Enabling private CDW environment using Azure Kubernetes Service (Preview)

Date published: 2021-05-03
Date modified: 2021-07-29

Legal Notice

Contents

Overview

Azure Kubernetes Service (AKS) simplifies container-based application deployment and management. When you create an AKS cluster, a control plane is automatically created and configured, and the Azure platform configures the secure communication between the control plane and compute nodes. In a private cluster, the API server for the control plane has an internal IP address. The Azure nodes and the Kubernetes control plane components do not have publicly routable IP addresses. When you deploy a private cluster, the network traffic between your API server and your node pools remains only on the private network.

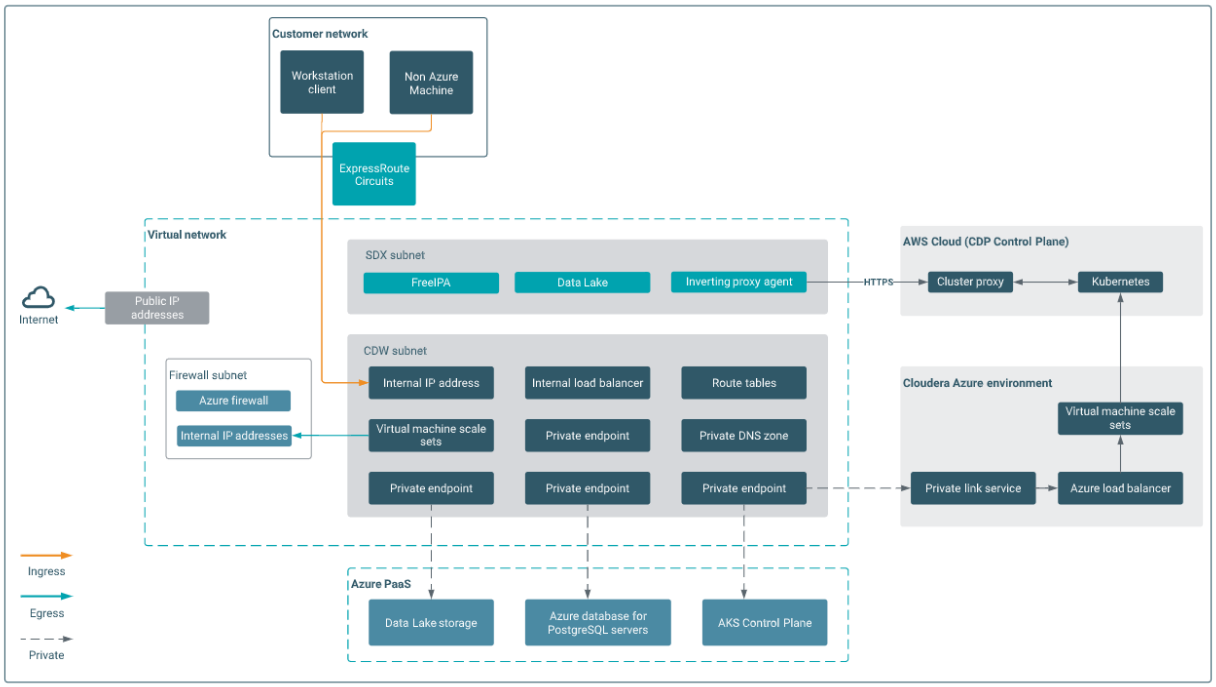
Cluster Connectivity Manager (CCM) {*version 2*} enables the CDP Control Plane to communicate with the Kubernetes control plane using an inverting proxy solution. Cloudera Data Warehouse (CDW) can communicate with the Kubernetes control plane and the other resources, such as virtual machines deployed in your network, by using a special established channel.

Large enterprises have a requirement to create private CDW environments in the public cloud so that the cloud environment is similar to the on-premises one, in terms of security and networking. For example, it is frequently not permissible to allow any public access to your environment, resources, and applications deployed in your cloud environment.

A private environment has the following criteria:

- Kubernetes API server for the control plane must have private API endpoints
- Hue, DAS, JDBC, or ODBC applications can only be accessed through private endpoints
- Egress Kubernetes traffic must be controlled using a transparent proxy
- Storage, SQL database, and other resources can only be accessed through private endpoints
- CDP Control Plane can only be accessed through private endpoints using Azure Private Link Service
- Ability to configure a DNS zone in your environment

Below is a typical network layout for private environments:



Setting up the environment for private cluster deployment

Before you create a private AKS cluster, you must complete the following:

1. Create a resource group for CDP from the Microsoft Azure portal.
   The resource group provides a management layer that enables you to create, update, and delete resources in your Azure account. For more information, see Create resource groups.

