ANOTHER BRICK IN THE FIREWALL

How to Secure your Azure Data Platform

Grace O’Halloran

INTRODUCTION

- Senior Data Engineering Consultant @ Advancing Analytics
- 6yrs working with Azure Data Platforms
- Microsoft Certified Azure Developer & Administrator

in Grace O'Halloran (grace-o-halloran)

@graceaohalloran

grace@advancinganalytics.co.uk

www.thinkingacloud.co.uk

https://github.com/gracedev94/GraceOH-CommunityContent

ANOTHER BRICK IN THE FIREWALL

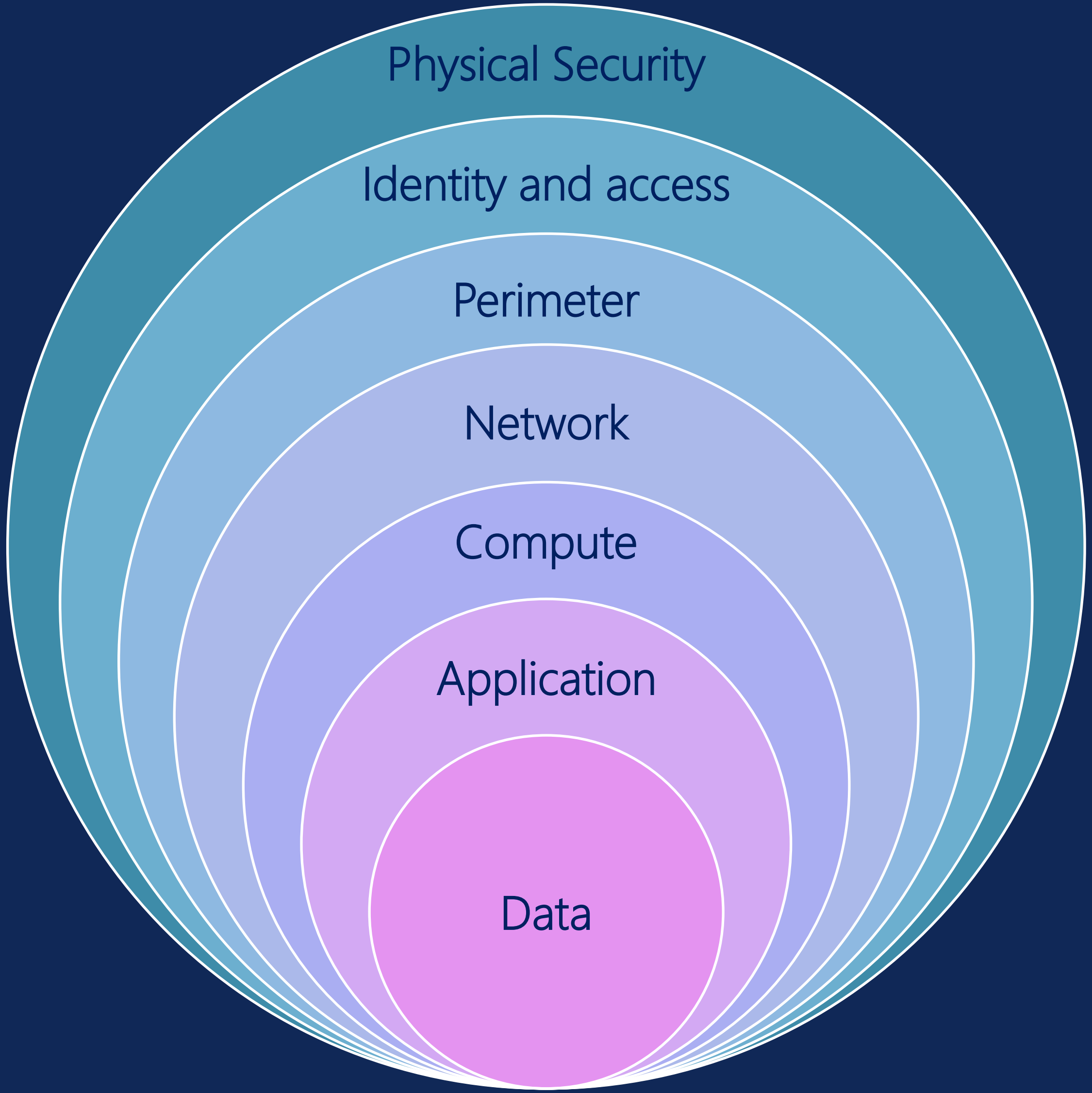
Why care about Network Security	Networks	Ingress & Egress	Azure Private Link	Data Platform Components
• I do data – why is this relevant to me?	• Hub-and-spoke Topology • Address Space Considerations	• Firewalls & UDRs • Network Security Groups • Secure Development Access	• Private Endpoints • Azure Private DNS	• Azure Data Factory: Integration Runtimes • Azure DevOps: Self-hosted Build Agents • Databricks: VNet Injections, SCC, Private Link.

WHY CARE ABOUT NETWORK SECURITY

I do data – why is this relevant to me?

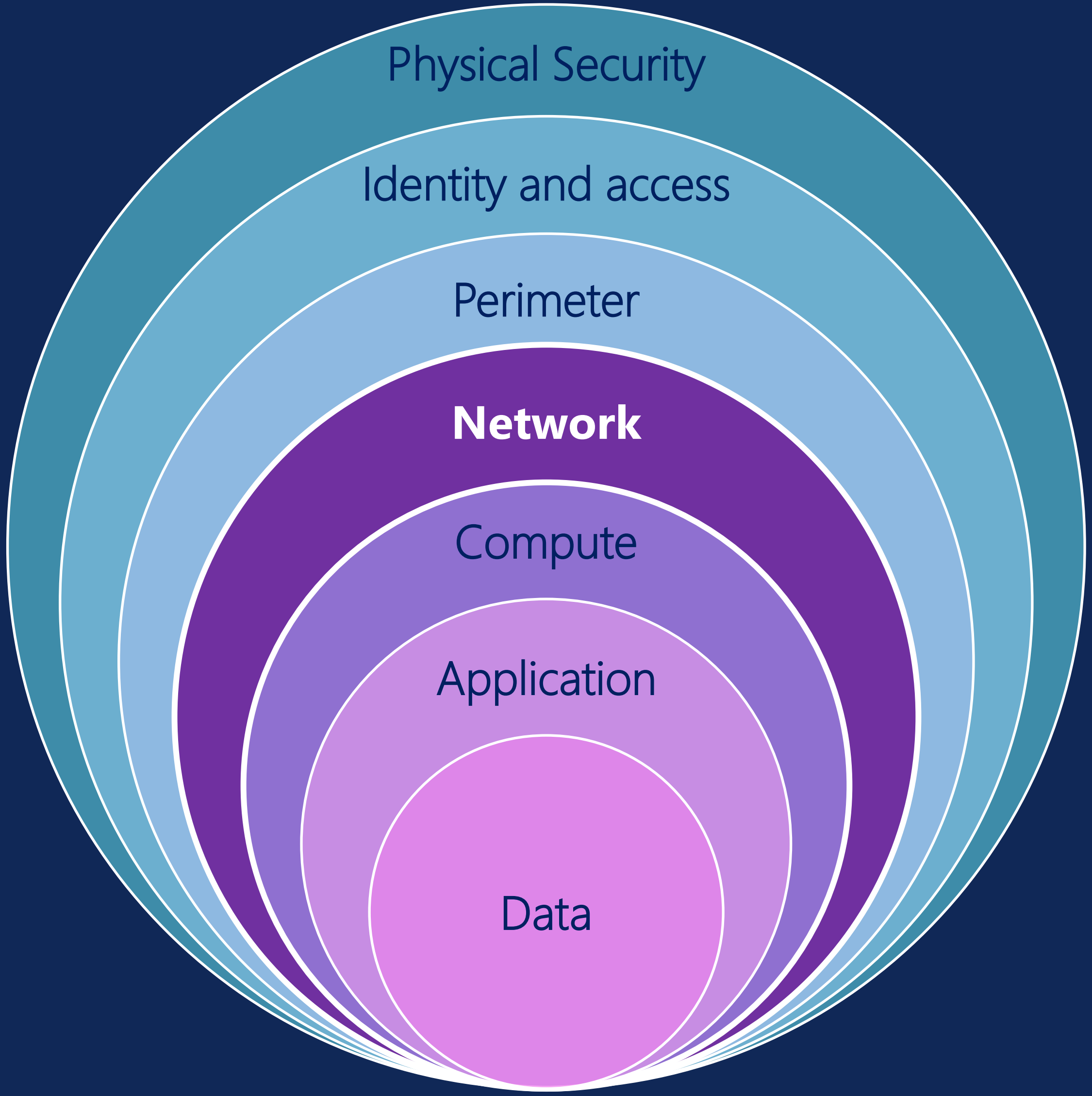
DEFENSE IN DEPTH

- Defense in depth is a cybersecurity approach that uses multiple layers of security measures to protect systems and data.
- Each layer adds a barrier, making it harder for attackers to breach.
- This strategy acknowledges that no single measure is enough and aims to reduce the impact of breaches.



