What is Azure Kubernetes Service?
Article
03/10/2023
5 minutes to read
38 contributors
In this article
Access, security, and monitoring
Clusters and nodes
Virtual networks and ingress
Development tooling integration
Show 4 more
Azure Kubernetes Service (AKS) simplifies deploying a managed Kubernetes cluster in Azure by offloading the operational overhead to Azure. As a hosted Kubernetes service, Azure handles critical tasks, like health monitoring and maintenance. When you create an AKS cluster, a control plane is automatically created and configured. This control plane is provided at no cost as a managed Azure resource abstracted from the user. You only pay for and manage the nodes attached to the AKS cluster.

You can create an AKS cluster using:

Azure CLI
Azure PowerShell
Azure portal
Template-driven deployment options, like Azure Resource Manager templates, Bicep, and Terraform.
When you deploy an AKS cluster, you specify the number and size of the nodes, and AKS deploys and configures the Kubernetes control plane and nodes. Advanced networking, Azure Active Directory (Azure AD) integration, monitoring, and other features can be configured during the deployment process.

For more information on Kubernetes basics, see Kubernetes core concepts for AKS.

Note

This service supports Azure Lighthouse, which lets service providers sign in to their own tenant to manage subscriptions and resource groups that customers have delegated.

AKS also supports Windows Server containers.

Access, security, and monitoring
For improved security and management, you can integrate with Azure AD to:

Use Kubernetes role-based access control (Kubernetes RBAC).
Monitor the health of your cluster and resources.
Identity and security management
Kubernetes RBAC
To limit access to cluster resources, AKS supports Kubernetes RBAC. Kubernetes RBAC controls access and permissions to Kubernetes resources and namespaces.

Azure AD
You can configure an AKS cluster to integrate with Azure AD. With Azure AD integration, you can set up Kubernetes access based on existing identity and group membership. Your existing Azure AD users and groups can be provided with an integrated sign-on experience and access to AKS resources.

For more information on identity, see Access and identity options for AKS.

To secure your AKS clusters, see Integrate Azure AD with AKS.

Integrated logging and monitoring
Azure Monitor for Container Health collects memory and processor performance metrics from containers, nodes, and controllers within your AKS clusters and deployed applications. You can review both container logs and the Kubernetes logs, which are:

Stored in an Azure Log Analytics workspace.
Available through the Azure portal, Azure CLI, or a REST endpoint.
For more information, see Monitor AKS container health.

Clusters and nodes
AKS nodes run on Azure virtual machines (VMs). With AKS nodes, you can connect storage to nodes and pods, upgrade cluster components, and use GPUs. AKS supports Kubernetes clusters that run multiple node pools to support mixed operating systems and Windows Server containers.

For more information about Kubernetes cluster, node, and node pool capabilities, see Kubernetes core concepts for AKS.

Cluster node and pod scaling
As demand for resources change, the number of cluster nodes or pods that run your services automatically scales up or down. You can adjust both the horizontal pod autoscaler or the cluster autoscaler to adjust to demands and only run necessary resources.

For more information, see Scale an AKS cluster.

Cluster node upgrades
AKS offers multiple Kubernetes versions. As new versions become available in AKS, you can upgrade your cluster using the Azure portal, Azure CLI, or Azure PowerShell. During the upgrade process, nodes are carefully cordoned and drained to minimize disruption to running applications.

To learn more about lifecycle versions, see Supported Kubernetes versions in AKS. For steps on how to upgrade, see Upgrade an AKS cluster.

GPU-enabled nodes
AKS supports the creation of GPU-enabled node pools. Azure currently provides single or multiple GPU-enabled VMs. GPU-enabled VMs are designed for compute-intensive, graphics-intensive, and visualization workloads.

For more information, see Using GPUs on AKS.

Confidential computing nodes (public preview)
AKS supports the creation of Intel SGX-based, confidential computing node pools (DCSv2 VMs). Confidential computing nodes allow containers to run in a hardware-based, trusted execution environment (enclaves). Isolation between containers, combined with code integrity through attestation, can help with your defense-in-depth container security strategy. Confidential computing nodes support both confidential containers (existing Docker apps) and enclave-aware containers.

For more information, see Confidential computing nodes on AKS.

Mariner nodes
Mariner is an open-source Linux distribution created by Microsoft, and it’s now available for preview as a container host on Azure Kubernetes Service (AKS). The Mariner

container host provides reliability and consistency from cloud to edge across the AKS, AKS-HCI, and Arc products. You can deploy Mariner node pools in a new cluster, add Mariner node pools to your existing Ubuntu clusters, or migrate your Ubuntu nodes to Mariner nodes.

For more information, see Use the Mariner container host on AKS

Storage volume support
To support application workloads, you can mount static or dynamic storage volumes for persistent data. Depending on the number of connected pods expected to share the storage volumes, you can use storage backed by:

Azure Disks for single pod access
Azure Files for multiple, concurrent pod access.
For more information, see Storage options for applications in AKS.

Virtual networks and ingress
An AKS cluster can be deployed into an existing virtual network. In this configuration, every pod in the cluster is assigned an IP address in the virtual network and can directly communicate with other pods in the cluster and other nodes in the virtual network.

Pods can also connect to other services in a peered virtual network and on-premises networks over ExpressRoute or site-to-site (S2S) VPN connections.

For more information, see the Network concepts for applications in AKS.

Ingress with HTTP application routing
The HTTP application routing add-on helps you easily access applications deployed to your AKS cluster. When enabled, the HTTP application routing solution configures an ingress controller in your AKS cluster.

As applications are deployed, publicly accessible DNS names are auto-configured. The HTTP application routing sets up a DNS zone and integrates it with the AKS cluster. You can then deploy Kubernetes ingress resources as normal.

To get started with Ingress traffic, see HTTP application routing.

Development tooling integration

Kubernetes has a rich ecosystem of development and management tools that work seamlessly with AKS. These tools include Helm and the Kubernetes extension for Visual Studio Code.

Azure provides several tools that help streamline Kubernetes.

Docker image support and private container registry
AKS supports the Docker image format. For private storage of your Docker images, you can integrate AKS with Azure Container Registry (ACR).

To create a private image store, see Azure Container Registry.

Kubernetes certification
AKS has been CNCF-certified as Kubernetes conformant.

Regulatory compliance
AKS is compliant with SOC, ISO, PCI DSS, and HIPAA. For more information, see Overview of Microsoft Azure compliance.

Next steps
Learn more about deploying and managing AKS.

Feedback