaDB allows you to choose the redundancy option for your database server.</p> <p>It can be set to a geo-redundant backup storage in which the data is not only stored within the region in which your server is hosted, but is also replicated to a paired region to provide recovery options in case of a region failure.</p> <p>Configuring geo-redundant storage for backup is only allowed when creating a server.</p> <p>(Related policy: <a href="#">Geo-redundant backup should be enabled for Azure Database for MariaDB ↗</a>)</p>	Low
<a href="#">Geo-redundant backup should be enabled for Azure Database for MySQL ↗</a>	<p>Azure Database for MySQL allows you to choose the redundancy option for your database server.</p> <p>It can be set to a geo-redundant backup storage in which the data is not only stored within the region in which your server is hosted, but is also replicated to a paired region to provide recovery options in case of a region failure.</p> <p>Configuring geo-redundant storage for backup is only allowed when creating a server.</p> <p>(Related policy: <a href="#">Geo-redundant backup should be enabled for Azure Database for MySQL ↗</a>)</p>	Low
<a href="#">Geo-redundant backup should be enabled for Azure Database for PostgreSQL ↗</a>	<p>Azure Database for PostgreSQL allows you to choose the redundancy option for your database server.</p>	Low

Recommendation	Description	Severity
<a href="#">enabled for Azure Database for PostgreSQL ↗</a>	<p>It can be set to a geo-redundant backup storage in which the data is not only stored within the region in which your server is hosted, but is also replicated to a paired region to provide recovery options in case of a region failure.</p> <p>Configuring geo-redundant storage for backup is only allowed when creating a server.</p> <p>(Related policy: <a href="#">Geo-redundant backup should be enabled for Azure Database for PostgreSQL ↗</a>)</p>	
<a href="#">GitHub repositories should have Code scanning enabled ↗</a>	<p>GitHub uses code scanning to analyze code in order to find security vulnerabilities and errors in code. Code scanning can be used to find, triage, and prioritize fixes for existing problems in your code. Code scanning can also prevent developers from introducing new problems. Scans can be scheduled for specific days and times, or scans can be triggered when a specific event occurs in the repository, such as a push. If code scanning finds a potential vulnerability or error in code, GitHub displays an alert in the repository. A vulnerability is a problem in a project's code that could be exploited to damage the confidentiality, integrity, or availability of the project.</p> <p>(No related policy)</p>	Medium
<a href="#">GitHub repositories should have Dependabot scanning enabled ↗</a>	<p>GitHub sends Dependabot alerts when it detects vulnerabilities in code dependencies that affect repositories. A vulnerability is a problem in a project's code that could be exploited to damage the confidentiality, integrity, or availability of the project or other projects that use its code. Vulnerabilities vary in type, severity, and method of attack. When code depends on a package that has a security vulnerability, this vulnerable dependency can cause a range of problems.</p> <p>(No related policy)</p>	Medium
<a href="#">GitHub repositories should have Secret scanning enabled ↗</a>	<p>GitHub scans repositories for known types of secrets, to prevent fraudulent use of secrets that were accidentally committed to repositories. Secret scanning will scan the entire Git history on all branches present in the GitHub repository for any secrets. Examples of secrets are tokens and private keys that a service provider can issue for authentication. If a secret is checked into a repository, anyone who has read access to the repository can use the secret to access the external service with those privileges. Secrets should be stored in a dedicated, secure location outside the repository for the project.</p> <p>(No related policy)</p>	High
<a href="#">Microsoft Defender for Azure SQL</a>	<p>Microsoft Defender for SQL is a unified package that provides advanced SQL security capabilities.</p> <p>It includes functionality for surfacing and mitigating potential</p>	High

Recommendation	Description	Severity
<a href="#">Database servers should be enabled ↗</a>	<p>database vulnerabilities, detecting anomalous activities that could indicate a threat to your database, and discovering and classifying sensitive data.</p> <p>Important: Protections from this plan are charged as shown on the <a href="#">Defender plans</a> page. If you don't have any Azure SQL Database servers in this subscription, you won't be charged. If you later create Azure SQL Database servers on this subscription, they'll automatically be protected and charges will begin. Learn about the <a href="#">pricing details per region</a> ↗.</p> <p>Learn more in <a href="#">Introduction to Microsoft Defender for SQL</a>.            (Related policy: <a href="#">Azure Defender for Azure SQL Database servers should be enabled</a> ↗)</p>	
<a href="#">Microsoft Defender for DNS should be enabled ↗</a>	<p>Microsoft Defender for DNS provides an additional layer of protection for your cloud resources by continuously monitoring all DNS queries from your Azure resources.</p> <p>Defender for DNS alerts you about suspicious activity at the DNS layer. Learn more in <a href="#">Introduction to Microsoft Defender for DNS</a>. Enabling this Defender plan results in charges. Learn about the pricing details per region on Defender for Cloud's pricing page: <a href="https://azure.microsoft.com/services/defender-for-cloud/#pricing">https://azure.microsoft.com/services/defender-for-cloud/#pricing</a> ↗.</p> <p>(No related policy)</p>	High
<a href="#">Microsoft Defender for open-source relational databases should be enabled ↗</a>	<p>Microsoft Defender for open-source relational databases detects anomalous activities indicating unusual and potentially harmful attempts to access or exploit databases. Learn more in <a href="#">Introduction to Microsoft Defender for open-source relational databases</a>.</p> <p>Important: Enabling this plan will result in charges for protecting your open-source relational databases. If you don't have any open-source relational databases in this subscription, no charges will be incurred. If you create any open-source relational databases on this subscription in the future, they will automatically be protected and charges will begin at that time.</p> <p>(No related policy)</p>	High
<a href="#">Microsoft Defender for Resource Manager should be enabled ↗</a>	<p>Microsoft Defender for Resource Manager automatically monitors the resource management operations in your organization. Defender for Cloud detects threats and alerts you about suspicious activity. Learn more in <a href="#">Introduction to Microsoft Defender for Resource Manager</a>. Enabling this Defender plan results in charges. Learn about the pricing details per region on Defender for Cloud's pricing page: <a href="https://azure.microsoft.com/services/defender-for-cloud/#pricing">https://azure.microsoft.com/services/defender-for-cloud/#pricing</a> ↗.</p> <p>(No related policy)</p>	High

Recommendation	Description	Severity
<a href="#">Microsoft Defender for SQL on machines should be enabled on workspaces ↗</a>	<p>Microsoft Defender for servers brings threat detection and advanced defenses for your Windows and Linux machines. With this Defender plan enabled on your subscriptions but not on your workspaces, you're paying for the full capability of Microsoft Defender for servers but missing out on some of the benefits.</p> <p>When you enable Microsoft Defender for servers on a workspace, all machines reporting to that workspace will be billed for Microsoft Defender for servers - even if they're in subscriptions without Defender plans enabled. Unless you also enable Microsoft Defender for servers on the subscription, those machines won't be able to take advantage of just-in-time VM access, adaptive application controls, and network detections for Azure resources.</p> <p>Learn more in <a href="#">Introduction to Microsoft Defender for servers</a>. (No related policy)</p>	Medium
<a href="#">Microsoft Defender for SQL servers on machines should be enabled ↗</a>	<p>Microsoft Defender for SQL is a unified package that provides advanced SQL security capabilities. It includes functionality for surfacing and mitigating potential database vulnerabilities, detecting anomalous activities that could indicate a threat to your database, and discovering and classifying sensitive data.</p> <p>Important: Remediating this recommendation will result in charges for protecting your SQL servers on machines. If you don't have any SQL servers on machines in this subscription, no charges will be incurred.</p> <p>If you create any SQL servers on machines on this subscription in the future, they will automatically be protected and charges will begin at that time.</p> <p><a href="#">Learn more about Microsoft Defender for SQL servers on machines</a>.</p> <p>(Related policy: <a href="#">Azure Defender for SQL servers on machines should be enabled ↗</a>)</p>	High
<a href="#">Microsoft Defender for SQL should be enabled for unprotected Azure SQL servers ↗</a>	<p>Microsoft Defender for SQL is a unified package that provides advanced SQL security capabilities. It surfaces and mitigates potential database vulnerabilities, and detects anomalous activities that could indicate a threat to your database.</p> <p>Microsoft Defender for SQL is billed as shown on <a href="#">pricing details per region</a> ↗.</p> <p>(Related policy: <a href="#">Advanced data security should be enabled on your SQL servers ↗</a>)</p>	High
<a href="#">Microsoft Defender for SQL should be enabled for</a>	<p>Microsoft Defender for SQL is a unified package that provides advanced SQL security capabilities. It surfaces and mitigates potential database vulnerabilities, and detects anomalous</p>	High

Recommendation	Description	Severity
<a href="#">unprotected SQL Managed Instances</a>	<p>activities that could indicate a threat to your database.</p> <p>Microsoft Defender for SQL is billed as shown on <a href="#">pricing details per region</a>.</p> <p>(Related policy: <a href="#">Advanced data security should be enabled on SQL Managed Instance</a>)</p>	
<a href="#">Microsoft Defender for Storage should be enabled</a>	<p>Microsoft Defender for storage detects unusual and potentially harmful attempts to access or exploit storage accounts.</p> <p>Important: Protections from this plan are charged as shown on the <a href="#">Defender plans</a> page. If you don't have any Azure Storage accounts in this subscription, you won't be charged. If you later create Azure Storage accounts on this subscription, they'll automatically be protected and charges will begin. Learn about the <a href="#">pricing details per region</a>.</p> <p>Learn more in <a href="#">Introduction to Microsoft Defender for Storage</a>.</p> <p>(Related policy: <a href="#">Azure Defender for Storage should be enabled</a>)</p>	High
<a href="#">Network Watcher should be enabled</a>	<p>Network Watcher is a regional service that enables you to monitor and diagnose conditions at a network scenario level in, to, and from Azure. Scenario level monitoring enables you to diagnose problems at an end-to-end network level view.</p> <p>Network diagnostic and visualization tools available with Network Watcher help you understand, diagnose, and gain insights to your network in Azure.</p> <p>(Related policy: <a href="#">Network Watcher should be enabled</a>)</p>	Low
<a href="#">Over-provisioned identities in subscriptions should be investigated to reduce the Permission Creep Index (PCI)</a>	<p>Over-provisioned identities in subscription should be investigated to reduce the Permission Creep Index (PCI) and to safeguard your infrastructure. Reduce the PCI by removing the unused high risk permission assignments. High PCI reflects risk associated with the identities with permissions that exceed their normal or required usage</p> <p>(No related policy)</p>	Medium
<a href="#">Private endpoint connections on Azure SQL Database should be enabled</a>	<p>Private endpoint connections enforce secure communication by enabling private connectivity to Azure SQL Database.</p> <p>(Related policy: <a href="#">Private endpoint connections on Azure SQL Database should be enabled</a>)</p>	Medium
<a href="#">Private endpoint should be enabled for MariaDB servers</a>	<p>Private endpoint connections enforce secure communication by enabling private connectivity to Azure Database for MariaDB.</p> <p>Configure a private endpoint connection to enable access to traffic coming only from known networks and prevent access from all other IP addresses, including within Azure.</p> <p>(Related policy: <a href="#">Private endpoint should be enabled for MariaDB servers</a>)</p>	Medium

Recommendation	Description	Severity
<a href="#">Private endpoint should be enabled for MySQL servers ↴</a>	<p>Private endpoint connections enforce secure communication by enabling private connectivity to Azure Database for MySQL.</p> <p>Configure a private endpoint connection to enable access to traffic coming only from known networks and prevent access from all other IP addresses, including within Azure.</p> <p>(Related policy: <a href="#">Private endpoint should be enabled for MySQL servers ↴</a>)</p>	Medium
<a href="#">Private endpoint should be enabled for PostgreSQL servers ↴</a>	<p>Private endpoint connections enforce secure communication by enabling private connectivity to Azure Database for PostgreSQL.</p> <p>Configure a private endpoint connection to enable access to traffic coming only from known networks and prevent access from all other IP addresses, including within Azure.</p> <p>(Related policy: <a href="#">Private endpoint should be enabled for PostgreSQL servers ↴</a>)</p>	Medium
<a href="#">Public network access on Azure SQL Database should be disabled ↴</a>	<p>Disabling the public network access property improves security by ensuring your Azure SQL Database can only be accessed from a private endpoint. This configuration denies all logins that match IP or virtual network based firewall rules.</p> <p>(Related policy: <a href="#">Public network access on Azure SQL Database should be disabled ↴</a>)</p>	Medium
<a href="#">Public network access should be disabled for Cognitive Services accounts ↴</a>	<p>This policy audits any Cognitive Services account in your environment with public network access enabled. Public network access should be disabled so that only connections from private endpoints are allowed.</p> <p>(Related policy: <a href="#">Public network access should be disabled for Cognitive Services accounts ↴</a>)</p>	Medium
<a href="#">Public network access should be disabled for MariaDB servers ↴</a>	<p>Disable the public network access property to improve security and ensure your Azure Database for MariaDB can only be accessed from a private endpoint. This configuration strictly disables access from any public address space outside of Azure IP range, and denies all logins that match IP or virtual network-based firewall rules.</p> <p>(Related policy: <a href="#">Public network access should be disabled for MariaDB servers ↴</a>)</p>	Medium
<a href="#">Public network access should be disabled for MySQL servers ↴</a>	<p>Disable the public network access property to improve security and ensure your Azure Database for MySQL can only be accessed from a private endpoint. This configuration strictly disables access from any public address space outside of Azure IP range, and denies all logins that match IP or virtual network-based firewall rules.</p> <p>(Related policy: <a href="#">Public network access should be disabled for MySQL servers ↴</a>)</p>	Medium