DEFENSE IN DEPTH

- Networking is one of the layers that can be controlled within the Azure environment.
- However, network security of a data platform needs to be paired with other security measures, such as AAD RBAC (Identity and access) and encryption (Data).
- When talking about security and access, it's important to distinguish between the different layers. This presentation will focus on the network layer.

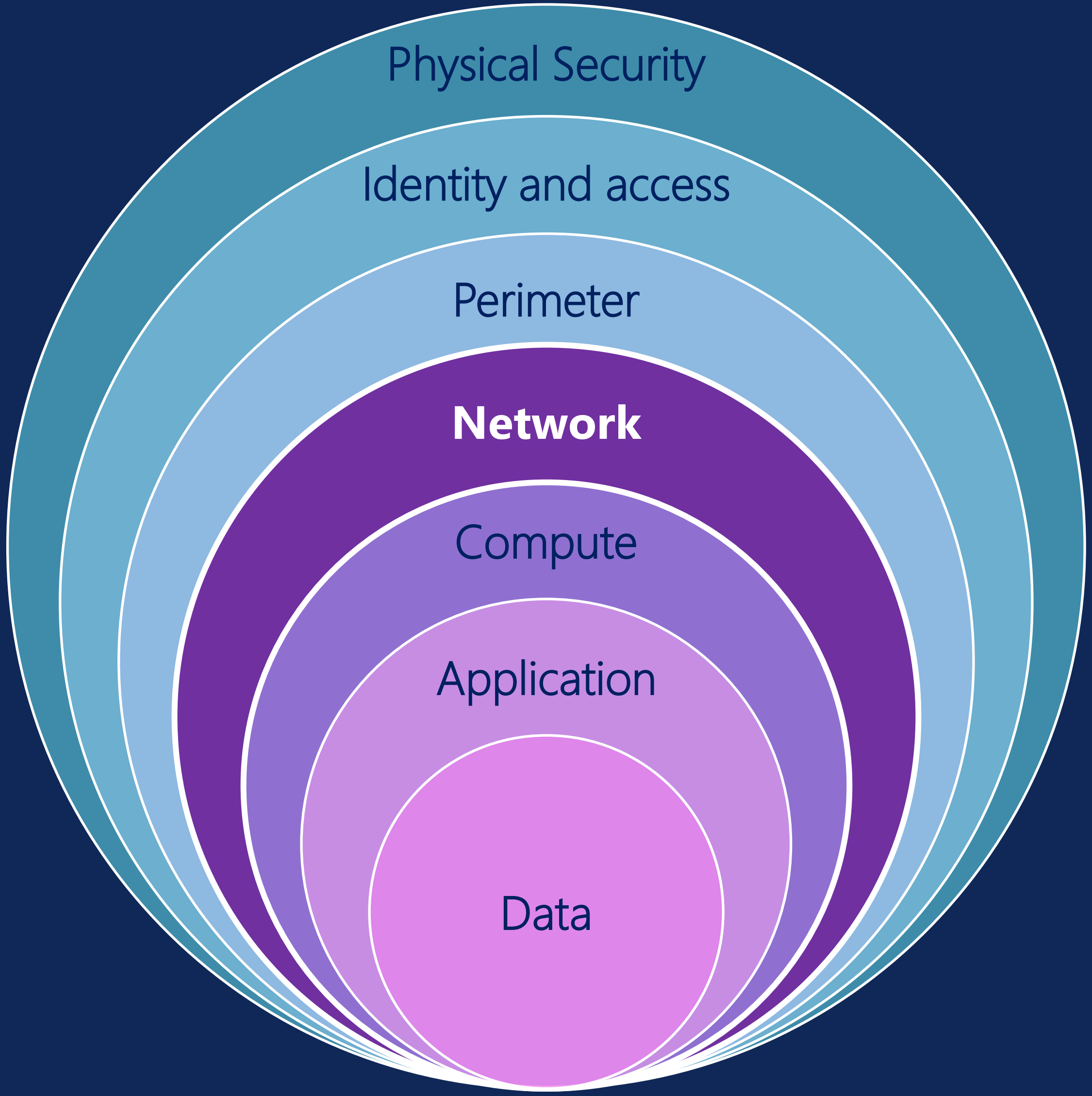


DEFENSE IN DEPTH

- It's important to apply defense in depth to all solutions, including data platforms.

- Many data roles require some level of infrastructure knowledge, especially in the cloud.

- It makes your life much easier if you understand the basics!

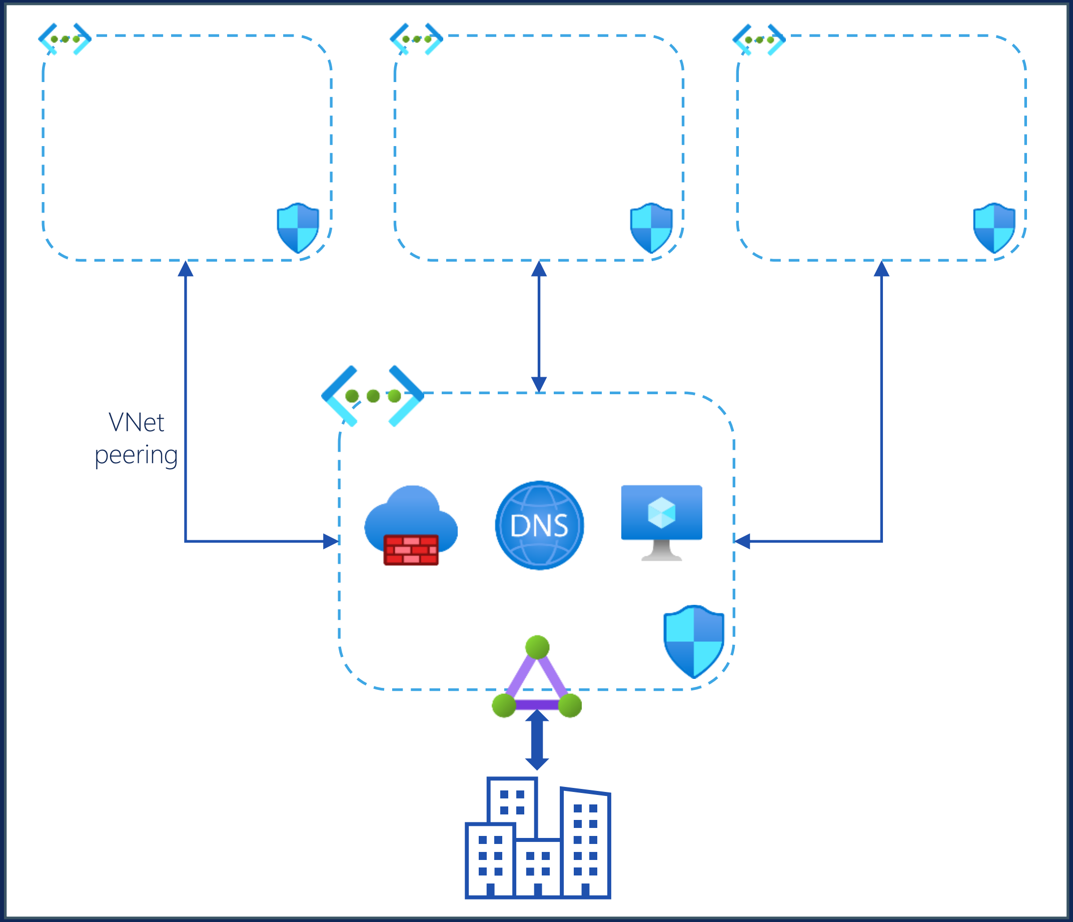


NETWORKS

Hub-and-Spoke Topology

HUB-AND-SPOKE TOPOLOGY

- A hub-and-spoke topology is a network architecture with a central hub connecting to multiple spokes or nodes.
- In Azure, this setup creates a secure network using a central hub as a gateway to manage access to resources and data.
- A data platform would typically reside inside a spoke but will interact with some central infrastructure in the hub.
- This topology is recommended as best practice by Microsoft.



NETWORKS

Address Space Considerations

ADDRESS SPACE CONSIDERATIONS

Avoid Conflicts

Ensure that any address space used for provisioning virtual networks is not already in use within the same network architecture.

Use IPAM

Use centralized IP Address Management to track address space. Ensure any newly provisioned networks are catalogued within IPAM.

Size Requirements

Consider how many components within a spoke will require IP addresses. Specifically consider compute sizing requirements.

