

Deployment

Below is a high-level guide to deploy the AI Hub Gateway Landing Zone main components.

Networking

For the AI Hub Gateway Landing Zone to be deployed, you will need to have/identify the following components:

- **Virtual network & subnet:** A virtual network to host the AI Hub Gateway Landing Zone.
 - APIM to be deployed in **internal mode** requires a subnet with /27 or larger with NSG that allows the critical rules.
 - **Private endpoints subnet(s):** Private endpoints for the AI services to be exposed through the AI Hub Gateway Landing Zone. Usually a /27 or larger subnet would be sufficient.
- **Private DNS zone:** A private DNS zone to resolve the private endpoints.
 - Internal APIM relies on **private DNS** to resolve the APIM endpoints, so a Azure Private DNS zone or other DNS solution is required.
 - **Private endpoints DNS zone:** A private DNS zone to resolve the private endpoints for the connected Azure AI services.
- **ExpressRoute or VPN:** If you are planning to connect to on-premises or other clouds, you will need to have an ExpressRoute or VPN connection.
- **DMZ appliances:** If you are planning to expose backend and gateway services on the internet, you need to have a Web Application Firewall (like Azure Front Door & Application Gateway) and network firewall (like Azure Firewall) to govern both ingress and egress traffic.

Azure API Management (APIM)

APIM is the central component of the AI Hub Gateway Landing Zone.

Recommended deployment of APIM to be in **internal mode** to ensure that the gateway is not exposed to the internet and to ensure that the gateway is only accessible through the private network.

internal mode requires a subnet with /27 or larger with NSG that allows the critical rules in addition to management public IP (with DNS label set)

This is a great starting point to deploy APIM in internal mode: [Deploy Azure API Management in internal mode](#)

Application Insights

Application Insights is a critical component of the AI Hub Gateway Landing Zone, and it is used to monitor the operational performance of the gateway.

To deploy Application Insights, you can use the following guide: [How to integrate Azure API Management with Azure Application Insights](#)

Event Hub

Event Hub is used to stream usage and charge-back data to target data and charge back platforms.

To deploy Event Hub, you can use the following guide: [Logging with Event Hub]<https://azure.github.io/apim-lab/apim-lab/6-analytics-monitoring/analytics-monitoring-6-3-event-hub.html>)

Deployment summary

When deployment of primary components is completed, you will have the following components deployed:

- **Azure API Management**
- **Application Insights**
- **Event Hub**

Network wiring also will be established to allow the gateway to access the AI services through private endpoints, internet access through DMZ appliances and backend systems through private network.