Recommendation	Description	Severity
<a href="#">Public network access should be disabled for PostgreSQL servers ↗</a>	<p>Disable the public network access property to improve security and ensure your Azure Database for PostgreSQL can only be accessed from a private endpoint. This configuration disables access from any public address space outside of Azure IP range, and denies all logins that match IP or virtual network-based firewall rules.</p> <p>(Related policy: <a href="#">Public network access should be disabled for PostgreSQL servers ↗</a>)</p>	Medium
<a href="#">Redis Cache should allow access only via SSL ↗</a>	<p>Enable only connections via SSL to Redis Cache. Use of secure connections ensures authentication between the server and the service and protects data in transit from network layer attacks such as man-in-the-middle, eavesdropping, and session-hijacking.</p> <p>(Related policy: <a href="#">Only secure connections to your Azure Cache for Redis should be enabled ↗</a>)</p>	High
<a href="#">SQL databases should have vulnerability findings resolved ↗</a>	<p>SQL Vulnerability assessment scans your database for security vulnerabilities, and exposes any deviations from best practices such as misconfigurations, excessive permissions, and unprotected sensitive data. Resolving the vulnerabilities found can greatly improve your database security posture. <a href="#">Learn more ↗</a></p> <p>(Related policy: <a href="#">Vulnerabilities on your SQL databases should be remediated ↗</a>)</p>	High
<a href="#">SQL managed instances should have vulnerability assessment configured ↗</a>	<p>Vulnerability assessment can discover, track, and help you remediate potential database vulnerabilities.</p> <p>(Related policy: <a href="#">Vulnerability assessment should be enabled on SQL Managed Instance ↗</a>)</p>	High
<a href="#">SQL servers on machines should have vulnerability findings resolved ↗</a>	<p>SQL Vulnerability assessment scans your database for security vulnerabilities, and exposes any deviations from best practices such as misconfigurations, excessive permissions, and unprotected sensitive data. Resolving the vulnerabilities found can greatly improve your database security posture. <a href="#">Learn more ↗</a></p> <p>(Related policy: <a href="#">Vulnerabilities on your SQL servers on machine should be remediated ↗</a>)</p>	High
<a href="#">SQL servers should have an Azure Active Directory administrator provisioned ↗</a>	<p>Provision an Azure AD administrator for your SQL server to enable Azure AD authentication. Azure AD authentication enables simplified permission management and centralized identity management of database users and other Microsoft services.</p> <p>(Related policy: <a href="#">An Azure Active Directory administrator should be provisioned for SQL servers ↗</a>)</p>	High

Recommendation	Description	Severity
<a href="#">SQL servers should have vulnerability assessment configured</a>	<p>Vulnerability assessment can discover, track, and help you remediate potential database vulnerabilities.</p> <p>(Related policy: <a href="#">Vulnerability assessment should be enabled on your SQL servers</a>)</p>	High
<a href="#">Storage account should use a private link connection</a>	<p>Private links enforce secure communication, by providing private connectivity to the storage account</p> <p>(Related policy: <a href="#">Storage account should use a private link connection</a>)</p>	Medium
<a href="#">Storage accounts should be migrated to new Azure Resource Manager resources</a>	<p>To benefit from new capabilities in Azure Resource Manager, you can migrate existing deployments from the Classic deployment model. Resource Manager enables security enhancements such as: stronger access control (RBAC), better auditing, ARM-based deployment and governance, access to managed identities, access to key vault for secrets, Azure AD-based authentication and support for tags and resource groups for easier security management. <a href="#">Learn more</a></p> <p>(Related policy: <a href="#">Storage accounts should be migrated to new Azure Resource Manager resources</a>)</p>	Low
<a href="#">Storage accounts should restrict network access using virtual network rules</a>	<p>Protect your storage accounts from potential threats using virtual network rules as a preferred method instead of IP-based filtering. Disabling IP-based filtering prevents public IPs from accessing your storage accounts.</p> <p>(Related policy: <a href="#">Storage accounts should restrict network access using virtual network rules</a>)</p>	Medium
<a href="#">Subscriptions should have a contact email address for security issues</a>	<p>To ensure the relevant people in your organization are notified when there is a potential security breach in one of your subscriptions, set a security contact to receive email notifications from Defender for Cloud.</p> <p>(Related policy: <a href="#">Subscriptions should have a contact email address for security issues</a>)</p>	Low
<a href="#">Transparent Data Encryption on SQL databases should be enabled</a>	<p>Enable transparent data encryption to protect data-at-rest and meet compliance requirements</p> <p>(Related policy: <a href="#">Transparent Data Encryption on SQL databases should be enabled</a>)</p>	Low
<a href="#">VM Image Builder templates should use private link</a>	<p>Audit VM Image Builder templates that do not have a virtual network configured. When a virtual network is not configured, a public IP is created and used instead, which may directly expose resources to the internet and increase the potential attack surface.</p> <p>(Related policy: <a href="#">VM Image Builder templates should use private link</a>)</p>	Medium

Recommendation	Description	Severity
<a href="#">Web Application Firewall (WAF) should be enabled for Application Gateway ↗</a>	<p>Deploy Azure Web Application Firewall (WAF) in front of public facing web applications for additional inspection of incoming traffic. Web Application Firewall (WAF) provides centralized protection of your web applications from common exploits and vulnerabilities such as SQL injections, Cross-Site Scripting, local and remote file executions. You can also restrict access to your web applications by countries/regions, IP address ranges, and other http(s) parameters via custom rules.</p> <p>(Related policy: <a href="#">Web Application Firewall (WAF) should be enabled for Application Gateway ↗</a>)</p>	Low
<a href="#">Web Application Firewall (WAF) should be enabled for Azure Front Door Service service ↗</a>	<p>Deploy Azure Web Application Firewall (WAF) in front of public facing web applications for additional inspection of incoming traffic. Web Application Firewall (WAF) provides centralized protection of your web applications from common exploits and vulnerabilities such as SQL injections, Cross-Site Scripting, local and remote file executions. You can also restrict access to your web applications by countries/regions, IP address ranges, and other http(s) parameters via custom rules.</p> <p>(Related policy: <a href="#">Web Application Firewall (WAF) should be enabled for Azure Front Door Service?service  ↗</a>)</p>	Low
<a href="#">Cognitive Services should use private link ↗</a>	<p>Azure Private Link lets you connect your virtual networks to Azure services without a public IP address at the source or destination. The Private Link platform handles the connectivity between the consumer and services over the Azure backbone network. By mapping private endpoints to Cognitive Services, you'll reduce the potential for data leakage. Learn more about <a href="#">private links ↗</a>.</p> <p>(Related policy: <a href="#">Cognitive Services should use private link ↗</a>)</p>	Medium
<a href="#">Azure Cosmos DB should disable public network access ↗</a>	<p>Disabling public network access improves security by ensuring that your Cosmos DB account isn't exposed on the public internet. Creating private endpoints can limit exposure of your Cosmos DB account. <a href="#">Learn more</a>. (Related policy: <a href="#">Azure Cosmos DB should disable public network access ↗</a>)</p>	Medium
<a href="#">Cosmos DB accounts should use private link ↗</a>	<p>Azure Private Link lets you connect your virtual network to Azure services without a public IP address at the source or destination. The Private Link platform handles the connectivity between the consumer and services over the Azure backbone network. By mapping private endpoints to your Cosmos DB account, data leakage risks are reduced. Learn more about <a href="#">private links</a>. (Related policy: <a href="#">Cosmos DB accounts should use private link ↗</a>)</p>	Medium
<a href="#">Azure SQL Database should be running</a>	<p>Setting TLS version to 1.2 or newer improves security by ensuring your Azure SQL Database can only be accessed from</p>	Medium

Recommendation	Description	Severity
<a href="#">TLS version 1.2 or newer ↗</a>	clients using TLS 1.2 or newer. Using versions of TLS less than 1.2 is not recommended since they have well documented security vulnerabilities. (Related policy: <a href="#">Azure SQL Database should be running TLS version 1.2 or newer ↗</a> )	
<a href="#">Azure SQL Managed Instances should disable public network access ↗</a>	Disabling public network access (public endpoint) on Azure SQL Managed Instances improves security by ensuring that they can only be accessed from inside their virtual networks or via Private Endpoints. Learn more about <a href="#">public network access ↗</a> . (Related policy: <a href="#">Azure SQL Managed Instances should disable public network access ↗</a> )	Medium
<a href="#">Storage accounts should prevent shared key access ↗</a>	Audit requirement of Azure Active Directory (Azure AD) to authorize requests for your storage account. By default, requests can be authorized with either Azure Active Directory credentials, or by using the account access key for Shared Key authorization. Of these two types of authorization, Azure AD provides superior security and ease of use over shared Key, and is recommended by Microsoft. (Related policy: <a href="#">policy ↗</a> )	Medium

## IdentityAndAccess recommendations

There are 32 recommendations in this category.

[Expand table](#)

Recommendation	Description	Severity
<a href="#">A maximum of 3 owners should be designated for subscriptions ↗</a>	To reduce the potential for breaches by compromised owner accounts, we recommend limiting the number of owner accounts to a maximum of 3 (Related policy: <a href="#">A maximum of 3 owners should be designated for your subscription ↗</a> )	High
<a href="#">Accounts with owner permissions on Azure resources should be MFA enabled ↗</a>	If you only use passwords to authenticate your users, you're leaving an attack vector open. Users often use weak passwords for multiple services. By enabling <a href="#">multifactor authentication</a> (MFA), you provide better security for your accounts, while still allowing your users to authenticate to almost any application with single sign-on (SSO). Multifactor authentication is a process by which users are prompted, during the sign-in process, for another form of identification. For example, a code may be sent to their cellphone, or they may be asked for a fingerprint scan. We recommend you to enable MFA for all accounts that have	High

Recommendation	Description	Severity
	<p><a href="#">owner permissions</a> on Azure resources, to prevent breach and attacks.</p> <p>More details and frequently asked questions are available here: <a href="#">Manage multifactor authentication (MFA) enforcement on your subscriptions</a></p> <p>(No related policy)</p>	
<a href="#">Accounts with read permissions on Azure resources should be MFA enabled</a>	<p>If you only use passwords to authenticate your users, you're leaving an attack vector open. Users often use weak passwords for multiple services. By enabling <a href="#">multifactor authentication</a> (MFA), you provide better security for your accounts, while still allowing your users to authenticate to almost any application with single sign-on (SSO). Multifactor authentication is a process by which users are prompted, during the sign-in process, for an additional form of identification. For example, a code may be sent to their cellphone, or they may be asked for a fingerprint scan. We recommend you to enable MFA for all accounts that have <a href="#">read permissions</a> on Azure resources, to prevent breach and attacks.</p> <p>More details and frequently asked questions are available here: <a href="#">Manage multifactor authentication (MFA) enforcement on your subscriptions</a></p> <p>(No related policy)</p>	High
<a href="#">Accounts with write permissions on Azure resources should be MFA enabled</a>	<p>If you only use passwords to authenticate your users, you are leaving an attack vector open. Users often use weak passwords for multiple services. By enabling <a href="#">multifactor authentication</a> (MFA), you provide better security for your accounts, while still allowing your users to authenticate to almost any application with single sign-on (SSO). Multifactor authentication is a process by which users are prompted, during the sign-in process, for an additional form of identification. For example, a code may be sent to their cellphone, or they may be asked for a fingerprint scan. We recommend you to enable MFA for all accounts that have <a href="#">write permissions</a> on Azure resources, to prevent breach and attacks.</p> <p>More details and frequently asked questions are available here: <a href="#">Manage multifactor authentication (MFA) enforcement on your subscriptions</a></p> <p>(No related policy)</p>	High
<a href="#">Azure Cosmos DB accounts should use Azure Active Directory as the only authentication method</a>	<p>The best way to authenticate to Azure services is by using Role-Based Access Control (RBAC). RBAC allows you to maintain the minimum privilege principle and supports the ability to revoke permissions as an effective method of response when compromised. You can configure your Azure</p>	Medium