2. (Optional) Create a private storage account and network access rules to block all internet traffic. See Create a storage account with a private endpoint.
   To configure network access rules, see Change the default network access rule.
   The private endpoints allow clients on a virtual network (VNET) to securely access data over a Private Link.

3. Create a VNET and a subnet.
   VNET is the fundamental building block for your private network in Azure. It enables Azure resources, such as Azure Virtual Machines, to securely communicate with each other and with the internet.
   For more information, see Create a virtual network using the Azure portal and Add a subnet.

4. (Optional) Configure the CDP Control Plane Private Link service.
   To ensure that all egress traffic including CDP Control Plane traffic travels only through private networks, you must configure CCM (v2) and CDP Control Plane private endpoints properly for the selected VNET.
   Contact your Cloudera account representative to register the CDP Control Plane Private Link service endpoints in your Azure VNET.

5. Configure custom DNS on the VNET to resolve Azure Private DNS zones.
   To resolve private endpoint DNS records, the VNET DNS servers must be capable of resolving Azure DNS records.

6. Disable network endpoint policies for private endpoints and Azure Private Link Service.
   For more information, see Disable network policies for private endpoints and Disable network policies for Private Link service source IP.

7. Configure the following firewall exceptions for CDW and AKS on the egress firewall, and for storage account endpoints, if the account is not Private-Link enabled:
   Outbound network access destinations
   Restrict egress traffic in AKS

8. Configure user-defined routing (UDR) on the VNET to forward all traffic to an egress firewall and link it to the subnet.
   For more information on using a user-defined route for egress, see Customize cluster egress with a User-Defined Route.

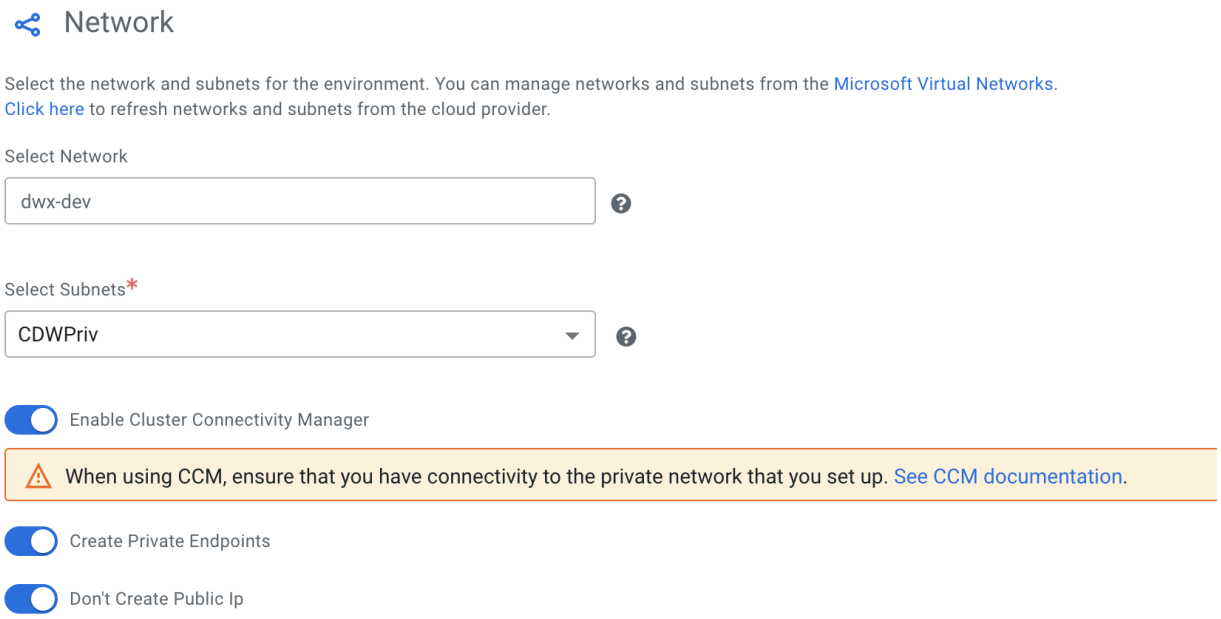
9. Create a CDP Azure environment in the VNET that you created earlier.
   For more information, see Working with Azure environments.

   Select private environment options for the PostgreSQL database, virtual machines, and CCM (v2) while registering the Azure environment in CDP.
   For more information, see Register an Azure environment.

   On the environment registration screen, the **Enable Cluster Connectivity Manager** option is enabled by default. It ensures that all traffic from Cloudera Control Plane to your cloud resources travels through a secured HTTPS tunnel. CDW Private AKS only works with CCM (v2).

   Do not create public IPs so that Azure VMs have private IP addresses only.