Allow for Growth

Always plan for growth within a network. Allow space for scaling of compute to meet changing requirements and allow space for spare subnets.

INGRESS & EGRESS

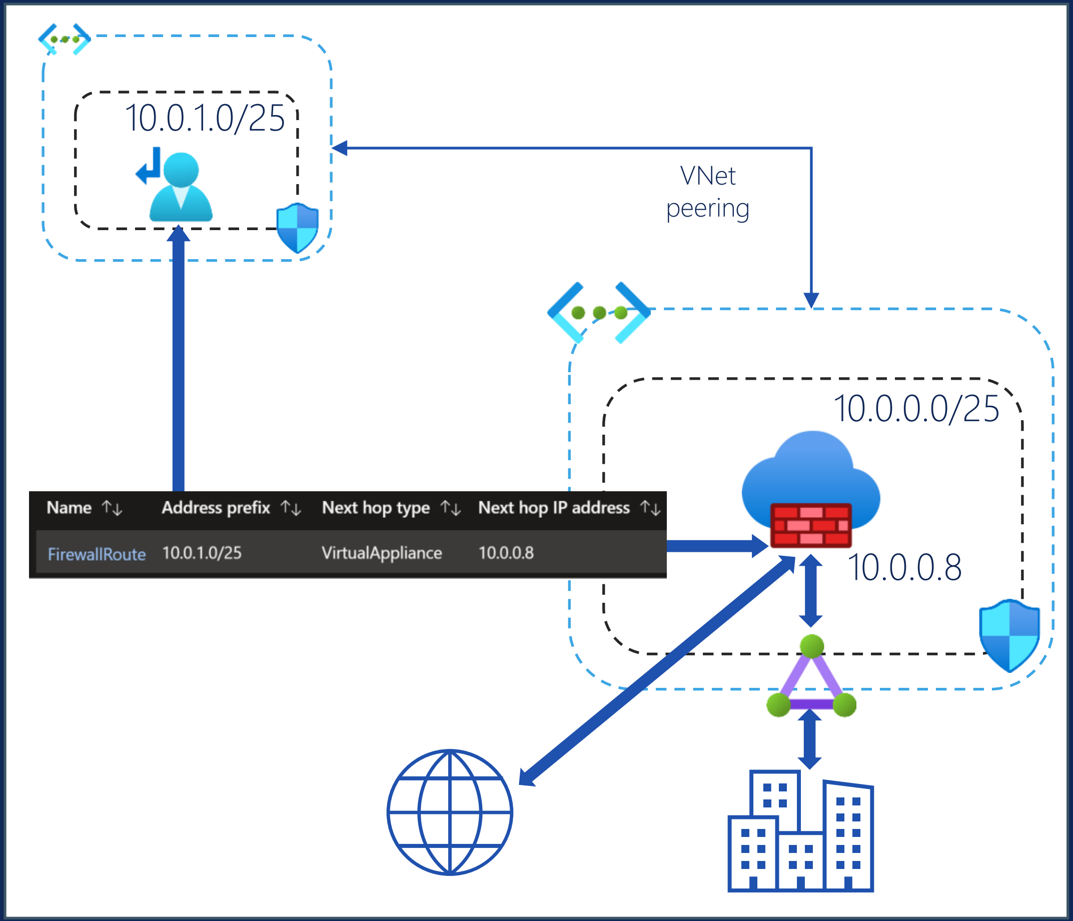
Firewalls & UDRs

"Ingress and Egress are fancy words for Inbound and Outbound."

- Grace O'Halloran, now.

FIREWALLS & USER DEFINED ROUTES

- In a hub-and-spoke topology, all ingress and egress goes through the hub. This provides a level of separation between the spokes and the outside world and minimises the number of entry and exit points in your network.
- It's best practice to have firewall in the hub to monitor, inspect and filter this traffic.
- A central firewall can also be used to inspect and restrict spoke to spoke traffic, when forcing all traffic to go via the hub using User Defined Routes (UDRs).
- UDRs are a powerful tool for completely customising the flow of traffic inside your network, down to subnet level.



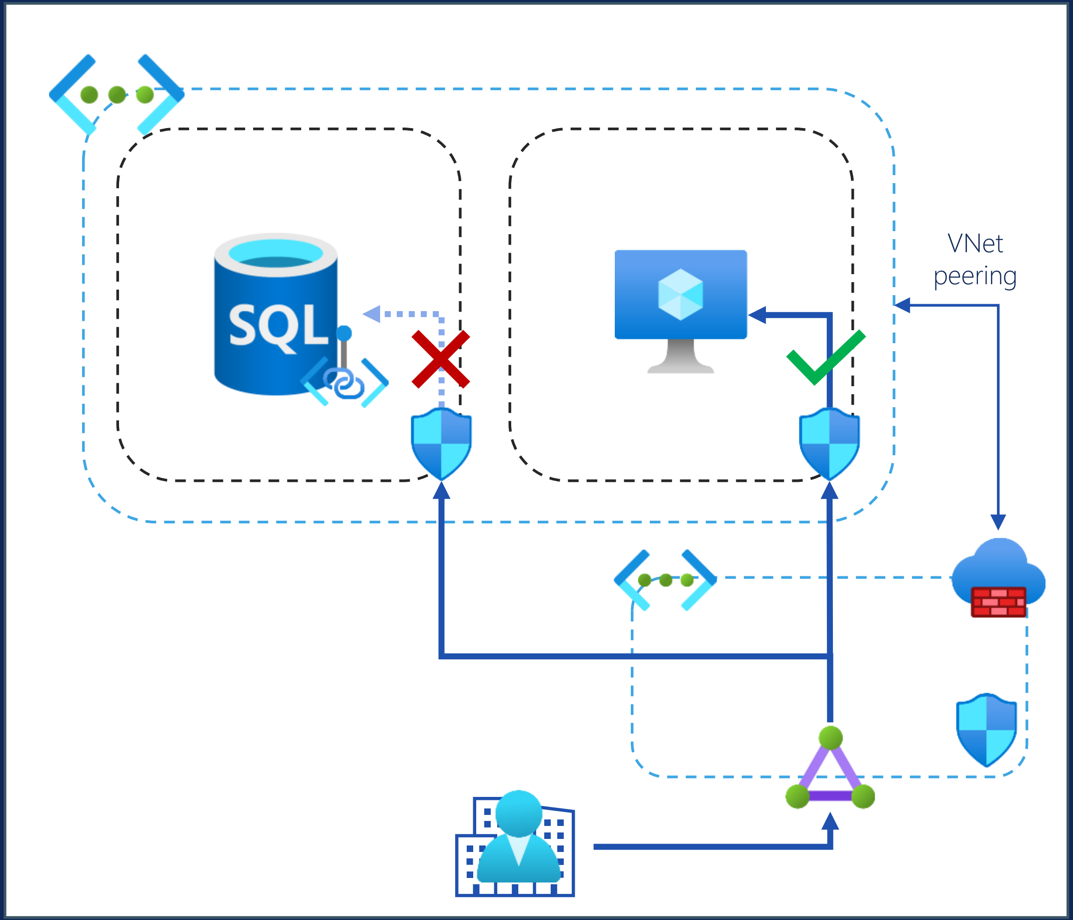
Name ↑↓	Address prefix ↑↓	Next hop type ↑↓	Next hop IP address ↑↓
FirewallRoute	10.0.1.0/25	VirtualAppliance	10.0.0.8

INGRESS & EGRESS

Network Security Groups

NETWORK SECURITY GROUPS

- Network Security Groups (NSGs) are firewalls for your subnets.
- You can specify Allow and Deny inbound and outbound rules for NSGs, based on IP addresses, ports, and protocols.
- Being able to restrict traffic at a subnet level provides even more segmentation for your network.



INGRESS & EGRESS

Secure Development Access

SECURE DEVELOPMENT ACCESS

Microsoft Virtualisation Tool

- Azure Virtual Desktop
- Windows Cloud PC

Third-party Virtualisation Tool

- Citrix
- VMWare

Azure Bastion

- Azure Bastion provides a host for users to securely connect to Azure VMs.

Jump Box

- Infra is responsible for maintaining the security of the jump box.