Recommendation	Description	Severity
	Cosmos DB account to enforce RBAC as the only authentication method. When the enforcement is configured, all other methods of access will be denied (primary/secondary keys and access tokens).  (No related policy)	
<a href="#">Blocked accounts with owner permissions on Azure resources should be removed ↴</a>	Accounts that have been blocked from signing in on Active Directory, should be removed from your Azure resources. These accounts can be targets for attackers looking to find ways to access your data without being noticed.  (No related policy)	High
<a href="#">Blocked accounts with read and write permissions on Azure resources should be remove ↴</a>	Accounts that have been blocked from signing in on Active Directory, should be removed from your Azure resources. These accounts can be targets for attackers looking to find ways to access your data without being noticed.  (No related policy)	High
<a href="#">Deprecated accounts should be removed from subscriptions ↴</a>	User accounts that have been blocked from signing in, should be removed from your subscriptions. These accounts can be targets for attackers looking to find ways to access your data without being noticed.  (Related policy: <a href="#">Deprecated accounts should be removed from your subscription ↴</a> )	High
<a href="#">Deprecated accounts with owner permissions should be removed from subscriptions ↴</a>	User accounts that have been blocked from signing in, should be removed from your subscriptions. These accounts can be targets for attackers looking to find ways to access your data without being noticed.  (Related policy: <a href="#">Deprecated accounts with owner permissions should be removed from your subscription ↴</a> )	High
<a href="#">Diagnostic logs in Key Vault should be enabled ↴</a>	Enable logs and retain them for up to a year. This enables you to recreate activity trails for investigation purposes when a security incident occurs or your network is compromised.  (Related policy: <a href="#">Diagnostic logs in Key Vault should be enabled ↴</a> )	Low
<a href="#">External accounts with owner permissions should be removed from subscriptions ↴</a>	Accounts with owner permissions that have different domain names (external accounts), should be removed from your subscription. This prevents unmonitored access. These accounts can be targets for attackers looking to find ways to access your data without being noticed.  (Related policy: <a href="#">External accounts with owner permissions should be removed from your subscription ↴</a> )	High
<a href="#">External accounts with read permissions</a>	Accounts with read permissions that have different domain names (external accounts), should be removed from your	High

Recommendation	Description	Severity
<a href="#">should be removed from subscriptions ↗</a>	<p>subscription. This prevents unmonitored access. These accounts can be targets for attackers looking to find ways to access your data without being noticed.</p> <p>(Related policy: <a href="#">External accounts with read permissions should be removed from your subscription ↗</a>)</p>	
<a href="#">External accounts with write permissions should be removed from subscriptions ↗</a>	<p>Accounts with write permissions that have different domain names (external accounts), should be removed from your subscription. This prevents unmonitored access. These accounts can be targets for attackers looking to find ways to access your data without being noticed.</p> <p>(Related policy: <a href="#">External accounts with write permissions should be removed from your subscription ↗</a>)</p>	High
<a href="#">Firewall should be enabled on Key Vault ↗</a>	<p>Key vault's firewall prevents unauthorized traffic from reaching your key vault and provides an additional layer of protection for your secrets. Enable the firewall to make sure that only traffic from allowed networks can access your key vault.</p> <p>(Related policy: <a href="#">Firewall should be enabled on Key Vault ↗</a>)</p>	Medium
<a href="#">Guest accounts with owner permissions on Azure resources should be removed ↗</a>	<p>Accounts with owner permissions that have been provisioned outside of the Azure Active Directory tenant (different domain names), should be removed from your Azure resources. Guest accounts aren't managed to the same standards as enterprise tenant identities. These accounts can be targets for attackers looking to find ways to access your data without being noticed.</p> <p>(No related policy)</p>	High
<a href="#">Guest accounts with read permissions on Azure resources should be removed ↗</a>	<p>Accounts with read permissions that have been provisioned outside of the Azure Active Directory tenant (different domain names), should be removed from your Azure resources. Guest accounts aren't managed to the same standards as enterprise tenant identities. These accounts can be targets for attackers looking to find ways to access your data without being noticed.</p> <p>(No related policy)</p>	High
<a href="#">Guest accounts with write permissions on Azure resources should be removed ↗</a>	<p>Accounts with write permissions that have been provisioned outside of the Azure Active Directory tenant (different domain names), should be removed from your Azure resources. Guest accounts aren't managed to the same standards as enterprise tenant identities. These accounts can be targets for attackers looking to find ways to access your data without being noticed.</p> <p>(No related policy)</p>	High

Recommendation	Description	Severity
<a href="#">Key Vault keys should have an expiration date ↗</a>	<p>Cryptographic keys should have a defined expiration date and not be permanent. Keys that are valid forever provide a potential attacker with more time to compromise the key. It's a recommended security practice to set expiration dates on cryptographic keys.</p> <p>(Related policy: <a href="#">Key Vault keys should have an expiration date ↗</a>)</p>	High
<a href="#">Key Vault secrets should have an expiration date ↗</a>	<p>Secrets should have a defined expiration date and not be permanent. Secrets that are valid forever provide a potential attacker with more time to compromise them. It's a recommended security practice to set expiration dates on secrets.</p> <p>(Related policy: <a href="#">Key Vault secrets should have an expiration date ↗</a>)</p>	High
<a href="#">Key vaults should have purge protection enabled ↗</a>	<p>Malicious deletion of a key vault can lead to permanent data loss. A malicious insider in your organization can potentially delete and purge key vaults. Purge protection protects you from insider attacks by enforcing a mandatory retention period for soft deleted key vaults. No one inside your organization or Microsoft will be able to purge your key vaults during the soft delete retention period.</p> <p>(Related policy: <a href="#">Key vaults should have purge protection enabled ↗</a>)</p>	Medium
<a href="#">Key vaults should have soft delete enabled ↗</a>	<p>Deleting a key vault without soft delete enabled permanently deletes all secrets, keys, and certificates stored in the key vault. Accidental deletion of a key vault can lead to permanent data loss. Soft delete allows you to recover an accidentally deleted key vault for a configurable retention period.</p> <p>(Related policy: <a href="#">Key vaults should have soft delete enabled ↗</a>)</p>	High
<a href="#">MFA should be enabled on accounts with owner permissions on subscriptions ↗</a>	<p>Multifactor authentication (MFA) should be enabled for all subscription accounts with owner permissions to prevent a breach of accounts or resources.</p> <p>(Related policy: <a href="#">MFA should be enabled on accounts with owner permissions on your subscription ↗</a>)</p>	High
<a href="#">MFA should be enabled on accounts with read permissions on subscriptions ↗</a>	<p>Multifactor authentication (MFA) should be enabled for all subscription accounts with read privileges to prevent a breach of accounts or resources.</p> <p>(Related policy: <a href="#">MFA should be enabled on accounts with read permissions on your subscription ↗</a>)</p>	High

Recommendation	Description	Severity
<a href="#">MFA should be enabled on accounts with write permissions on subscriptions ↗</a>	<p>Multifactor authentication (MFA) should be enabled for all subscription accounts with write privileges to prevent a breach of accounts or resources.</p> <p>(Related policy: <a href="#">MFA should be enabled accounts with write permissions on your subscription ↗</a>)</p>	High
<a href="#">Microsoft Defender for Key Vault should be enabled ↗</a>	<p>Microsoft Defender for Cloud includes Microsoft Defender for Key Vault, providing an additional layer of security intelligence.</p> <p>Microsoft Defender for Key Vault detects unusual and potentially harmful attempts to access or exploit Key Vault accounts.</p> <p>Important: Protections from this plan are charged as shown on the <a href="#">Defender plans</a> page. If you don't have any key vaults in this subscription, you won't be charged. If you later create key vaults on this subscription, they'll automatically be protected and charges will begin. Learn about the <a href="#">pricing details per region ↗</a>.</p> <p>Learn more in <a href="#">Introduction to Microsoft Defender for Key Vault</a>.</p> <p>(Related policy: <a href="#">Azure Defender for Key Vault should be enabled ↗</a>)</p>	High
<a href="#">Private endpoint should be configured for Key Vault ↗</a>	<p>Private link provides a way to connect Key Vault to your Azure resources without sending traffic over the public internet. Private link provides defense in depth protection against data exfiltration.</p> <p>(Related policy: <a href="#">Private endpoint should be configured for Key Vault ↗</a>)</p>	Medium
<a href="#">Storage account public access should be disallowed ↗</a>	<p>Anonymous public read access to containers and blobs in Azure Storage is a convenient way to share data, but might present security risks. To prevent data breaches caused by undesired anonymous access, Microsoft recommends preventing public access to a storage account unless your scenario requires it.</p> <p>(Related policy: <a href="#">Storage account public access should be disallowed ↗</a>)</p>	Medium
<a href="#">There should be more than one owner assigned to subscriptions ↗</a>	<p>Designate more than one subscription owner in order to have administrator access redundancy.</p> <p>(Related policy: <a href="#">There should be more than one owner assigned to your subscription ↗</a>)</p>	High
<a href="#">Validity period of certificates stored in Azure Key Vault should not exceed 12 months ↗</a>	<p>Ensure your certificates do not have a validity period that exceeds 12 months.</p> <p>(Related policy: <a href="#">Certificates should have the specified maximum validity period ↗</a>)</p>	Medium

Recommendation	Description	Severity
<a href="#">Azure overprovisioned identities should have only the necessary permissions (Preview)</a>	Overprovisioned identities, or over permissioned identities, don't use many of their granted permissions. Regularly right-size permissions of these identities to reduce the risk of permissions misuse, either accidental or malicious. This action decreases the potential blast radius during a security incident.	Medium
<a href="#">Super identities in your Azure environment should be removed (Preview)</a>	Super Identity is any human or workload identity such as users, Service Principals and serverless functions that have admin permissions and can perform any action on any resource across the infrastructure. Super Identities are extremely high risk, as any malicious or accidental permissions misuse can result in catastrophic service disruption, service degradation, or data leakage. Super Identities pose a huge threat to cloud infrastructure. Too many super identities can create excessive risks and increase the blast radius during a breach.	Medium
<a href="#">Unused identities in your Azure environment should be removed (Preview)</a>	Inactive Identities are the identities that have not performed any action on any infrastructure resources in the last 90 days. Inactive identities pose a significant risk to your organization as they could be used by attackers to gain access and execute tasks in your environment.	Medium

## IoT recommendations

There are **4** recommendations in this category.

[Expand table](#)

Recommendation	Description	Severity
<a href="#">Default IP Filter Policy should be Deny</a>	IP Filter Configuration should have rules defined for allowed traffic and should deny all other traffic by default (No related policy)	Medium
<a href="#">Diagnostic logs in IoT Hub should be enabled</a>	Enable logs and retain them for up to a year. This enables you to recreate activity trails for investigation purposes when a security incident occurs or your network is compromised. (Related policy: <a href="#">Diagnostic logs in IoT Hub should be enabled</a> )	Low
<a href="#">Identical Authentication Credentials</a>	Identical authentication credentials to the IoT Hub used by multiple devices. This could indicate an illegitimate device	High

Recommendation	Description	Severity
	impersonating a legitimate device. It also exposes the risk of device impersonation by an attacker  (No related policy)	
IP Filter rule large IP range ↗	An Allow IP Filter rule's source IP range is too large. Overly permissive rules might expose your IoT hub to malicious intenders  (No related policy)	Medium

## Networking recommendations

There are 13 recommendations in this category.

