Cloudera Data Warehouse Public Cloud

Cloud Support

Date published: 2024-01-01
Date modified: 2025-10-21

CLOUDERA

https://docs.cloudera.com/

Legal Notice

Contents

Cloudera Data Warehouse on cloud supported providers

Cloudera Data Warehouse on cloud service supports Amazon Web Services (AWS) and Microsoft Azure. Basic configuration for environments that use these public cloud providers is performed during registration in Management Console. To configure specific Cloudera Data Warehouse features, you use the Cloudera Data Warehouse service UI.

Amazon Web Services support in Cloudera Data Warehouse on cloud

This topic provides an overview of Amazon Web Services (AWS) support in Cloudera Data Warehouse on cloud.

⚠️ **Important:** Ensure your AWS environment satisfies the items listed in the AWS environments requirements checklist.

The "environment" concept in Cloudera is closely related to the virtual private network in your AWS account. Registering an environment provides Cloudera with access to your AWS account and identifies the resources there that Cloudera services can access or provision. A single environment is contained within a single AWS region, so all resources deployed by Cloudera are deployed within that region within one specific virtual network. Once you've registered an environment with Cloudera Management Console, you can start provisioning resources such as clusters, which run on the physical infrastructure in an AWS data center.

The following diagram shows the components of a Cloudera environment on AWS:
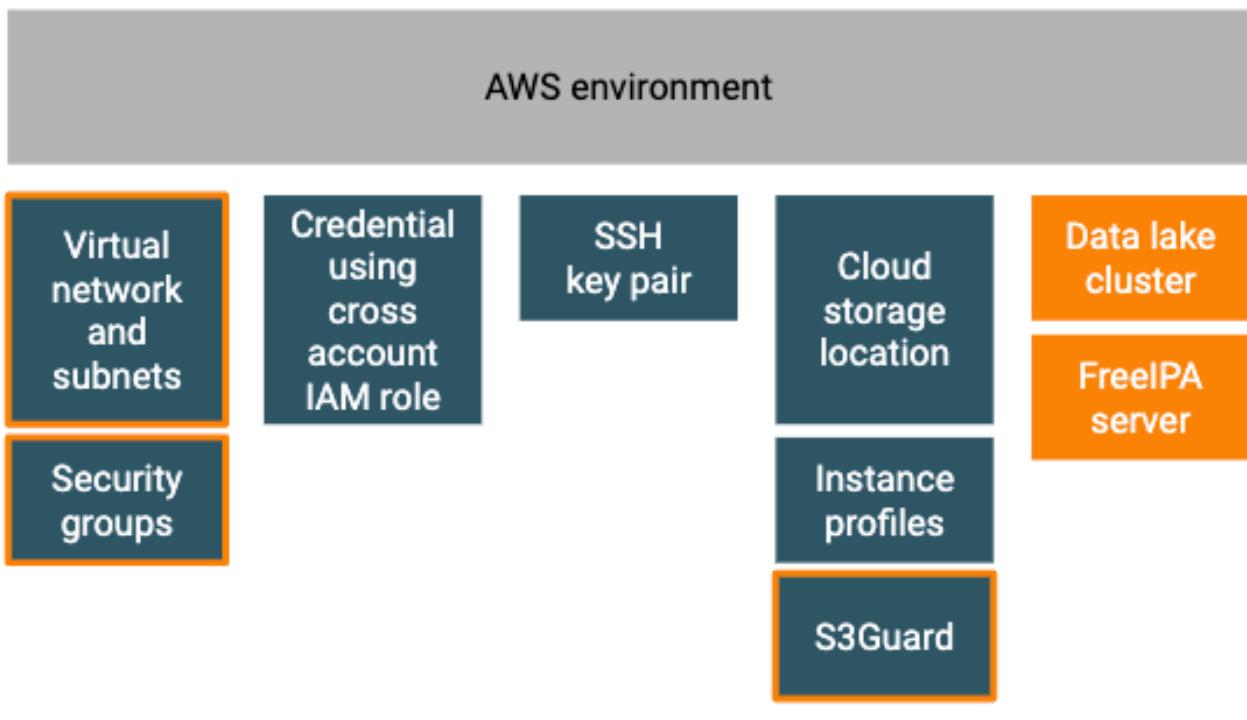
The diagram includes all major user-created and Cloudera-created components of an AWS environment:

- Items in dark blue boxes with orange outlines can be automatically provisioned by Cloudera in your AWS account. Or you can optionally pre-create them in your AWS account and then provide them when registering an environment in Cloudera.
- Items in dark blue boxes must be pre-created by your Cloudera administrator prior to environment registration and then provided during environment registration in Cloudera.
- Items in orange boxes are automatically provisioned on AWS by Cloudera as part of environment registration.

Note: Items that are user-created are not terminated when a Cloudera environment is deleted.

Information that describes how to register an AWS environment with Cloudera is linked to at the end of this page.

Cloudera Data Warehouse service features for AWS environments

Cloudera Data Warehouse offers the following additional features for AWS environments that are used for Database Catalogs and Virtual Warehouses:

- Restrict access to Kubernetes endpoints and to service endpoints of the Kubernetes cluster at the load balancer lever by specifying a list of IP CIDRs that are allowed access. For more information, see "Restricting endpoint access in AWS," which is linked to at the end of this page.
- Set up private networking in AWS environments, which uses AWS private subnets. Private subnets receive no direct inbound connections from the internet, providing private network connectivity for workload endpoints in the Cloudera Data Warehouse service. For more information, see "Set up private networking in AWS," which is linked to at the end of this page.