AZURE PRIVATE LINK

Private Endpoints

AZURE PRIVATE LINK

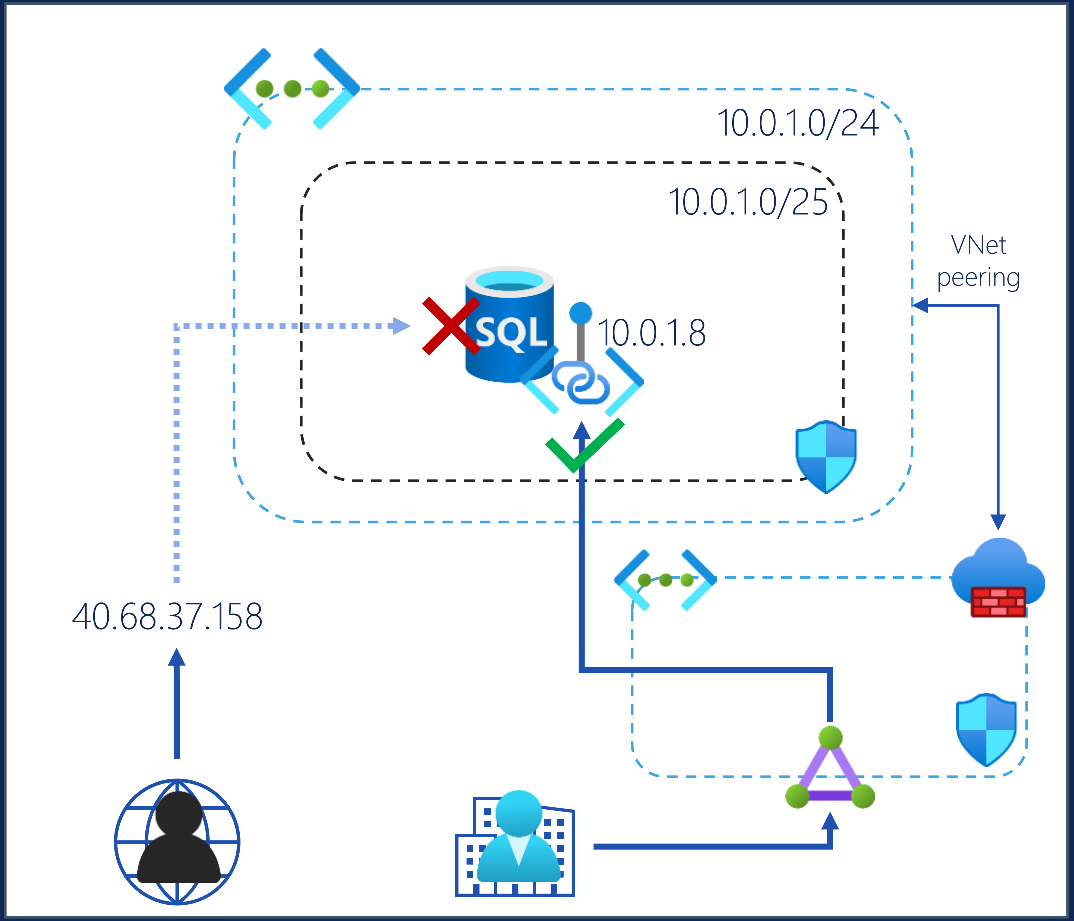
- Azure Private Link is the name of the Microsoft technology which underpins Azure Private Endpoints.

- Azure Private Endpoints can be used to securely access Azure PaaS resources instead of using the default public endpoints.

- Any default public endpoint of a resource (some resources have multiple endpoints) can be replaced by a private endpoint.

- They “bring services into your VNet”, by associating a private IP address from your VNet to the endpoint of your resource.

- Azure Private Endpoints are an Azure resource in their own right, and when deployed a NIC is also created which holds the private IP address.

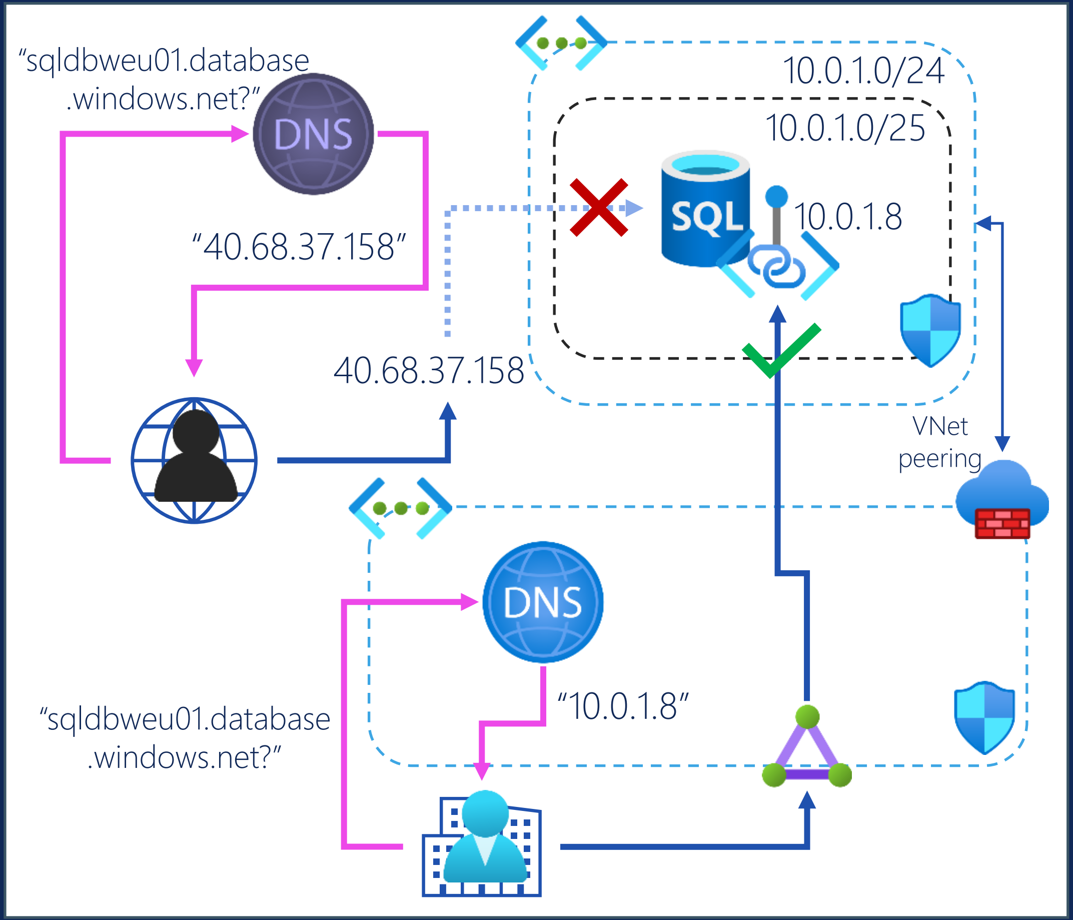


AZURE PRIVATE LINK

Azure Private DNS

AZURE PRIVATE DNS

- Azure Private Link does not work without integration with Azure Private DNS.
- Whilst possible to have a custom DNS solution, the Microsoft recommended route is to use Azure Private DNS Zones, which are an Azure resource.
- Public DNS is like a huge public phonebook that allows us to look up associated public IP addresses with domains. This is a public service hosted on public servers.
- Private DNS Zones are like smaller, private phonebooks, like the contacts list in your phone. These allow those with access to look up associated private IP addresses with domains.



AZURE PRIVATE DNS

- Private DNS Zones should be part of a central DNS solution, meaning only one set of Private DNS Zones are required for an Azure estate.

- One Private DNS Zone resource is required per Azure domain you wish to use Private Endpoints for.

- Private DNS Zones used for Private Link must have specific names which map to the domains they are used for.

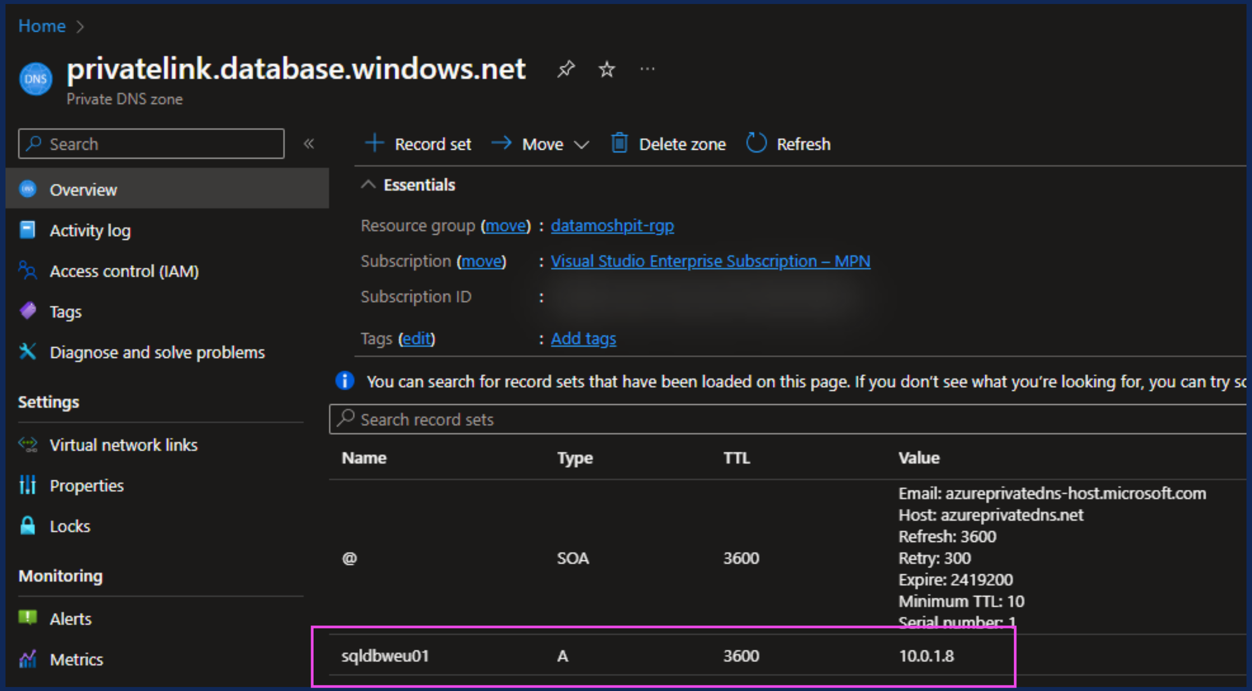
- For example:

Azure SQL Server resource name:	sqldbweu01
Azure SQL Server public endpoint:	sqldbweu01.database.windows.net
Domain:	database.windows.net
Azure Private DNS Zone required:	privatelink.database.windows.net

- Any other SQL Servers which are deployed in the Azure tenant can reuse the same Private DNS Zone, since the domain will be the same.

AZURE PRIVATE DNS

- Private DNS Zones are empty when deployed. They must be populated with A Records; a type of DNS Record which provides the mapping between IP addresses and domains.
- A Records are specific to a particular Private Endpoint.
- Since Private DNS Zones can (and should) be reused, they may contain multiple A Records pertaining to multiple different Private Endpoints.



privatelink.database.windows.net

Private DNS zone

Resource group (move) : datamoshpit-rgp

Subscription (move) : Visual Studio Enterprise Subscription – MPN

Subscription ID : [redacted]

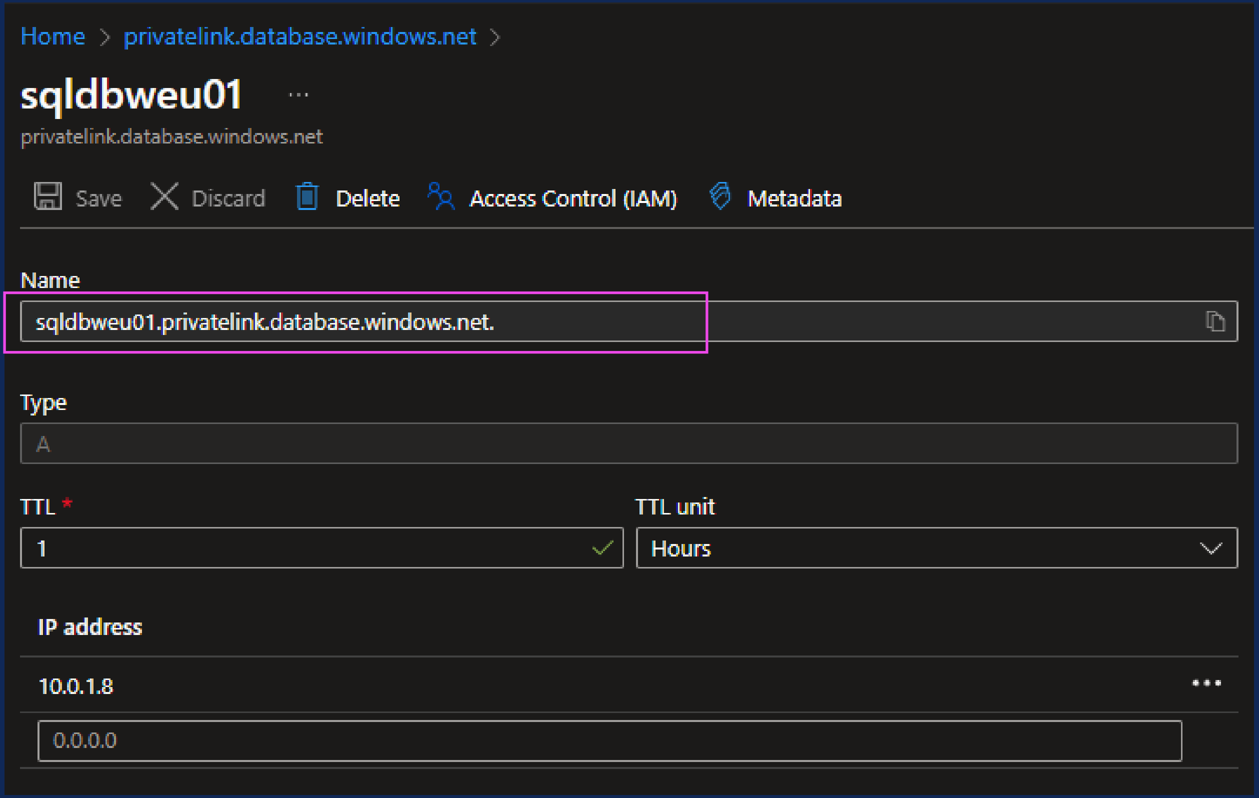
Tags (edit) : Add tags

You can search for record sets that have been loaded on this page. If you don't see what you're looking for, you can try sc

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
sqldbweu01	A	3600	10.0.1.8

AZURE PRIVATE DNS

- Private DNS Zones are empty when deployed. They must be populated with A Records; a type of DNS Record which provides the mapping between IP addresses and domains.
- A Records are specific to a particular Private Endpoint.
- Since Private DNS Zones can (and should) be reused, they may contain multiple A Records pertaining to multiple different Private Endpoints.



sqldbweu01 ...

privatelink.database.windows.net

Name: sqldbweu01.privatelink.database.windows.net.

Type: A

TTL *: 1

TTL unit: Hours

IP address

10.0.1.8

0.0.0.0

DATA PLATFORM COMPONENTS

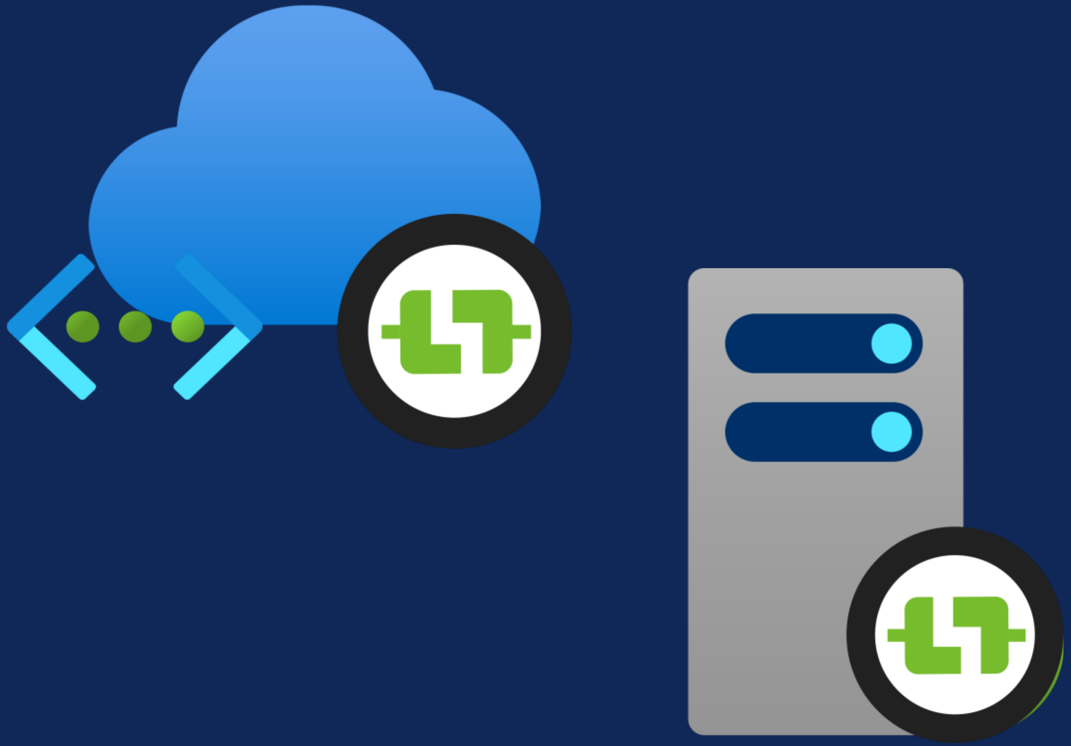
Azure Data Factory: Integration Runtimes

ADF: INTEGRATION RUNTIMES

- When inside a private network, you must consider how your compute elements will have access to your resources.