[+] Expand table

Recommendation	Description	Severity
Access to storage accounts with firewall and virtual network configurations should be restricted ↗	Review the settings of network access in your storage account firewall settings. We recommended configuring network rules so that only applications from allowed networks can access the storage account. To allow connections from specific internet or on-premise clients, access can be granted to traffic from specific Azure virtual networks or to public internet IP address ranges.  (Related policy: <a href="#">Storage accounts should restrict network access ↗</a> )	Low
Adaptive network hardening recommendations should be applied on internet facing virtual machines ↗	Defender for Cloud has analyzed the internet traffic communication patterns of the virtual machines listed below, and determined that the existing rules in the NSGs associated to them are overly-permissive, resulting in an increased potential attack surface.  This typically occurs when this IP address doesn't communicate regularly with this resource. Alternatively, the IP address has been flagged as malicious by Defender for Cloud's threat intelligence sources. Learn more in <a href="#">Improve your network security posture with adaptive network hardening</a> .  (Related policy: <a href="#">Adaptive network hardening recommendations should be applied on internet facing virtual machines ↗</a> )	High
All network ports should be restricted on network security groups	Defender for Cloud has identified some of your network security groups' inbound rules to be too permissive. Inbound rules should not allow access from 'Any' or	High

Recommendation	Description	Severity
<a href="#">associated to your virtual machine ↗</a>	'Internet' ranges. This can potentially enable attackers to target your resources. (Related policy: <a href="#">All network ports should be restricted on network security groups associated to your virtual machine ↗</a> )	
<a href="#">Azure DDoS Protection Standard should be enabled ↗</a>	Defender for Cloud has discovered virtual networks with Application Gateway resources unprotected by the DDoS protection service. These resources contain public IPs. Enable mitigation of network volumetric and protocol attacks. (Related policy: <a href="#">Azure DDoS Protection Standard should be enabled ↗</a> )	Medium
<a href="#">Internet-facing virtual machines should be protected with network security groups ↗</a>	Protect your VM from potential threats by restricting access to it with a network security group (NSG). NSGs contain a list of Access Control List (ACL) rules that allow or deny network traffic to your VM from other instances, in or outside the same subnet. To keep your machine as secure as possible, the VM access to the internet must be restricted and an NSG should be enabled on the subnet. VMs with 'High' severity are internet-facing VMs. (Related policy: <a href="#">Internet-facing virtual machines should be protected with network security groups ↗</a> )	High
<a href="#">IP forwarding on your virtual machine should be disabled ↗</a>	Defender for Cloud has discovered that IP forwarding is enabled on some of your virtual machines. Enabling IP forwarding on a virtual machine's NIC allows the machine to receive traffic addressed to other destinations. IP forwarding is rarely required (e.g., when using the VM as a network virtual appliance), and therefore, this should be reviewed by the network security team. (Related policy: <a href="#">IP Forwarding on your virtual machine should be disabled ↗</a> )	Medium
<a href="#">Machines should have ports closed that might expose attack vectors ↗</a>	Azure's terms of use <a href="#">↗</a> prohibit the use of Azure services in ways that could damage, disable, overburden, or impair any Microsoft server or the network. This recommendation lists exposed ports that need to be closed for your continued security. It also illustrates the potential threat to each port. (No related policy)	High
<a href="#">Management ports of virtual machines should be protected with just-in-time network access control ↗</a>	Defender for Cloud has identified some overly-permissive inbound rules for management ports in your Network Security Group. Enable just-in-time access control to protect your VM from internet-based brute-force attacks. Learn more in <a href="#">Understanding just-in-time (JIT) VM access</a> .	High

Recommendation	Description	Severity
	(Related policy: <a href="#">Management ports of virtual machines should be protected with just-in-time network access control ↗</a> )	
<a href="#">Management ports should be closed on your virtual machines ↗</a>	<p>Open remote management ports are exposing your VM to a high level of risk from Internet-based attacks. These attacks attempt to brute force credentials to gain admin access to the machine.</p> <p>(Related policy: <a href="#">Management ports should be closed on your virtual machines ↗</a>)</p>	Medium
<a href="#">Non-internet-facing virtual machines should be protected with network security groups ↗</a>	<p>Protect your non-internet-facing virtual machine from potential threats by restricting access to it with a network security group (NSG). NSGs contain a list of Access Control List (ACL) rules that allow or deny network traffic to your VM from other instances, whether or not they're on the same subnet.</p> <p>Note that to keep your machine as secure as possible, the VM's access to the internet must be restricted and an NSG should be enabled on the subnet.</p> <p>(Related policy: <a href="#">Non-internet-facing virtual machines should be protected with network security groups ↗</a>)</p>	Low
<a href="#">Secure transfer to storage accounts should be enabled ↗</a>	<p>Secure transfer is an option that forces your storage account to accept requests only from secure connections (HTTPS). Use of HTTPS ensures authentication between the server and the service and protects data in transit from network layer attacks such as man-in-the-middle, eavesdropping, and session-hijacking.</p> <p>(Related policy: <a href="#">Secure transfer to storage accounts should be enabled ↗</a>)</p>	High
<a href="#">Subnets should be associated with a network security group ↗</a>	<p>Protect your subnet from potential threats by restricting access to it with a network security group (NSG). NSGs contain a list of Access Control List (ACL) rules that allow or deny network traffic to your subnet. When an NSG is associated with a subnet, the ACL rules apply to all the VM instances and integrated services in that subnet, but don't apply to internal traffic inside the subnet. To secure resources in the same subnet from one another, enable NSG directly on the resources as well.</p> <p>Note that the following subnet types will be listed as not applicable: GatewaySubnet, AzureFirewallSubnet, AzureBastionSubnet.</p> <p>(Related policy: <a href="#">Subnets should be associated with a Network Security Group ↗</a>)</p>	Low
<a href="#">Virtual networks should be protected by Azure Firewall ↗</a>	<p>Some of your virtual networks aren't protected with a firewall. Use Azure Firewall to restrict access to your virtual networks and prevent potential threats. <a href="#">Learn more about Azure Firewall</a></p>	Low

Recommendation	Description	Severity
	Azure Firewall <a href="#">↗</a> . (Related policy: All Internet traffic should be routed via your deployed Azure Firewall <a href="#">↗</a> )	

## API recommendations

[\[+\] Expand table](#)

Recommendation	Description & related policy	Severity
Microsoft Defender for APIs should be enabled	Enable the Defender for APIs plan to discover and protect API resources against attacks and security misconfigurations. <a href="#">Learn more</a>	High
Azure API Management APIs should be onboarded to Defender for APIs.	Onboarding APIs to Defender for APIs requires compute and memory utilization on the Azure API Management service. Monitor performance of your Azure API Management service while onboarding APIs, and scale out your Azure API Management resources as needed.	High
API endpoints that are unused should be disabled and removed from the Azure API Management service	As a security best practice, API endpoints that haven't received traffic for 30 days are considered unused, and should be removed from the Azure API Management service. Keeping unused API endpoints might pose a security risk. These might be APIs that should have been deprecated from the Azure API Management service, but have accidentally been left active. Such APIs typically do not receive the most up-to-date security coverage.	Low
API endpoints in Azure API Management should be authenticated	API endpoints published within Azure API Management should enforce authentication to help minimize security risk. Authentication mechanisms are sometimes implemented incorrectly or are missing. This allows attackers to exploit implementation flaws and to access data. For APIs published in Azure API Management, this recommendation assesses authentication through verifying the presence of Azure API Management subscription keys for APIs or products where subscription is required, and the execution of policies for validating <a href="#">JWT</a> , <a href="#">client certificates</a> , and <a href="#">Microsoft Entra</a> tokens. If none of these authentication mechanisms are executed during the API call, the API will receive this recommendation.	High

## API management recommendations

Recommendation	Description & related policy	Severity
API Management subscriptions should not be scoped to all APIs	API Management subscriptions should be scoped to a product or an individual API instead of all APIs, which could result in excessive data exposure.	Medium
API Management calls to API backends should not bypass certificate thumbprint or name validation	API Management should validate the backend server certificate for all API calls. Enable SSL certificate thumbprint and name validation to improve the API security.	Medium
API Management direct management endpoint should not be enabled	The direct management REST API in Azure API Management bypasses Azure Resource Manager role-based access control, authorization, and throttling mechanisms, thus increasing the vulnerability of your service.	Low
API Management APIs should use only encrypted protocols	APIs should be available only through encrypted protocols, like HTTPS or WSS. Avoid using unsecured protocols, such as HTTP or WS to ensure security of data in transit.	High
API Management secret named values should be stored in Azure Key Vault	Named values are a collection of name and value pairs in each API Management service. Secret values can be stored either as encrypted text in API Management (custom secrets) or by referencing secrets in Azure Key Vault. Reference secret named values from Azure Key Vault to improve security of API Management and secrets. Azure Key Vault supports granular access management and secret rotation policies.	Medium
API Management should disable public network access to the service configuration endpoints	To improve the security of API Management services, restrict connectivity to service configuration endpoints, like direct access management API, Git configuration management endpoint, or self-hosted gateways configuration endpoint.	Medium
API Management minimum API version should be set to 2019-12-01 or higher	To prevent service secrets from being shared with read-only users, the minimum API version should be set to 2019-12-01 or higher.	Medium
API Management calls to API backends should be authenticated	Calls from API Management to backends should use some form of authentication, whether via certificates or credentials. Does not apply to Service Fabric backends.	Medium

# AI recommendations

 Expand table

Recommendation	Description & related policy	Severity
Resource logs in Azure Machine Learning Workspaces should be enabled (Preview)	Resource logs enable recreating activity trails to use for investigation purposes when a security incident occurs or when your network is compromised.	Medium
Azure Machine Learning Workspaces should disable public network access (Preview)	Disabling public network access improves security by ensuring that the Machine Learning Workspaces aren't exposed on the public internet. You can control exposure of your workspaces by creating private endpoints instead. For more information, see <a href="#">Configure a private endpoint for an Azure Machine Learning workspace</a> .	Medium
Azure Machine Learning Computes should be in a virtual network (Preview)	Azure Virtual Networks provide enhanced security and isolation for your Azure Machine Learning Compute Clusters and Instances, as well as subnets, access control policies, and other features to further restrict access. When a compute is configured with a virtual network, it is not publicly addressable and can only be accessed from virtual machines and applications within the virtual network.	Medium
Azure Machine Learning Computes should have local authentication methods disabled (Preview)	Disabling local authentication methods improves security by ensuring that Machine Learning Computes require Azure Active Directory identities exclusively for authentication. For more information, see <a href="#">Azure Policy Regulatory Compliance controls for Azure Machine Learning</a> .	Medium
Azure Machine Learning compute instances should be recreated to get the latest software updates (Preview)	Ensure Azure Machine Learning compute instances run on the latest available operating system. Security is improved and vulnerabilities reduced by running with the latest security patches. For more information, see <a href="#">Vulnerability management for Azure Machine Learning</a> .	Medium
Resource logs in Azure Databricks Workspaces should be enabled (Preview)	Resource logs enable recreating activity trails to use for investigation purposes when a security incident occurs or when your network is compromised.	Medium
Azure Databricks Workspaces should disable public network access (Preview)	Disabling public network access improves security by ensuring that the resource isn't exposed on the public internet. You can control exposure of your resources by	Medium

Recommendation	Description & related policy	Severity
Azure Databricks Clusters should disable public IP (Preview)	Disabling public IP of clusters in Azure Databricks Workspaces improves security by ensuring that the clusters aren't exposed on the public internet. For more information, see <a href="#">Secure cluster connectivity</a> .	Medium
Azure Databricks Workspaces should be in a virtual network (Preview)	Azure Virtual Networks provide enhanced security and isolation for your Azure Databricks Workspaces, as well as subnets, access control policies, and other features to further restrict access. For more information, see <a href="#">Deploy Azure Databricks in your Azure virtual network</a> .	Medium
Azure Databricks Workspaces should use private link (Preview)	Azure Private Link lets you connect your virtual networks to Azure services without a public IP address at the source or destination. The Private Link platform handles the connectivity between the consumer and services over the Azure backbone network. By mapping private endpoints to Azure Databricks workspaces, you can reduce data leakage risks. For more information, see <a href="#">Create the workspace and private endpoints in the Azure portal UI</a> .	Medium

## Deprecated recommendations

[+] [Expand table](#)

Recommendation	Description & related policy	Severity
Access to App Services should be restricted	Restrict access to your App Services by changing the networking configuration, to deny inbound traffic from ranges that are too broad. (Related policy: [Preview]: Access to App Services should be restricted)	High
The rules for web applications on IaaS NSGs should be hardened	Harden the network security group (NSG) of your virtual machines that are running web applications, with NSG rules that are overly permissive with regard to web application ports. (Related policy: The NSGs rules for web applications on IaaS should be hardened)	High
Pod Security Policies should be defined to reduce the attack vector by removing unnecessary application privileges (Preview)	Define Pod Security Policies to reduce the attack vector by removing unnecessary application privileges. It is recommended to configure pod security policies so pods can only access resources which they are allowed to access.	Medium

Recommendation	Description & related policy	Severity
	(Related policy: [Preview]: Pod Security Policies should be defined on Kubernetes Services)	
Install Azure Security Center for IoT security module to get more visibility into your IoT devices	Install Azure Security Center for IoT security module to get more visibility into your IoT devices.	Low
Your machines should be restarted to apply system updates	Restart your machines to apply the system updates and secure the machine from vulnerabilities. (Related policy: System updates should be installed on your machines)	Medium
Monitoring agent should be installed on your machines	This action installs a monitoring agent on the selected virtual machines. Select a workspace for the agent to report to. (No related policy)	High
Java should be updated to the latest version for web apps	<p>Periodically, newer versions are released for Java software either due to security flaws or to include additional functionality.</p> <p>Using the latest Java version for web apps is recommended to benefit from security fixes, if any, and/or new functionalities of the latest version.</p> <p>(Related policy: Ensure that 'Java version' is the latest, if used as a part of the Web app)</p>	Medium
Python should be updated to the latest version for function apps	<p>Periodically, newer versions are released for Python software either due to security flaws or to include additional functionality.</p> <p>Using the latest Python version for function apps is recommended to benefit from security fixes, if any, and/or new functionalities of the latest version.</p> <p>(Related policy: Ensure that 'Python version' is the latest, if used as a part of the Function app)</p>	Medium
Python should be updated to the latest version for web apps	<p>Periodically, newer versions are released for Python software either due to security flaws or to include additional functionality.</p> <p>Using the latest Python version for web apps is recommended to benefit from security fixes, if any, and/or new functionalities of the latest version.</p> <p>(Related policy: Ensure that 'Python version' is the latest, if used as a part of the Web app)</p>	Medium
Java should be updated to the latest version for function apps	Periodically, newer versions are released for Java software either due to security flaws or to include additional functionality.	Medium

Recommendation	Description & related policy	Severity
	<p>Using the latest Java version for function apps is recommended to benefit from security fixes, if any, and/or new functionalities of the latest version.</p> <p>(Related policy: Ensure that 'Java version' is the latest, if used as a part of the Function app)</p>	
PHP should be updated to the latest version for web apps	<p>Periodically, newer versions are released for PHP software either due to security flaws or to include additional functionality.</p> <p>Using the latest PHP version for web apps is recommended to benefit from security fixes, if any, and/or new functionalities of the latest version.</p> <p>(Related policy: Ensure that 'PHP version' is the latest, if used as a part of the WEB app)</p>	Medium

## Next steps

To learn more about recommendations, see the following:

- [What are security policies, initiatives, and recommendations?](#)
- [Review your security recommendations](#)

# Security recommendations for AWS resources - a reference guide

Article • 06/27/2023

This article lists the recommendations you might see in Microsoft Defender for Cloud if you've connected an AWS account from the [Environment settings](#) page. The recommendations shown in your environment depend on the resources you're protecting and your customized configuration.