   Enable the **Create Private Endpoints** option. By default, the PostgreSQL Azure database provisioned for your Data Lake is reachable through a service endpoint (public IP address). To increase security, you must select to have it reachable through a private endpoint instead of a service endpoint.
   For more information, see Enabling private endpoint for PostgreSQL on Azure.



Activating private AKS in CDW

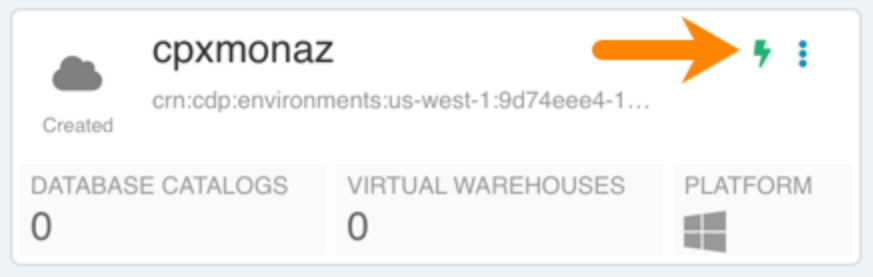
Activating an environment with the CDW service sets up the Kubernetes cluster, which provides the compute resources for the Database Catalog. In addition, activating an environment enables the CDW service to use the existing Data Lake that was set up for the environment, including all data, metadata, and security policies.

Before you begin

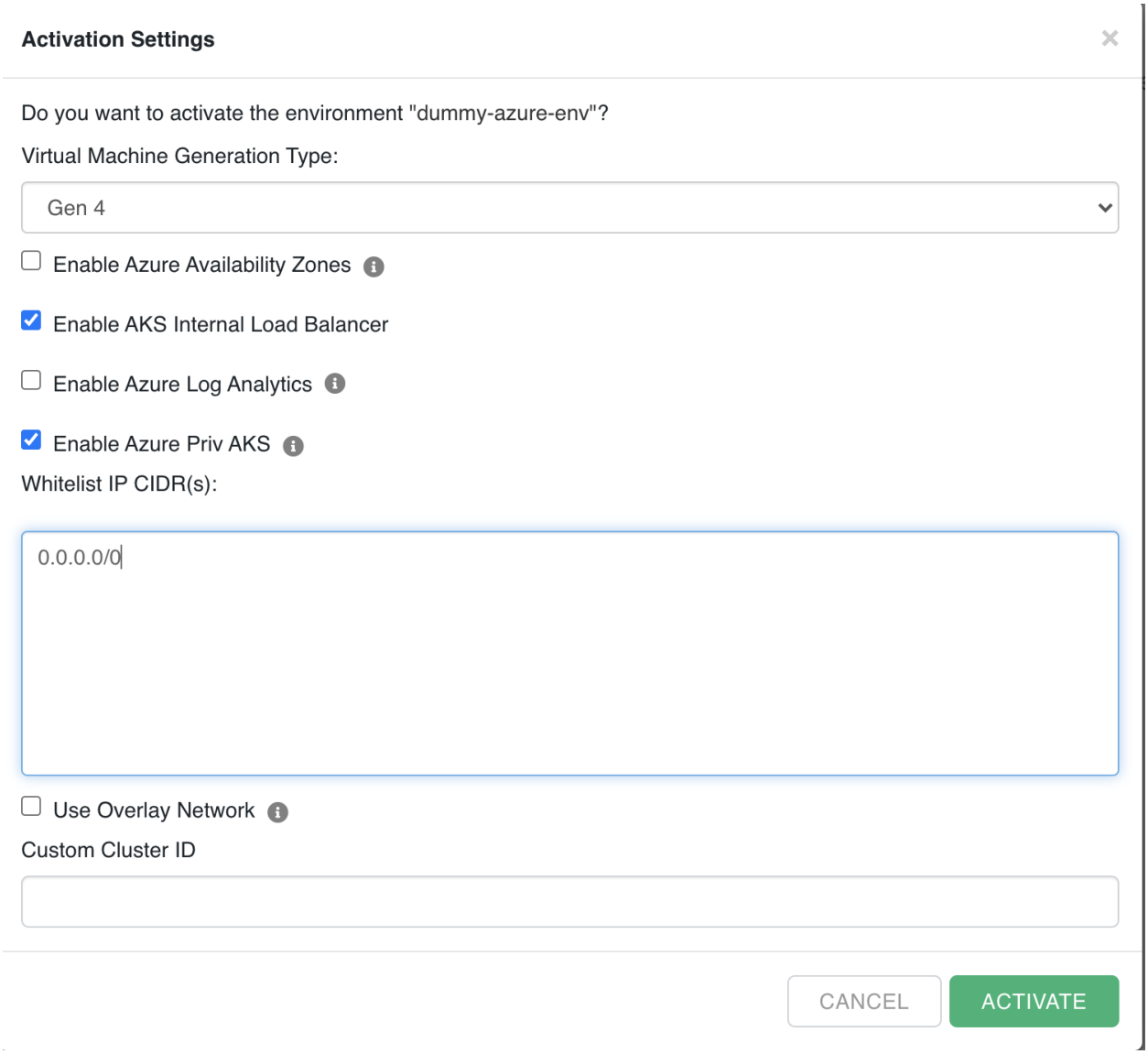
> **Note:** If there is a private DNS zone (`privatelink.postgres.database.azure.com`) already linked to the CDW VNET, you must remove it before environment activation.