- The compute used in Azure Data Factory (ADF) is the Integration Runtime (IR). The default Azure hosted IR will not have access to your privately secured resources.

- There are two options:
  1. Azure IR with Managed VNet
  2. Self-hosted IR (SHIR)



ADF: INTEGRATION RUNTIMES

Azure IR with Managed VNet

Use the Azure-hosted IR with the Managed VNet enabled in order to secure the compute inside a private network.

You must use Managed Private Endpoints to allow your IR access to your protected resources.

Pros

- Fully managed and serverless
- Elastic scaling
- No maintaining of firewall rules

Cons

- No control over address space
- Requires additional private endpoints
- Can increase cost
- Doesn’t work easily with on-prem connectivity

Self-hosted IR (SHIR)

SHIRs are created by installing an IR application on your own machine, this can be an on-prem server or an Azure VM.

The SHIR server will utilise existing Private Endpoints to securely connect to your protected resources.

Pros

- High Availability options
- Runtime costs are cheaper
- Allows for easy connectivity to on-prem data sources

Cons

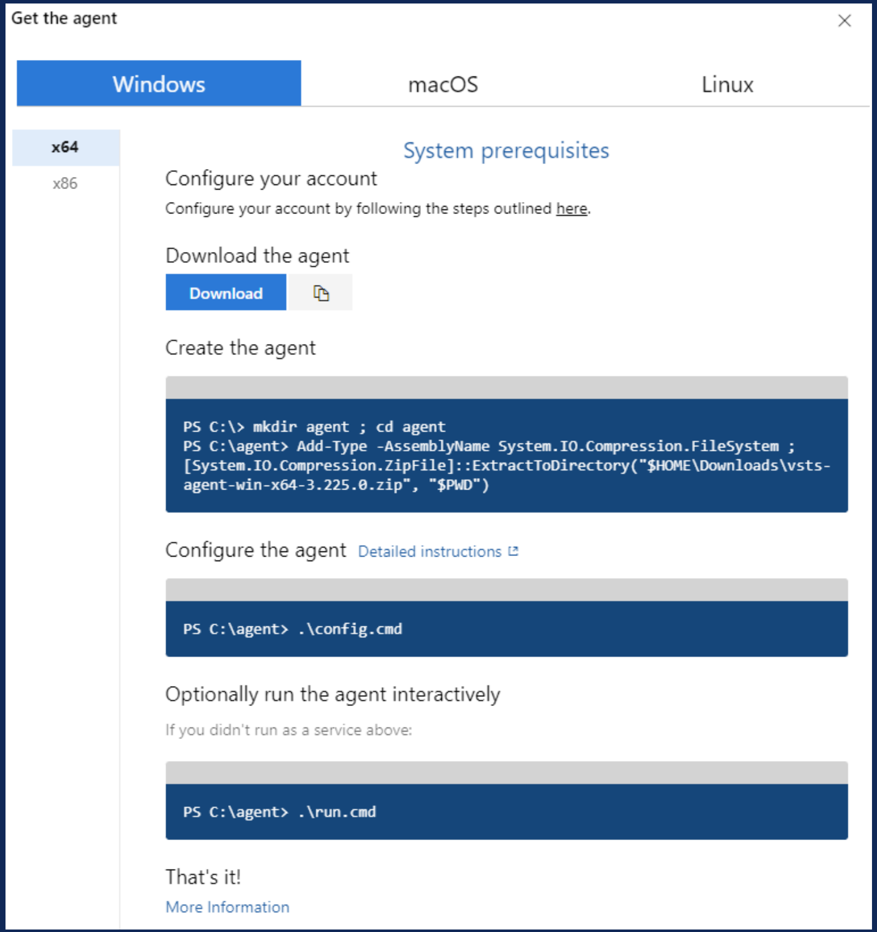
- Requires pre-existing infrastructure
- Responsible for providing and maintaining the server
- Maintenance of firewall rules
- Pay for compute resource

DATA PLATFORM COMPONENTS

Azure DevOps: Self-hosted Build Agents

AZURE DEVOPS: BUILD AGENTS

- When your platform is inside a private network, you must consider your CICD compute.
- In Azure DevOps, you cannot use the default Microsoft-hosted build agent, as it will not have access to your endpoints.
- You must instead use a Self-hosted Build Agent. Similar to the Self-hosted Integration Runtime, you install the build agent on your own machine.



System prerequisites

Configure your account

Configure your account by following the steps outlined here.

Download the agent

Create the agent

```
PS C:\> mkdir agent ; cd agent
PS C:\agent> Add-Type -AssemblyName System.IO.Compression.FileSystem ;
[System.IO.Compression.ZipFile]::ExtractToDirectory("$HOME\Downloads\vsts-
agent-win-x64-3.225.0.zip", "$PWD")
```

Configure the agent Detailed instructions

```
PS C:\agent> .\config.cmd
```

Optionally run the agent interactively

If you didn't run as a service above:

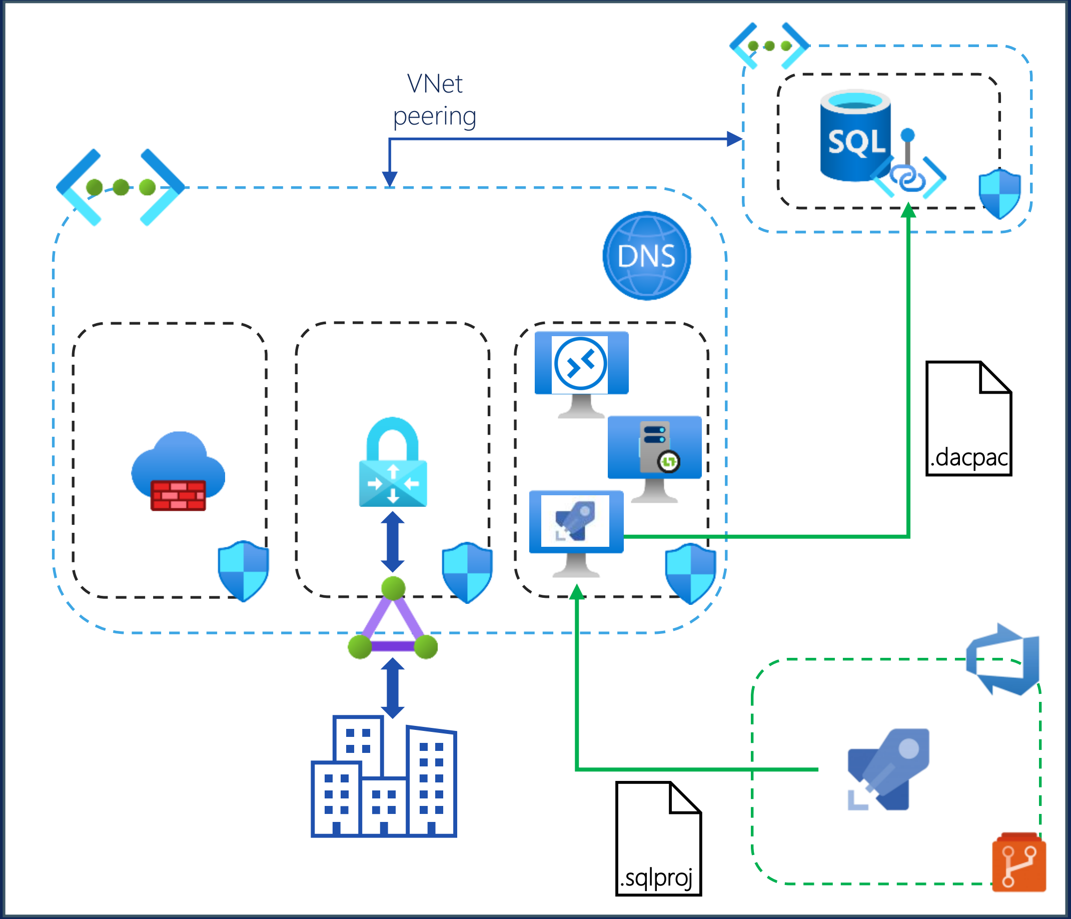
```
PS C:\agent> .\run.cmd
```

That's it!

More Information

AZURE DEVOPS: BUILD AGENTS

- You are responsible for ensuring network connectivity between your Build Agent server and your platform's endpoints.
- It's not uncommon to have a central collection of Azure VMs in the hub used for things such as DevOps Build Agents.



DATA PLATFORM COMPONENTS

Databricks: VNet Injection, Secure Cluster Connectivity, Private Link.

DATABRICKS

Managed VNet

Clusters are protected by a VNet but is managed by Databricks.

VNet Injection

Clusters are protected by your own VNet.