To learn about how to respond to these recommendations, see [Remediate recommendations in Defender for Cloud](#).

Your secure score is based on the number of security recommendations you've completed. To decide which recommendations to resolve first, look at the severity of each one and its potential impact on your secure score.

## AWS Compute recommendations

There are 18 AWS recommendations in this category.

Recommendation	Description	Severity
Amazon EC2 instances managed by Systems Manager should have a patch compliance status of COMPLIANT after a patch installation ↗	This control checks whether the compliance status of the Amazon EC2 Systems Manager patch compliance is COMPLIANT or NON_COMPLIANT after the patch installation on the instance. It only checks instances that are managed by AWS Systems Manager Patch Manager. It does not check whether the patch was applied within the 30-day limit prescribed by PCI DSS requirement '6.2'. It also does not validate whether the patches applied were classified as security patches. You should create patching groups with the appropriate baseline settings and ensure in-scope systems are managed by those patch groups in Systems Manager. For more information about patch groups, see the <a href="#">AWS Systems Manager User Guide</a> ↗.	Medium

Recommendation	Description	Severity
<a href="#">Amazon EFS should be configured to encrypt file data at rest using AWS KMS</a>	<p>This control checks whether Amazon Elastic File System is configured to encrypt the file data using AWS KMS. The check fails in the following cases:</p> <ul style="list-style-type: none"> <li>* "Encrypted" is set to "false" in the <a href="#">DescribeFileSystems</a>.</li> </ul> <p>The "KmsKeyId" key in the <a href="#">DescribeFileSystems</a> response does not match the KmsKeyId parameter for <a href="#">efs-encrypted-check</a>. Note that this control does not use the "KmsKeyId" parameter for <a href="#">efs-encrypted-check</a>. It only checks the value of "Encrypted".</p> <p>For an added layer of security for your sensitive data in Amazon EFS, you should create encrypted file systems.</p> <p>Amazon EFS supports encryption for file systems at-rest. You can enable encryption of data at rest when you create an Amazon EFS file system.</p> <p>To learn more about Amazon EFS encryption, see <a href="#">Data encryption in Amazon EFS</a> in the Amazon Elastic File System User Guide.</p>	Medium
<a href="#">Amazon EFS volumes should be in backup plans</a>	<p>This control checks whether Amazon Elastic File System (Amazon EFS) file systems are added to the backup plans in AWS Backup. The control fails if Amazon EFS file systems are not included in the backup plans.</p> <p>Including EFS file systems in the backup plans helps you to protect your data from deletion and data loss.</p>	Medium
<a href="#">Application Load Balancer deletion protection should be enabled</a>	<p>This control checks whether an Application Load Balancer has deletion protection enabled. The control fails if deletion protection is not configured.</p> <p>Enable deletion protection to protect your Application Load Balancer from deletion.</p>	Medium
<a href="#">Auto Scaling groups associated with a load balancer should use health checks</a>	<p>Auto Scaling groups that are associated with a load balancer are using Elastic Load Balancing health checks.</p> <p>PCI DSS does not require load balancing or highly available configurations. This is recommended by AWS best practices.</p>	Low
<a href="#">AWS accounts should have Azure Arc auto provisioning enabled</a>	<p>For full visibility of the security content from Microsoft Defender for servers, EC2 instances should be connected to Azure Arc. To ensure that all eligible EC2 instances automatically receive Azure Arc, enable auto-provisioning from Defender for Cloud at the AWS account level. Learn more about <a href="#">Azure Arc</a>, and <a href="#">Microsoft Defender for Servers</a>.</p>	High

Recommendation	Description	Severity
<a href="#">CloudFront distributions should have origin failover configured</a>	<p>This control checks whether an Amazon CloudFront distribution is configured with an origin group that has two or more origins. CloudFront origin failover can increase availability. Origin failover automatically redirects traffic to a secondary origin if the primary origin is unavailable or if it returns specific HTTP response status codes.</p>	Medium
<a href="#">CodeBuild GitHub or Bitbucket source repository URLs should use OAuth</a>	<p>This control checks whether the GitHub or Bitbucket source repository URL contains either personal access tokens or a user name and password.</p> <p>Authentication credentials should never be stored or transmitted in clear text or appear in the repository URL. Instead of personal access tokens or user name and password, you should use OAuth to grant authorization for accessing GitHub or Bitbucket repositories.</p> <p>Using personal access tokens or a user name and password could expose your credentials to unintended data exposure and unauthorized access.</p>	High
<a href="#">CodeBuild project environment variables should not contain credentials</a>	<p>This control checks whether the project contains the environment variables <code>AWS_ACCESS_KEY_ID</code> and <code>AWS_SECRET_ACCESS_KEY</code>.</p> <p>Authentication credentials <code>AWS_ACCESS_KEY_ID</code> and <code>AWS_SECRET_ACCESS_KEY</code> should never be stored in clear text, as this could lead to unintended data exposure and unauthorized access.</p>	High
<a href="#">DynamoDB Accelerator (DAX) clusters should be encrypted at rest</a>	<p>This control checks whether a DAX cluster is encrypted at rest. Encrypting data at rest reduces the risk of data stored on disk being accessed by a user not authenticated to AWS. The encryption adds another set of access controls to limit the ability of unauthorized users to access to the data.</p> <p>For example, API permissions are required to decrypt the data before it can be read.</p>	Medium
<a href="#">DynamoDB tables should automatically scale capacity with demand</a>	<p>This control checks whether an Amazon DynamoDB table can scale its read and write capacity as needed. This control passes if the table uses either on-demand capacity mode or provisioned mode with auto scaling configured.</p> <p>Scaling capacity with demand avoids throttling exceptions, which helps to maintain availability of your applications.</p>	Medium
<a href="#">EC2 instances should be connected to Azure Arc</a>	<p>Connect your EC2 instances to Azure Arc in order to have full visibility to Microsoft Defender for Servers security content. Learn more about <a href="#">Azure Arc</a>, and about <a href="#">Microsoft Defender for Servers</a> on hybrid-cloud environment.</p>	High

Recommendation	Description	Severity
<a href="#">EC2 instances should be managed by AWS Systems Manager</a>	<p>Status of the Amazon EC2 Systems Manager patch compliance is 'COMPLIANT' or 'NON_COMPLIANT' after the patch installation on the instance.</p> <p>Only instances that are managed by AWS Systems Manager Patch Manager are checked. Patches that were applied within the 30-day limit prescribed by PCI DSS requirement '6' are not checked.</p>	Medium
<a href="#">Instances managed by Systems Manager should have an association</a>	<p>This control checks whether the status of the AWS Systems Manager association compliance is COMPLIANT or NON_COMPLIANT after the association is run on an instance. The control passes if the association compliance status is COMPLIANT.</p>	Low
<a href="#">compliance status of COMPLIANT</a>	<p>A State Manager association is a configuration that is assigned to your managed instances. The configuration defines the state that you want to maintain on your instances. For example, an association can specify that antivirus software must be installed and running on your instances, or that certain ports must be closed.</p> <p>After you create one or more State Manager associations, compliance status information is immediately available to you in the console or in response to AWS CLI commands or corresponding Systems Manager API operations. For associations, "Configuration" Compliance shows statuses of Compliant or Non-compliant and the severity level assigned to the association, such as "Critical" or "Medium". To learn more about State Manager association compliance, see <a href="#">About State Manager association compliance</a> in the AWS Systems Manager User Guide.</p> <p>You must configure your in-scope EC2 instances for Systems Manager association. You must also configure the patch baseline for the security rating of the vendor of patches, and set the autoapproval date to meet PCI DSS '3.2.1' requirement '6.2'. For additional guidance on how to <a href="#">Create an association</a>, see Create an association in the AWS Systems Manager User Guide. For additional information on working with patching in Systems Manager, see <a href="#">AWS Systems Manager Patch Manager</a> in the AWS Systems Manager User Guide.</p>	

Recommendation	Description	Severity
<a href="#">Lambda functions should have a dead-letter queue configured</a>	<p>This control checks whether a Lambda function is configured with a dead-letter queue. The control fails if the Lambda function is not configured with a dead-letter queue.</p> <p>As an alternative to an on-failure destination, you can configure your function with a dead-letter queue to save discarded events for further processing.</p> <p>A dead-letter queue acts the same as an on-failure destination. It is used when an event fails all processing attempts or expires without being processed.</p> <p>A dead-letter queue allows you to look back at errors or failed requests to your Lambda function to debug or identify unusual behavior.</p> <p>From a security perspective, it is important to understand why your function failed and to ensure that your function does not drop data or compromise data security as a result.</p> <p>For example, if your function cannot communicate to an underlying resource, that could be a symptom of a denial of service (DoS) attack elsewhere in the network.</p>	Medium
<a href="#">Lambda functions should use supported runtimes</a>	<p>This control checks that the Lambda function settings for runtimes match the expected values set for the supported runtimes for each language. This control checks for the following runtimes:</p> <p><code>nodejs14.x, nodejs12.x, nodejs10.x, python3.8, python3.7, python3.6, ruby2.7, ruby2.5, java11, java8, java8.al2, go1.x, dotnetcore3.1, dotnetcore2.1</code></p> <p><a href="#">Lambda runtimes</a> are built around a combination of operating system, programming language, and software libraries that are subject to maintenance and security updates. When a runtime component is no longer supported for security updates, Lambda deprecates the runtime. Even though you cannot create functions that use the deprecated runtime, the function is still available to process invocation events. Make sure that your Lambda functions are current and do not use out-of-date runtime environments.</p> <p>To learn more about the supported runtimes that this control checks for the supported languages, see <a href="#">AWS Lambda runtimes</a> in the AWS Lambda Developer Guide.</p>	Medium
<a href="#">Management ports of EC2 instances should be protected with just-in-time network access control</a>	<p>Microsoft Defender for Cloud has identified some overly-permissive inbound rules for management ports in your network. Enable just-in-time access control to protect your Instances from internet-based brute-force attacks. <a href="#">Learn more</a>.</p>	High

Recommendation	Description	Severity
Unused EC2 security groups should be removed ↗	Security groups should be attached to Amazon EC2 instances or to an ENI. A healthy finding can indicate there are unused Amazon EC2 security groups.	Low

## AWS Container recommendations

There are 3 AWS recommendations in this category.