Steps

1. In the CDW service, expand the **Environments** column by clicking the **More…** menu on the left side of the page.
2. Locate the environment that you want to activate.
3. Click the activation icon to launch the **Activation Settings** dialog box:



4. In the **Activation Settings** dialog box, you can enable environment features:



   a. (Optional) Select **Enable Azure Availability Zones**.
   b. (Mandatory) Choose the appropriate subnet from the drop-down list.
   c. (Mandatory) Select **Enable AKS Internal Load Balancer** to make endpoint access private. You must select this option to access Hue, DAS, and JDBC endpoints.

d. (Optional) Select **Enable Azure Log Analytics** and then select the workspace from the adjacent drop-down list.
e. (Mandatory) Select **Enable Azure Priv AKS** to create a private AKS cluster with UDR support.
f. (Mandatory) Enter `0.0.0.0/0` in the **Whitelist IP CIDRs** text box.
g. (Optional) Select **Use Overlay Network** if IP address exhaustion is a concern for your deployment.
h. (Optional) Specify a custom cluster ID for your environment and resources.

5. Click **Activate**.

Related links
Private endpoint for PostgreSQL

Use a custom private DNS Zone for AKS

When you set up the environment for private cluster deployment, you configure custom DNS on the VNET to resolve Azure Private DNS zones. If this default AKS Private DNS zone creation behavior is not suitable for your needs, you can use the CDP CLI to further customize the zones. Activating CDW clusters with the private AKS option enables multiple different AKS features at once to support private CDW deployments.

Prerequisites

To successfully activate the CDW cluster with the custom DNS zone you must complete the following tasks:

- Create the DNS zone `privatelink.<region>.azmk8s.io`.
- Grant at least the private dns zone contributor and vnet contributor roles to the service principal (Azure Enterprise Application) used for environment creation.
- If the Private DNS Zone is in a different subscription than the AKS cluster, you need to register Microsoft.ContainerServices in both the subscriptions.
- Make sure the Zone is not linked to the CDW subnet.

For more information, see Azure documentation.

Syntax

The following syntax shows how to use the CDP CLI to configure a DNS zone for AKS:

```
cdp dw create-cluster --<create_cluster_options> \
--environment-crn <CLOUDERA_RESOURCE_NAME_OF_ENVIRONMENT> \
```

```
--azure-options <property>=<value>,<property>=<value> ... \
privateDNSZoneAKS=<PRIVATE_DNS_ZONE_RESOURCE_ID>
```

Customize AKS deployment options

You use the following properties and values to customize the deployment using the --azure-options parameter:

- enableUDR=true/false
  You can customize an AKS cluster with a unique outboundType of either type loadBalancer or userDefinedRouting. If you set enableUDR to true, CDW creates AKS clusters with outboundType of userDefinedRouting. Link to Azure documentation.
- enablePrivateSQL=true/false
  If you set this option to true, Private Link allows you to create private endpoints for Azure Database for PostgreSQL - Single server to bring the server inside your Virtual Network (VNet). Link to Azure documentation.
- enablePrivateAks=true/false
  By using a private cluster, you can ensure network traffic between your API server and your node pools remains on the private network. Enabling this option instructs CDW services to create an AKS cluster with a private API. Link to Azure documentation.

-- create-cluster options are:

- *--use-private-load-balancer*
- *--no-use-private-load-balancer*
  An internal load balancer makes a Kubernetes service accessible only to applications running in the same virtual network as the Kubernetes cluster. Link to Azure documentation.

Example command

The following command configures a DNS zone for AKS:

```
cdp dw create-cluster --environment-crn --use-private-load-balancer \
--azure-options enableUDR=true,enablePrivateSQL=true,\
enablePrivateAks=true \
privateDNSZoneAKS=/subscriptions/subid/resourceGroups/rgname/providers/Microsoft.Network/privateDnsZones/privatelink.westus2.azmk8s.io
```