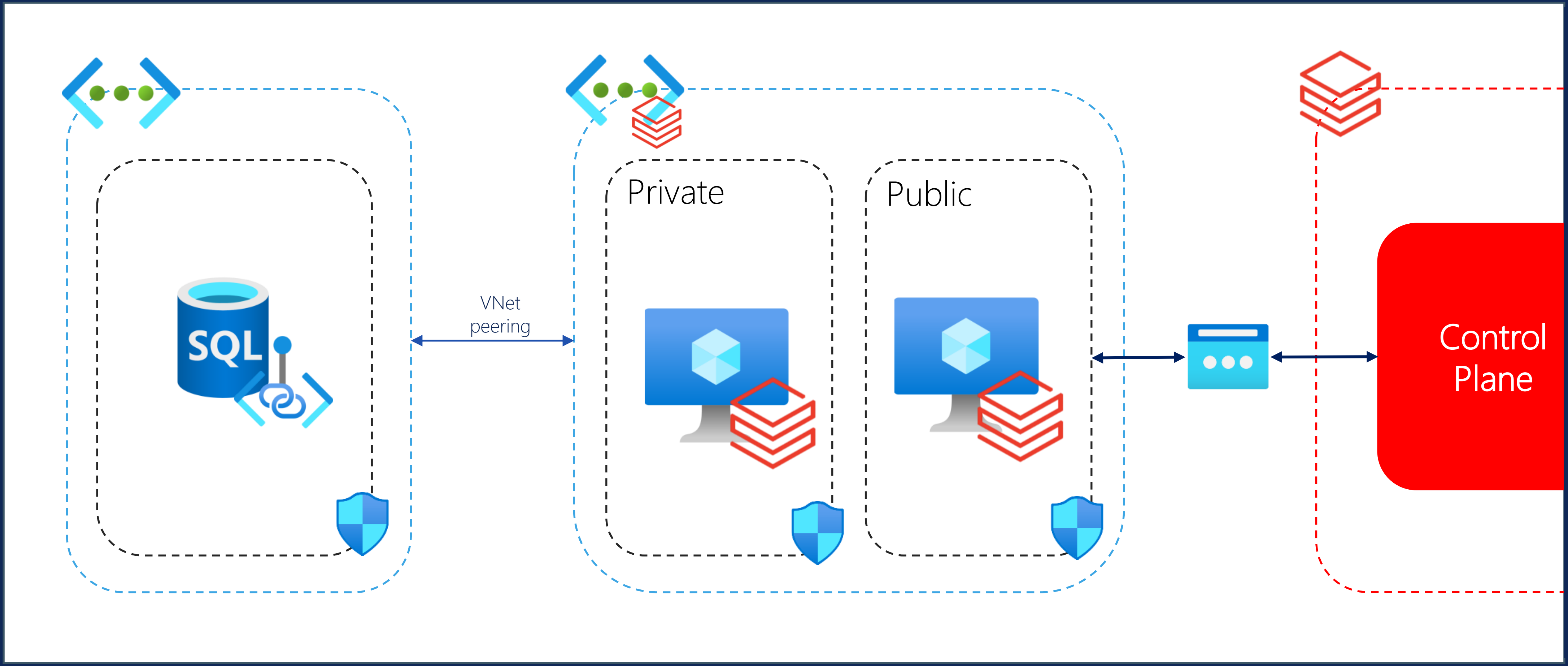
Secure Cluster Connectivity

Also known as "No Public IP". Clusters have no public IP addresses and your VNet (data plane) has no open ports.

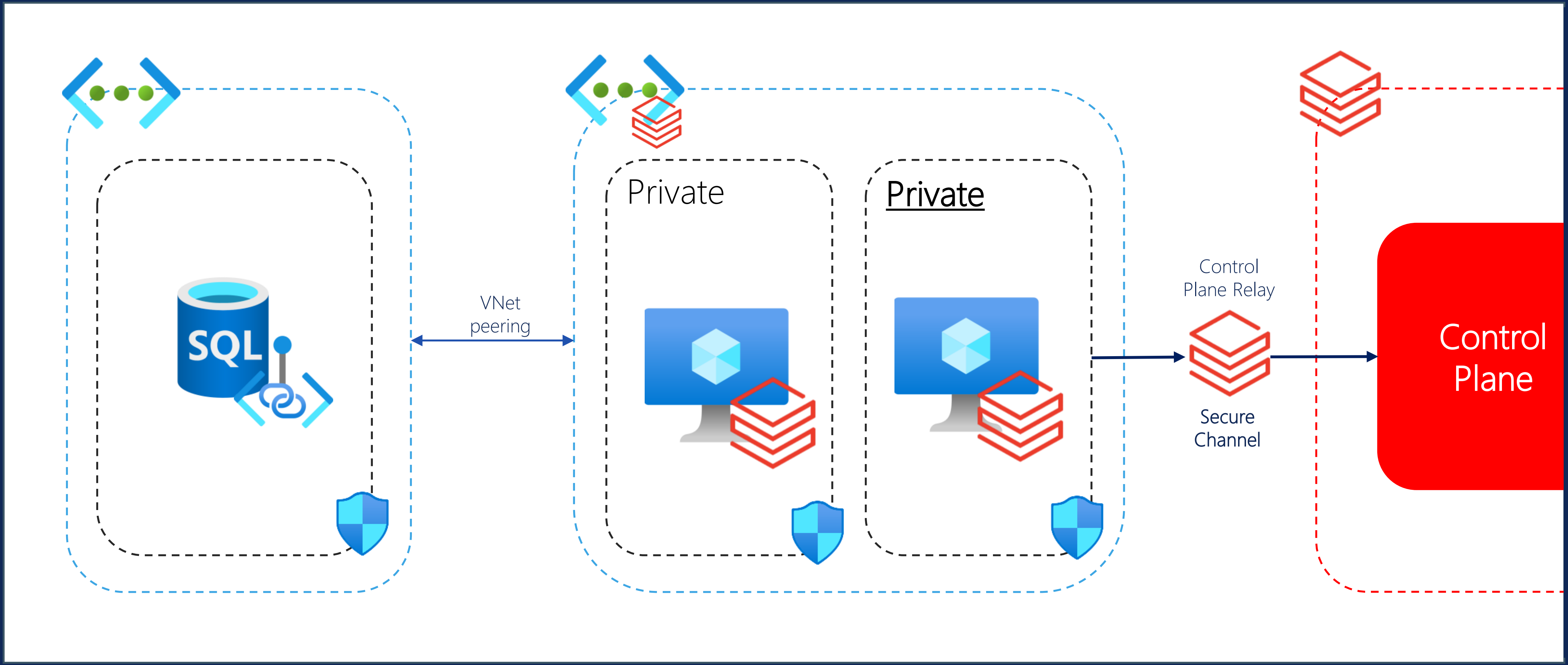
Private Link

Traffic between the clusters and Databricks control plane stays private, instead of traversing the Microsoft backbone. Also protects front-end connectivity.