Recommendation	Description	Severity
EKS clusters should grant the required AWS permissions to Microsoft Defender for Cloud ↗	<p>Microsoft Defender for Containers provides protections for your EKS clusters.</p> <p>To monitor your cluster for security vulnerabilities and threats, Defender for Containers needs permissions for your AWS account. These permissions will be used to enable Kubernetes control plane logging on your cluster and establish a reliable pipeline between your cluster and Defender for Cloud's backend in the cloud.</p> <p>Learn more about <a href="#">Microsoft Defender for Cloud's security features for containerized environments</a>.</p>	High
EKS clusters should have Microsoft Defender's extension for Azure Arc installed ↗	<p>Microsoft Defender's <a href="#">cluster extension</a> provides security capabilities for your EKS clusters. The extension collects data from a cluster and its nodes to identify security vulnerabilities and threats.</p> <p>The extension works with <a href="#">Azure Arc-enabled Kubernetes</a>.</p> <p>Learn more about <a href="#">Microsoft Defender for Cloud's security features for containerized environments</a>.</p>	High
Microsoft Defender for Containers should be enabled on AWS connectors ↗	<p>Microsoft Defender for Containers provides real-time threat protection for containerized environments and generates alerts about suspicious activities.</p> <p>Use this information to harden the security of Kubernetes clusters and remediate security issues.</p> <p><b>Important:</b> When you've enabled Microsoft Defender for Containers and deployed Azure Arc to your EKS clusters, the protections - and charges - will begin. If you don't deploy Azure Arc on a cluster, Defender for Containers will not protect it and no charges will be incurred for this Microsoft Defender plan for that cluster.</p>	High

# Data plane recommendations

All the data plane recommendations listed [here](#) are supported under AWS after [enabling the Azure policy extension](#).

## AWS Data recommendations

There are **66** AWS recommendations in this category.

Recommendation	Description	Severity
<a href="#">Amazon Aurora clusters should have backtracking enabled</a> ↗	<p>This control checks whether Amazon Aurora clusters have backtracking enabled.</p> <p>Backups help you to recover more quickly from a security incident. They also strengthen the resilience of your systems.</p> <p>Aurora backtracking reduces the time to recover a database to a point in time. It doesn't require a database restore to do so.</p> <p>For more information about backtracking in Aurora, see <a href="#">Backtracking an Aurora DB cluster</a> ↗ in the Amazon Aurora User Guide.</p>	Medium
<a href="#">Amazon EBS snapshots shouldn't be publicly restorable</a> ↗	Amazon EBS snapshots shouldn't be publicly restorable by everyone unless explicitly allowed, to avoid accidental exposure of data. Additionally, permission to change Amazon EBS configurations should be restricted to authorized AWS accounts only.	High
<a href="#">Amazon ECS task definitions should have secure networking modes and user definitions</a> ↗	<p>This control checks whether an active Amazon ECS task definition that has host networking mode also has privileged or user container definitions.</p> <p>The control fails for task definitions that have host network mode and container definitions where privileged=false or is empty and user=root or is empty.</p> <p>If a task definition has elevated privileges, it is because the customer has specifically opted in to that configuration.</p> <p>This control checks for unexpected privilege escalation when a task definition has host networking enabled but the customer hasn't opted in to elevated privileges.</p>	High

Recommendation	Description	Severity
<a href="#">Amazon Elasticsearch Service domains should encrypt data sent between nodes</a>	This control checks whether Amazon ES domains have node-to-node encryption enabled. HTTPS (TLS) can be used to help prevent potential attackers from eavesdropping on or manipulating network traffic using person-in-the-middle or similar attacks. Only encrypted connections over HTTPS (TLS) should be allowed. Enabling node-to-node encryption for Amazon ES domains ensures that intra-cluster communications are encrypted in transit. There can be a performance penalty associated with this configuration. You should be aware of and test the performance trade-off before enabling this option.	Medium
<a href="#">Amazon Elasticsearch Service domains should have encryption at rest enabled</a>	It's important to enable encryption of the rest of Amazon ES domains to protect sensitive data	Medium
<a href="#">Amazon RDS database should be encrypted using customer managed key</a>	This check identifies RDS databases that are encrypted with default KMS keys and not with customer managed keys. As a leading practice, use customer managed keys to encrypt the data on your RDS databases and maintain control of your keys and data on sensitive workloads.	Medium
<a href="#">Amazon RDS instance should be configured with automatic backup settings</a>	This check identifies RDS instances, which aren't set with the automatic backup setting. If Automatic Backup is set, RDS creates a storage volume snapshot of your DB instance, backing up the entire DB instance and not just individual databases, which provide for point-in-time recovery. The automatic backup will happen during the specified backup window time and keeps the backups for a limited period of time as defined in the retention period. It's recommended to set automatic backups for your critical RDS servers that will help in the data restoration process.	Medium
<a href="#">Amazon Redshift clusters should have audit logging enabled</a>	This control checks whether an Amazon Redshift cluster has audit logging enabled.  Amazon Redshift audit logging provides additional information about connections and user activities in your cluster. This data can be stored and secured in Amazon S3 and can be helpful in security audits and investigations. For more information, see <a href="#">Database audit logging</a> in the <i>Amazon Redshift Cluster Management Guide</i> .	Medium

Recommendation	Description	Severity
<a href="#">Amazon Redshift clusters should have automatic snapshots enabled</a>	This control checks whether Amazon Redshift clusters have automated snapshots enabled. It also checks whether the snapshot retention period is greater than or equal to seven. Backups help you to recover more quickly from a security incident. They strengthen the resilience of your systems. Amazon Redshift takes periodic snapshots by default. This control checks whether automatic snapshots are enabled and retained for at least seven days. For more details on Amazon Redshift automated snapshots, see <a href="#">Automated snapshots</a> in the <i>Amazon Redshift Cluster Management Guide</i> .	Medium
<a href="#">Amazon Redshift clusters should prohibit public access</a>	We recommend Amazon Redshift clusters to avoid public accessibility by evaluating the 'publiclyAccessible' field in the cluster configuration item.	High
<a href="#">Amazon Redshift should have automatic upgrades to major versions enabled</a>	This control checks whether automatic major version upgrades are enabled for the Amazon Redshift cluster. Enabling automatic major version upgrades ensures that the latest major version updates to Amazon Redshift clusters are installed during the maintenance window. These updates might include security patches and bug fixes. Keeping up to date with patch installation is an important step in securing systems.	Medium
<a href="#">Amazon SQS queues should be encrypted at rest</a>	This control checks whether Amazon SQS queues are encrypted at rest. Server-side encryption (SSE) allows you to transmit sensitive data in encrypted queues. To protect the content of messages in queues, SSE uses keys managed in AWS KMS. For more information, see <a href="#">Encryption at rest</a> in the Amazon Simple Queue Service Developer Guide.	Medium
<a href="#">An RDS event notifications subscription should be configured for critical cluster events</a>	This control checks whether an Amazon RDS event subscription exists that has notifications enabled for the following source type, event category key-value pairs. DBCluster: ["maintenance" and "failure"]. RDS event notifications use Amazon SNS to make you aware of changes in the availability or configuration of your RDS resources. These notifications allow for rapid response. For more information about RDS event notifications, see <a href="#">Using Amazon RDS event notification in the Amazon RDS User Guide</a> .	Low

Recommendation	Description	Severity
<a href="#">An RDS event notifications subscription should be configured for critical database instance events ↗</a>	<p>This control checks whether an Amazon RDS event subscription exists with notifications enabled for the following source type, event category key-value pairs. DBInstance: ["maintenance", "configuration change" and "failure"].</p> <p>RDS event notifications use Amazon SNS to make you aware of changes in the availability or configuration of your RDS resources. These notifications allow for rapid response.</p> <p>For more information about RDS event notifications, see <a href="#">Using Amazon RDS event notification ↗</a> in the Amazon RDS User Guide.</p>	Low
<a href="#">An RDS event notifications subscription should be configured for critical database parameter group events ↗</a>	<p>This control checks whether an Amazon RDS event subscription exists with notifications enabled for the following source type, event category key-value pairs. DBParameterGroup: ["configuration", "change"].</p> <p>RDS event notifications use Amazon SNS to make you aware of changes in the availability or configuration of your RDS resources. These notifications allow for rapid response.</p> <p>For more information about RDS event notifications, see <a href="#">Using Amazon RDS event notification ↗</a> in the Amazon RDS User Guide.</p>	Low
<a href="#">An RDS event notifications subscription should be configured for critical database security group events ↗</a>	<p>This control checks whether an Amazon RDS event subscription exists with notifications enabled for the following source type, event category key-value pairs. DBSecurityGroup: ["configuration", "change", "failure"].</p> <p>RDS event notifications use Amazon SNS to make you aware of changes in the availability or configuration of your RDS resources. These notifications allow for a rapid response.</p> <p>For more information about RDS event notifications , see <a href="#">Using Amazon RDS event notification ↗</a> in the Amazon RDS User Guide.</p>	Low
<a href="#">API Gateway REST and WebSocket API logging should be enabled ↗</a>	<p>This control checks whether all stages of an Amazon API Gateway REST or WebSocket API have logging enabled.</p> <p>The control fails if logging isn't enabled for all methods of a stage or if logging Level is neither ERROR nor INFO.</p> <p>API Gateway REST or WebSocket API stages should have relevant logs enabled. API Gateway REST and WebSocket API execution logging provides detailed records of requests made to API Gateway REST and WebSocket API stages.</p> <p>The stages include API integration backend responses, Lambda authorizer responses, and the requestId for AWS integration endpoints.</p>	Medium

Recommendation	Description	Severity
<a href="#">API Gateway REST API cache data should be encrypted at rest</a>	<p>This control checks whether all methods in API Gateway REST API stages that have cache enabled are encrypted. The control fails if any method in an API Gateway REST API stage is configured to cache and the cache isn't encrypted.</p> <p>Encrypting data at rest reduces the risk of data stored on disk being accessed by a user not authenticated to AWS. It adds another set of access controls to limit unauthorized users ability access the data. For example, API permissions are required to decrypt the data before it can be read.</p> <p>API Gateway REST API caches should be encrypted at rest for an added layer of security.</p>	Medium
<a href="#">API Gateway REST API stages should be configured to use SSL certificates for backend authentication</a>	<p>This control checks whether Amazon API Gateway REST API stages have SSL certificates configured.</p> <p>Backend systems use these certificates to authenticate that incoming requests are from API Gateway.</p> <p>API Gateway REST API stages should be configured with SSL certificates to allow backend systems to authenticate that requests originate from API Gateway.</p>	Medium
<a href="#">API Gateway REST API stages should have AWS X-Ray tracing enabled</a>	<p>This control checks whether AWS X-Ray active tracing is enabled for your Amazon API Gateway REST API stages.</p> <p>X-Ray active tracing enables a more rapid response to performance changes in the underlying infrastructure. Changes in performance could result in a lack of availability of the API.</p> <p>X-Ray active tracing provides real-time metrics of user requests that flow through your API Gateway REST API operations and connected services.</p>	Low
<a href="#">API Gateway should be associated with an AWS WAF web ACL</a>	<p>This control checks whether an API Gateway stage uses an AWS WAF web access control list (ACL).</p> <p>This control fails if an AWS WAF web ACL isn't attached to a REST API Gateway stage.</p> <p>AWS WAF is a web application firewall that helps protect web applications and APIs from attacks. It enables you to configure an ACL, which is a set of rules that allow, block, or count web requests based on customizable web security rules and conditions that you define.</p> <p>Ensure that your API Gateway stage is associated with an AWS WAF web ACL to help protect it from malicious attacks.</p>	Medium

Recommendation	Description	Severity
<a href="#">Application and Classic Load Balancers logging should be enabled ↗</a>	<p>This control checks whether the Application Load Balancer and the Classic Load Balancer have logging enabled. The control fails if access_logs.s3.enabled is false.</p> <p>Elastic Load Balancing provides access logs that capture detailed information about requests sent to your load balancer. Each log contains information such as the time the request was received, the client's IP address, latencies, request paths, and server responses. You can use these access logs to analyze traffic patterns and to troubleshoot issues.</p> <p>To learn more, see <a href="#">Access logs for your Classic Load Balancer ↗</a> in User Guide for Classic Load Balancers.</p>	Medium
<a href="#">Attached EBS volumes should be encrypted at-rest ↗</a>	<p>This control checks whether the EBS volumes that are in an attached state are encrypted. To pass this check, EBS volumes must be in use and encrypted. If the EBS volume isn't attached, then it isn't subject to this check.</p> <p>For an added layer of security of your sensitive data in EBS volumes, you should enable EBS encryption at rest. Amazon EBS encryption offers a straightforward encryption solution for your EBS resources that doesn't require you to build, maintain, and secure your own key management infrastructure. It uses AWS KMS customer master keys (CMK) when creating encrypted volumes and snapshots.</p> <p>To learn more about Amazon EBS encryption, see <a href="#">Amazon EBS encryption ↗</a> in the Amazon EC2 User Guide for Linux Instances.</p>	Medium
<a href="#">AWS Database Migration Service replication instances shouldn't be public ↗</a>	<p>To protect your replicated instances from threats. A private replication instance should have a private IP address that you can't access outside of the replication network.</p> <p>A replication instance should have a private IP address when the source and target databases are in the same network, and the network is connected to the replication instance's VPC using a VPN, AWS Direct Connect, or VPC peering.</p> <p>You should also ensure that access to your AWS DMS instance configuration is limited to only authorized users.</p> <p>To do this, restrict users' IAM permissions to modify AWS DMS settings and resources.</p>	High