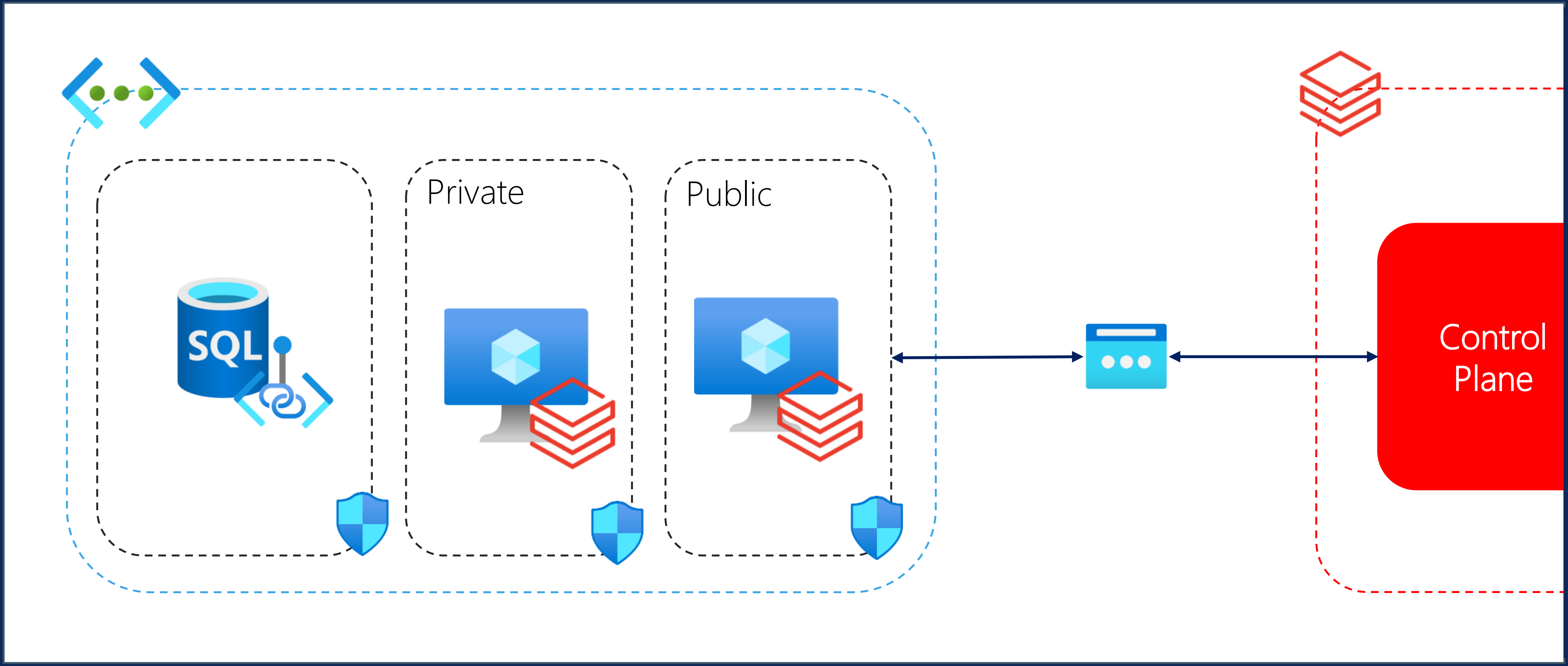
DATABRICKS: MANAGED VNET



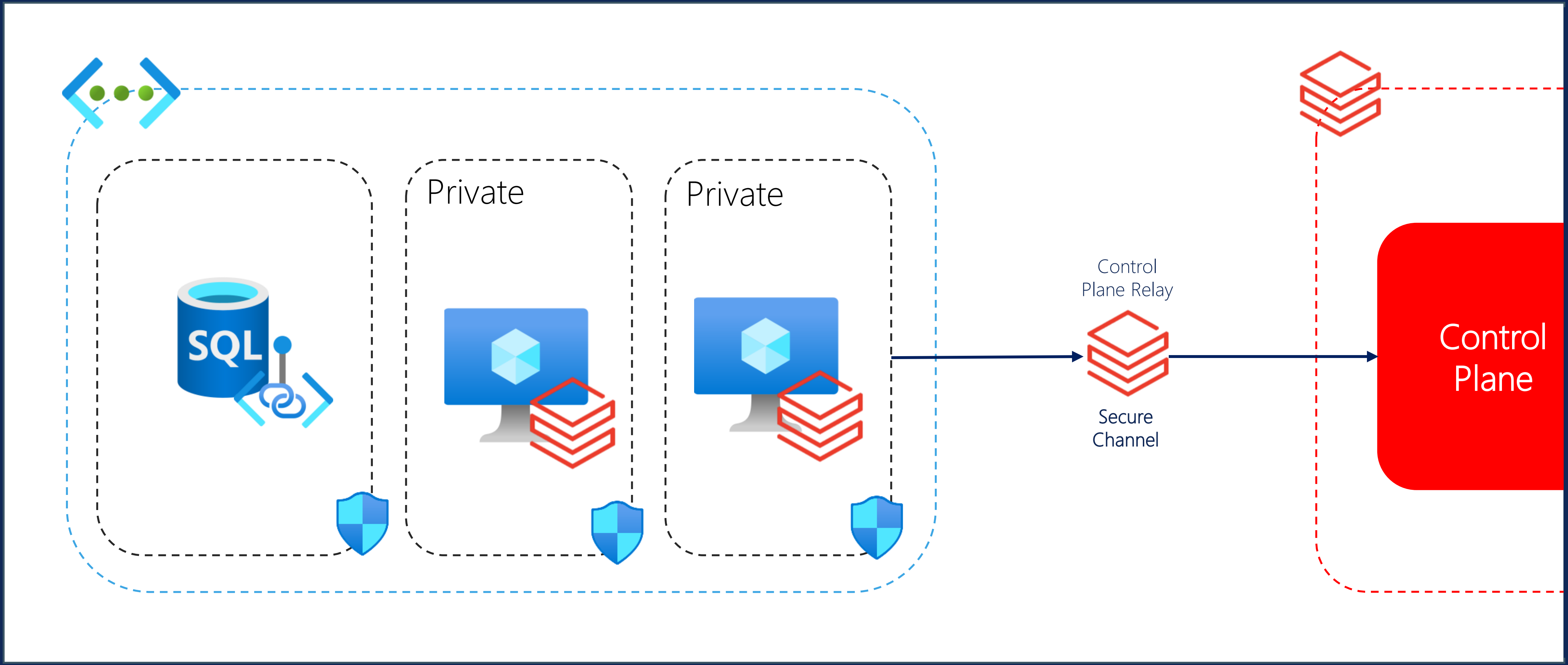
DATABRICKS: MANAGED VNET & SCC



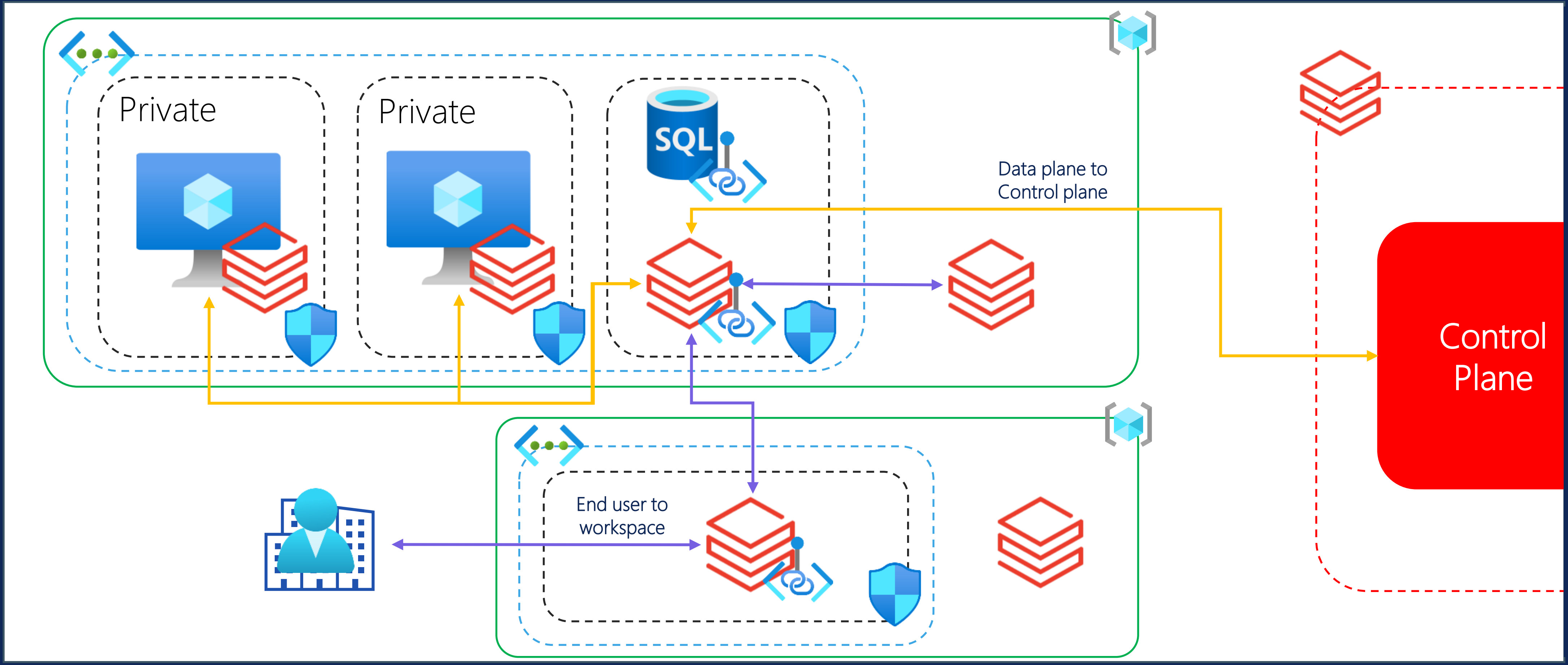
DATABRICKS: VNET INJECTION



DATABRICKS: VNET INJECTION & SCC



DATABRICKS: VNET INJECTION & PRIVATE LINK



THANK YOU

Any questions?

Grace O'Halloran (grace-o-halloran)

@graceaohalloran

grace@advancinganalytics.co.uk

www.thinkingacloud.co.uk

https://github.com/gracedev94/GraceOH-CommunityContent

https://www.youtube.com/watch?v=FdmE82BloS4