Recommendation	Description	Severity
<a href="#">Classic Load Balancer listeners should be configured with HTTPS or TLS termination ↗</a>	<p>This control checks whether your Classic Load Balancer listeners are configured with HTTPS or TLS protocol for front-end (client to load balancer) connections. The control is applicable if a Classic Load Balancer has listeners. If your Classic Load Balancer doesn't have a listener configured, then the control doesn't report any findings.</p> <p>The control passes if the Classic Load Balancer listeners are configured with TLS or HTTPS for front-end connections.</p> <p>The control fails if the listener isn't configured with TLS or HTTPS for front-end connections.</p> <p>Before you start to use a load balancer, you must add one or more listeners. A listener is a process that uses the configured protocol and port to check for connection requests. Listeners can support both HTTP and HTTPS/TLS protocols. You should always use an HTTPS or TLS listener, so that the load balancer does the work of encryption and decryption in transit.</p>	Medium
<a href="#">Classic Load Balancers should have connection draining enabled ↗</a>	<p>This control checks whether Classic Load Balancers have connection draining enabled.</p> <p>Enabling connection draining on Classic Load Balancers ensures that the load balancer stops sending requests to instances that are de-registering or unhealthy. It keeps the existing connections open. This is useful for instances in Auto Scaling groups, to ensure that connections aren't severed abruptly.</p>	Medium
<a href="#">CloudFront distributions should have AWS WAF enabled ↗</a>	<p>This control checks whether CloudFront distributions are associated with either AWS WAF or AWS WAFv2 web ACLs. The control fails if the distribution isn't associated with a web ACL.</p> <p>AWS WAF is a web application firewall that helps protect web applications and APIs from attacks. It allows you to configure a set of rules, called a web access control list (web ACL), that allow, block, or count web requests based on customizable web security rules and conditions that you define. Ensure your CloudFront distribution is associated with an AWS WAF web ACL to help protect it from malicious attacks.</p>	Medium

Recommendation	Description	Severity
<a href="#">CloudFront distributions should have logging enabled</a> ↗	<p>This control checks whether server access logging is enabled on CloudFront distributions. The control fails if access logging isn't enabled for a distribution.</p> <p>CloudFront access logs provide detailed information about every user request that CloudFront receives. Each log contains information such as the date and time the request was received, the IP address of the viewer that made the request, the source of the request, and the port number of the request from the viewer. These logs are useful for applications such as security and access audits and forensics investigation. For more guidance on how to analyze access logs, see <a href="#">Querying Amazon CloudFront logs</a> in the Amazon Athena User Guide.</p>	Medium
<a href="#">CloudFront distributions should require encryption in transit</a> ↗	<p>This control checks whether an Amazon CloudFront distribution requires viewers to use HTTPS directly or whether it uses redirection. The control fails if ViewerProtocolPolicy is set to allow-all for defaultCacheBehavior or for cacheBehaviors.</p> <p>HTTPS (TLS) can be used to help prevent potential attackers from using person-in-the-middle or similar attacks to eavesdrop on or manipulate network traffic. Only encrypted connections over HTTPS (TLS) should be allowed. Encrypting data in transit can affect performance. You should test your application with this feature to understand the performance profile and the impact of TLS.</p>	Medium
<a href="#">CloudTrail logs should be encrypted at rest using KMS CMKs</a> ↗	<p>We recommended to configure CloudTrail use SSE-KMS. Configuring CloudTrail to use SSE-KMS provides more confidentiality controls on log data as a given user must have S3 read permission on the corresponding log bucket and must be granted decrypt permission by the CMK policy.</p>	Medium
<a href="#">Connections to Amazon Redshift clusters should be encrypted in transit</a> ↗	<p>This control checks whether connections to Amazon Redshift clusters are required to use encryption in transit. The check fails if the Amazon Redshift cluster parameter require_SSL isn't set to '1'. TLS can be used to help prevent potential attackers from using person-in-the-middle or similar attacks to eavesdrop on or manipulate network traffic. Only encrypted connections over TLS should be allowed. Encrypting data in transit can affect performance. You should test your application with this feature to understand the performance profile and the impact of TLS.</p>	Medium

Recommendation	Description	Severity
<a href="#">Connections to Elasticsearch domains should be encrypted using TLS 1.2</a>	<p>This control checks whether connections to Elasticsearch domains are required to use TLS 1.2. The check fails if the Elasticsearch domain TLSSecurityPolicy isn't Policy-Min-TLS-1-2-2019-07.</p>	Medium
	<p>HTTPS (TLS) can be used to help prevent potential attackers from using person-in-the-middle or similar attacks to eavesdrop on or manipulate network traffic. Only encrypted connections over HTTPS (TLS) should be allowed. Encrypting data in transit can affect performance. You should test your application with this feature to understand the performance profile and the impact of TLS. TLS 1.2 provides several security enhancements over previous versions of TLS.</p>	
<a href="#">DynamoDB tables should have point-in-time recovery enabled</a>	<p>This control checks whether point-in-time recovery (PITR) is enabled for an Amazon DynamoDB table.</p> <p>Backups help you to recover more quickly from a security incident. They also strengthen the resilience of your systems.</p> <p>DynamoDB point-in-time recovery automates backups for DynamoDB tables. It reduces the time to recover from accidental delete or write operations.</p> <p>DynamoDB tables that have PITR enabled can be restored to any point in time in the last 35 days.</p>	Medium
<a href="#">EBS default encryption should be enabled</a>	<p>This control checks whether account-level encryption is enabled by default for Amazon Elastic Block Store(Amazon EBS).</p> <p>The control fails if the account level encryption isn't enabled.</p> <p>When encryption is enabled for your account, Amazon EBS volumes and snapshot copies are encrypted at rest. This adds an more layer of protection for your data.</p> <p>For more information, see <a href="#">Encryption by default</a> in the Amazon EC2 User Guide for Linux Instances.</p> <p>Note that following instance types don't support encryption: R1, C1, and M1.</p>	Medium
<a href="#">Elastic Beanstalk environments should have enhanced health reporting enabled</a>	<p>This control checks whether enhanced health reporting is enabled for your AWS Elastic Beanstalk environments.</p> <p>Elastic Beanstalk enhanced health reporting enables a more rapid response to changes in the health of the underlying infrastructure. These changes could result in a lack of availability of the application.</p> <p>Elastic Beanstalk enhanced health reporting provides a status descriptor to gauge the severity of the identified issues and identify possible causes to investigate. The Elastic Beanstalk health agent, included in supported Amazon Machine Images (AMIs), evaluates logs and metrics of environment EC2 instances.</p>	Low

Recommendation	Description	Severity
<a href="#">Elastic Beanstalk managed platform updates should be enabled ↗</a>	<p>This control checks whether managed platform updates are enabled for the Elastic Beanstalk environment.</p> <p>Enabling managed platform updates ensures that the latest available platform fixes, updates, and features for the environment are installed. Keeping up to date with patch installation is an important step in securing systems.</p>	High
<a href="#">Elastic Load Balancer shouldn't have ACM certificate expired or expiring in 90 days. ↗</a>	<p>This check identifies Elastic Load Balancers (ELB) which are using ACM certificates expired or expiring in 90 days. AWS Certificate Manager (ACM) is the preferred tool to provision, manage, and deploy your server certificates. With ACM you can request a certificate or deploy an existing ACM or external certificate to AWS resources. As a best practice, it's recommended to reimport expiring/expired certificates while preserving the ELB associations of the original certificate.</p>	High
<a href="#">Elasticsearch domain error logging to CloudWatch Logs should be enabled ↗</a>	<p>This control checks whether Elasticsearch domains are configured to send error logs to CloudWatch Logs.</p> <p>You should enable error logs for Elasticsearch domains and send those logs to CloudWatch Logs for retention and response.</p> <p>Domain error logs can assist with security and access audits, and can help to diagnose availability issues.</p>	Medium
<a href="#">Elasticsearch domains should be configured with at least three dedicated master nodes ↗</a>	<p>This control checks whether Elasticsearch domains are configured with at least three dedicated master nodes. This control fails if the domain doesn't use dedicated master nodes. This control passes if Elasticsearch domains have five dedicated master nodes. However, using more than three master nodes might be unnecessary to mitigate the availability risk, and will result in more cost.</p> <p>An Elasticsearch domain requires at least three dedicated master nodes for high availability and fault-tolerance. Dedicated master node resources can be strained during data node blue/green deployments because there are more nodes to manage.</p> <p>Deploying an Elasticsearch domain with at least three dedicated master nodes ensures sufficient master node resource capacity and cluster operations if a node fails.</p>	Medium
<a href="#">Elasticsearch domains should have at least three data nodes ↗</a>	<p>This control checks whether Elasticsearch domains are configured with at least three data nodes and zoneAwarenessEnabled is true.</p> <p>An Elasticsearch domain requires at least three data nodes for high availability and fault-tolerance. Deploying an Elasticsearch domain with at least three data nodes ensures cluster operations if a node fails.</p>	Medium

Recommendation	Description	Severity
<a href="#">Elasticsearch domains should have audit logging enabled</a> ↗	<p>This control checks whether Elasticsearch domains have audit logging enabled. This control fails if an Elasticsearch domain doesn't have audit logging enabled.</p> <p>Audit logs are highly customizable. They allow you to track user activity on your Elasticsearch clusters, including authentication successes and failures, requests to OpenSearch, index changes, and incoming search queries.</p>	Medium
<a href="#">Enhanced monitoring should be configured for RDS DB instances and clusters</a> ↗	<p>This control checks whether enhanced monitoring is enabled for your RDS DB instances.</p> <p>In Amazon RDS, Enhanced Monitoring enables a more rapid response to performance changes in underlying infrastructure. These performance changes could result in a lack of availability of the data. Enhanced Monitoring provides real-time metrics of the operating system that your RDS DB instance runs on. An agent is installed on the instance. The agent can obtain metrics more accurately than is possible from the hypervisor layer.</p> <p>Enhanced Monitoring metrics are useful when you want to see how different processes or threads on a DB instance use the CPU.</p> <p>For more information, see <a href="#">Enhanced Monitoring</a> ↗ in the <i>Amazon RDS User Guide</i>.</p>	Low
<a href="#">Ensure rotation for customer created CMKs is enabled</a> ↗	<p>AWS Key Management Service (KMS) allows customers to rotate the backing key which is key material stored within the KMS which is tied to the key ID of the Customer Created customer master key (CMK).</p> <p>It's the backing key that is used to perform cryptographic operations such as encryption and decryption.</p> <p>Automated key rotation currently retains all prior backing keys so that decryption of encrypted data can take place transparently.</p> <p>It's recommended that CMK key rotation be enabled.</p> <p>Rotating encryption keys helps reduce the potential impact of a compromised key as data encrypted with a new key can't be accessed with a previous key that may have been exposed.</p>	Medium

Recommendation	Description	Severity
<a href="#">Ensure S3 bucket access logging is enabled on the CloudTrail S3 bucket ↗</a>	<p>S3 Bucket Access Logging generates a log that contains access records. Ensure S3 bucket access logging is enabled on the CloudTrail S3 bucket for each request made to your S3 bucket. An access log record contains details about the request, such as the request type, the resources specified in the request worked, and the time and date the request was processed.</p> <p>It's recommended that bucket access logging be enabled on the CloudTrail S3 bucket.</p> <p>By enabling S3 bucket logging on target S3 buckets, it's possible to capture all events, which may affect objects within target buckets. Configuring logs to be placed in a separate bucket allows access to log information, which can be useful in security and incident response workflows.</p>	Low
<a href="#">Ensure the S3 bucket used to store CloudTrail logs isn't publicly accessible ↗</a>	<p>CloudTrail logs a record of every API call made in your AWS account. These log files are stored in an S3 bucket.</p> <p>It's recommended that the bucket policy, or access control list (ACL), applied to the S3 bucket that CloudTrail logs to prevents public access to the CloudTrail logs.</p> <p>Allowing public access to CloudTrail log content may aid an adversary in identifying weaknesses in the affected account's use or configuration.</p>	High
<a href="#">IAM shouldn't have expired SSL/TLS certificates ↗</a>	<p>This check identifies expired SSL/TLS certificates. To enable HTTPS connections to your website or application in AWS, you need an SSL/TLS server certificate. You can use ACM or IAM to store and deploy server certificates. Removing expired SSL/TLS certificates eliminates the risk that an invalid certificate will be deployed accidentally to a resource such as AWS Elastic Load Balancer (ELB), which can damage the credibility of the application/website behind the ELB. This check generates alerts if there are any expired SSL/TLS certificates stored in AWS IAM. As a best practice, it's recommended to delete expired certificates.</p>	High
<a href="#">Imported ACM certificates should be renewed after a specified time period ↗</a>	<p>This control checks whether ACM certificates in your account are marked for expiration within 30 days. It checks both imported certificates and certificates provided by AWS Certificate Manager. ACM can automatically renew certificates that use DNS validation. For certificates that use email validation, you must respond to a domain validation email.</p> <p>ACM also doesn't automatically renew certificates that you import. You must renew imported certificates manually.</p> <p>For more information about managed renewal for ACM certificates, see <a href="#">Managed renewal for ACM certificates ↗</a> in the AWS Certificate Manager User Guide.</p>	Medium

Recommendation	Description	Severity
<a href="#">Over-provisioned identities in accounts should be investigated to reduce the Permission Creep Index (PCI) ↗</a>	Over-provisioned identities in accounts should be investigated to reduce the Permission Creep Index (PCI) and to safeguard your infrastructure. Reduce the PCI by removing the unused high risk permission assignments. High PCI reflects risk associated with the identities with permissions that exceed their normal or required usage	Medium
<a href="#">RDS automatic minor version upgrades should be enabled ↗</a>	This control checks whether automatic minor version upgrades are enabled for the RDS database instance. Enabling automatic minor version upgrades ensures that the latest minor version updates to the relational database management system (RDBMS) are installed. These upgrades might include security patches and bug fixes. Keeping up to date with patch installation is an important step in securing systems.	High
<a href="#">RDS cluster snapshots and database snapshots should be encrypted at rest ↗</a>	This control checks whether RDS DB snapshots are encrypted. This control is intended for RDS DB instances. However, it can also generate findings for snapshots of Aurora DB instances, Neptune DB instances, and Amazon DocumentDB clusters. If these findings aren't useful, then you can suppress them. Encrypting data at rest reduces the risk that an unauthenticated user gets access to data that is stored on disk. Data in RDS snapshots should be encrypted at rest for an added layer of security.	Medium
<a href="#">RDS clusters should have deletion protection enabled ↗</a>	This control checks whether RDS clusters have deletion protection enabled. This control is intended for RDS DB instances. However, it can also generate findings for Aurora DB instances, Neptune DB instances, and Amazon DocumentDB clusters. If these findings aren't useful, then you can suppress them. Enabling cluster deletion protection is an more layer of protection against accidental database deletion or deletion by an unauthorized entity. When deletion protection is enabled, an RDS cluster can't be deleted. Before a deletion request can succeed, deletion protection must be disabled.	Low
<a href="#">RDS DB clusters should be configured for multiple Availability Zones ↗</a>	RDS DB clusters should be configured for multiple the data that is stored. Deployment to multiple Availability Zones allows for automate Availability Zones to ensure availability of ed failover in the event of an Availability Zone availability issue and during regular RDS maintenance events.	Medium

Recommendation	Description	Severity
<a href="#">RDS DB clusters should be configured to copy tags to snapshots ↴</a>	<p>Identification and inventory of your IT assets is a crucial aspect of governance and security.</p> <p>You need to have visibility of all your RDS DB clusters so that you can assess their security posture and act on potential areas of weakness.</p> <p>Snapshots should be tagged in the same way as their parent RDS database clusters.</p> <p>Enabling this setting ensures that snapshots inherit the tags of their parent database clusters.</p>	Low
<a href="#">RDS DB instances should be configured to copy tags to snapshots ↴</a>	<p>This control checks whether RDS DB instances are configured to copy all tags to snapshots when the snapshots are created.</p> <p>Identification and inventory of your IT assets is a crucial aspect of governance and security.</p> <p>You need to have visibility of all your RDS DB instances so that you can assess their security posture and take action on potential areas of weakness.</p> <p>Snapshots should be tagged in the same way as their parent RDS database instances. Enabling this setting ensures that snapshots inherit the tags of their parent database instances.</p>	Low
<a href="#">RDS DB instances should be configured with multiple Availability Zones ↴</a>	<p>This control checks whether high availability is enabled for your RDS DB instances.</p> <p>RDS DB instances should be configured for multiple Availability Zones (AZs). This ensures the availability of the data stored.</p> <p>Multi-AZ deployments allow for automated failover if there's an issue with Availability Zone availability and during regular RDS maintenance.</p>	Medium
<a href="#">RDS DB instances should have deletion protection enabled ↴</a>	<p>This control checks whether your RDS DB instances that use one of the listed database engines have deletion protection enabled.</p> <p>Enabling instance deletion protection is an more layer of protection against accidental database deletion or deletion by an unauthorized entity.</p> <p>While deletion protection is enabled, an RDS DB instance can't be deleted. Before a deletion request can succeed, deletion protection must be disabled.</p>	Low

Recommendation	Description	Severity
<a href="#">RDS DB instances should have encryption at rest enabled ↗</a>	<p>This control checks whether storage encryption is enabled for your Amazon RDS DB instances.</p> <p>This control is intended for RDS DB instances. However, it can also generate findings for Aurora DB instances, Neptune DB instances, and Amazon DocumentDB clusters. If these findings aren't useful, then you can suppress them.</p> <p>For an added layer of security for your sensitive data in RDS DB instances, you should configure your RDS DB instances to be encrypted at rest. To encrypt your RDS DB instances and snapshots at rest, enable the encryption option for your RDS DB instances. Data that is encrypted at rest includes the underlying storage for DB instances, its automated backups, read replicas, and snapshots.</p> <p>RDS encrypted DB instances use the open standard AES-256 encryption algorithm to encrypt your data on the server that hosts your RDS DB instances. After your data is encrypted, Amazon RDS handles authentication of access and decryption of your data transparently with a minimal impact on performance. You don't need to modify your database client applications to use encryption.</p> <p>Amazon RDS encryption is currently available for all database engines and storage types. Amazon RDS encryption is available for most DB instance classes. To learn about DB instance classes that don't support Amazon RDS encryption, see <a href="#">Encrypting Amazon RDS resources ↗</a> in the <i>Amazon RDS User Guide</i>.</p>	Medium
<a href="#">RDS DB Instances should prohibit public access ↗</a>	<p>We recommend that you also ensure that access to your RDS instance's configuration is limited to authorized users only, by restricting users' IAM permissions to modify RDS instances' settings and resources.</p>	High
<a href="#">RDS snapshots should prohibit public access ↗</a>	<p>We recommend only allowing authorized principals to access the snapshot and change Amazon RDS configuration.</p>	High

Recommendation	Description	Severity
<a href="#">Remove unused Secrets Manager secrets ↗</a>	<p>This control checks whether your secrets have been accessed within a specified number of days. The default value is 90 days. If a secret wasn't accessed within the defined number of days, this control fails.</p> <p>Deleting unused secrets is as important as rotating secrets. Unused secrets can be abused by their former users, who no longer need access to these secrets. Also, as more users get access to a secret, someone might have mishandled and leaked it to an unauthorized entity, which increases the risk of abuse.</p> <p>Deleting unused secrets helps revoke secret access from users who no longer need it. It also helps to reduce the cost of using Secrets Manager. Therefore, it's essential to routinely delete unused secrets.</p>	Medium
<a href="#">S3 buckets should have cross-region replication enabled ↗</a>	<p>Enabling S3 cross-region replication ensures that multiple versions of the data are available in different distinct Regions. This allows you to protect your S3 bucket against DDoS attacks and data corruption events.</p>	Low
<a href="#">S3 buckets should have server-side encryption enabled ↗</a>	<p>Enable server-side encryption to protect data in your S3 buckets. Encrypting the data can prevent access to sensitive data in the event of a data breach.</p>	Medium
<a href="#">Secrets Manager secrets configured with automatic rotation should rotate successfully ↗</a>	<p>This control checks whether an AWS Secrets Manager secret rotated successfully based on the rotation schedule. The control fails if <b>RotationOccurringAsScheduled</b> is <b>false</b>. The control doesn't evaluate secrets that don't have rotation configured.</p> <p>Secrets Manager helps you improve the security posture of your organization. Secrets include database credentials, passwords, and third-party API keys. You can use Secrets Manager to store secrets centrally, encrypt secrets automatically, control access to secrets, and rotate secrets safely and automatically.</p> <p>Secrets Manager can rotate secrets. You can use rotation to replace long-term secrets with short-term ones. Rotating your secrets limits how long an unauthorized user can use a compromised secret. For this reason, you should rotate your secrets frequently.</p> <p>In addition to configuring secrets to rotate automatically, you should ensure that those secrets rotate successfully based on the rotation schedule.</p> <p>To learn more about rotation, see <a href="#">Rotating your AWS Secrets Manager secrets ↗</a> in the AWS Secrets Manager User Guide.</p>	Medium

Recommendation	Description	Severity
<a href="#">Secrets Manager secrets should be rotated within a specified number of days ↗</a>	<p>This control checks whether your secrets have been rotated at least once within 90 days.</p> <p>Rotating secrets can help you to reduce the risk of an unauthorized use of your secrets in your AWS account. Examples include database credentials, passwords, third-party API keys, and even arbitrary text. If you don't change your secrets for a long period of time, the secrets are more likely to be compromised.</p> <p>As more users get access to a secret, it can become more likely that someone mishandled and leaked it to an unauthorized entity. Secrets can be leaked through logs and cache data. They can be shared for debugging purposes and not changed or revoked once the debugging completes. For all these reasons, secrets should be rotated frequently.</p> <p>You can configure your secrets for automatic rotation in AWS Secrets Manager. With automatic rotation, you can replace long-term secrets with short-term ones, significantly reducing the risk of compromise.</p> <p>Security Hub recommends that you enable rotation for your Secrets Manager secrets. To learn more about rotation, see <a href="#">Rotating your AWS Secrets Manager secrets ↗</a> in the AWS Secrets Manager User Guide.</p>	Medium
<a href="#">SNS topics should be encrypted at rest using AWS KMS ↗</a>	<p>This control checks whether an SNS topic is encrypted at rest using AWS KMS.</p> <p>Encrypting data at rest reduces the risk of data stored on disk being accessed by a user not authenticated to AWS. It also adds another set of access controls to limit the ability of unauthorized users to access the data.</p> <p>For example, API permissions are required to decrypt the data before it can be read. SNS topics should be encrypted at-rest for an added layer of security. For more information, see <a href="#">Encryption at rest ↗</a> in the Amazon Simple Notification Service Developer Guide.</p>	Medium
<a href="#">VPC flow logging should be enabled in all VPCs ↗</a>	VPC Flow Logs provide visibility into network traffic that passes through the VPC and can be used to detect anomalous traffic or insight during security events.	Medium

## AWS IdentityAndAccess recommendations

There are 46 AWS recommendations in this category.

Recommendation	Description	Severity

Recommendation	Description	Severity
<a href="#">Amazon Elasticsearch Service domains should be in a VPC</a>	<p>VPC cannot contain domains with a public endpoint.</p> <p>Note: this does not evaluate the VPC subnet routing configuration to determine public reachability.</p>	High
<a href="#">Amazon S3 permissions granted to other AWS accounts in bucket policies should be restricted</a>	<p>Implementing least privilege access is fundamental to reducing security risk and the impact of errors or malicious intent. If an S3 bucket policy allows access from external accounts, it could result in data exfiltration by an insider threat or an attacker. The 'blacklistedactionpatterns' parameter allows for successful evaluation of the rule for S3 buckets. The parameter grants access to external accounts for action patterns that are not included in the 'blacklistedactionpatterns' list.</p>	High
<a href="#">Avoid the use of the "root" account</a>	<p>The "root" account has unrestricted access to all resources in the AWS account. It is highly recommend that the use of this account be avoided.</p> <p>The "root" account is the most privileged AWS account. Minimizing the use of this account and adopting the principle of least privilege for access management will reduce the risk of accidental changes and unintended disclosure of highly privileged credentials.</p>	High
<a href="#">AWS KMS keys should not be unintentionally deleted</a>	<p>This control checks whether KMS keys are scheduled for deletion. The control fails if a KMS key is scheduled for deletion.</p> <p>KMS keys cannot be recovered once deleted. Data encrypted under a KMS key is also permanently unrecoverable if the KMS key is deleted. If meaningful data has been encrypted under a KMS key scheduled for deletion, consider decrypting the data or re-encrypting the data under a new KMS key unless you are intentionally performing a cryptographic erasure.</p> <p>When a KMS key is scheduled for deletion, a mandatory waiting period is enforced to allow time to reverse the deletion, if it was scheduled in error. The default waiting period is 30 days, but it can be reduced to as short as 7 days when the KMS key is scheduled for deletion. During the waiting period, the scheduled deletion can be canceled and the KMS key will not be deleted.</p> <p>For additional information regarding deleting KMS keys, see <a href="#">Deleting KMS keys</a> in the AWS Key Management Service Developer Guide.</p>	High