These features require additional configuration in the Cloudera Data Warehouse service UI to use them.

Related Information
AWS account requirements
AWS credentials
How to register AWS environments
Restricting endpoint access in AWS
Set up private networking in AWS

Microsoft Azure support in Cloudera Data Warehouse on cloud

This topic provides an overview of Azure support in Cloudera Data Warehouse on cloud.

⚠️ **Important:** Ensure your Azure environment satisfies the items listed in the Azure environments requirements checklist.

Like Amazon Web Services support, Cloudera environments are closely related to the virtual private network concept in your Azure account as well. Also registering an environment provides Cloudera access to resources in your Azure account and a single Cloudera environment is contained within a single region in Azure. All resources that are deployed by Cloudera are deployed within that region and within one specific virtual network. After you have registered an environment in Cloudera, you can start provisioning Cloudera resources such as clusters, which run on the physical infrastructure in an Azure data center.

The following diagram shows the components of a Cloudera environment on Azure:
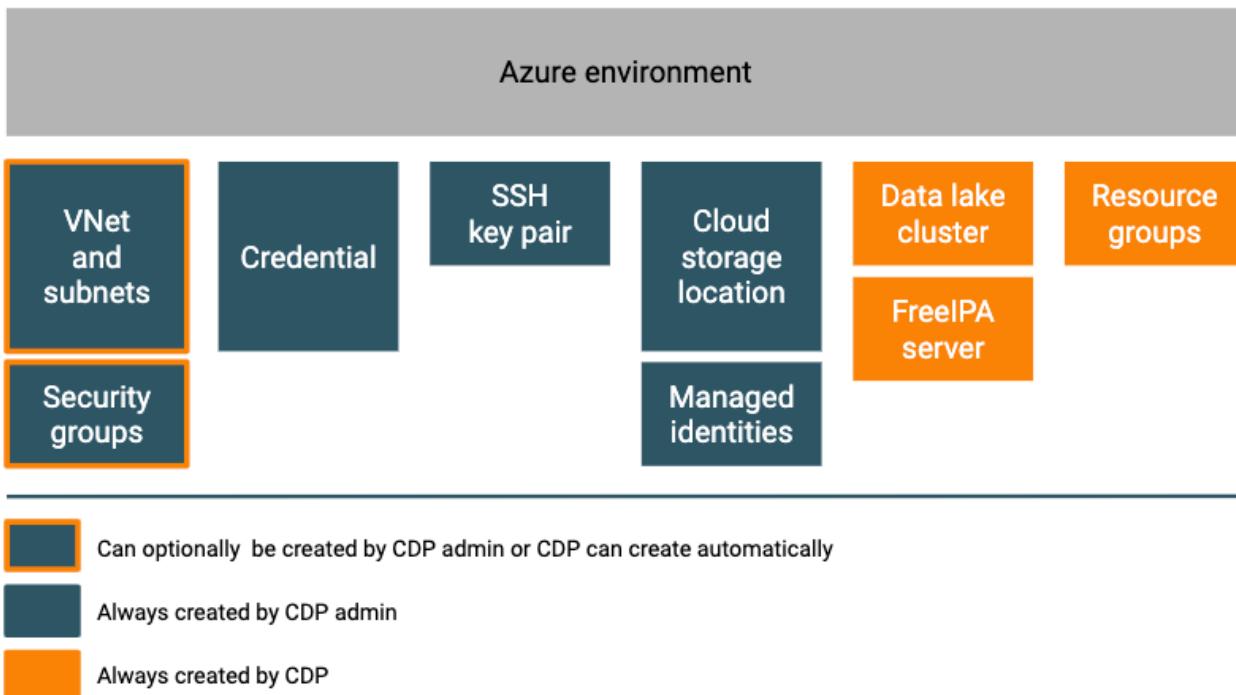
The diagram illustrates all major user-created and Cloudera-created components of an environment:

- The items in dark blue boxes with orange outlines can either be automatically provisioned by Cloudera on your Azure account, or you can optionally pre-create them and specify them when registering an environment.
- The items in dark blue boxes must be pre-created by your Cloudera administrator prior to environment registration and then specified when registering an environment.

- The items in orange boxes are automatically provisioned on Azure by Cloudera as part of environment provisioning.

Note: The items that are user-created are not terminated during environment deletion.

Information that describes how to register an Azure environment with Cloudera is linked to at the end of this page.

Related Information

Azure account requirements

Azure credentials

How to register Azure environments

Custom image repositories

If your organization must control the acquisition and provisioning of images in your cloud account, custom image repositories are available. If your organization does not allow internet access, or restricts image repositories to only those within your virtual private network (VPC in AWS or VNet in Azure), you can bring your own repository to Cloudera.

Using your custom repository, you can host and scan Cloudera Data Warehouse images. You gain complete control over which images are provisioned in your cloud account and how you acquire the images.

The following registries are supported for use as a custom image repository in Cloudera Data Warehouse:

- Azure Container Registry (ACR)
- Amazon Elastic Container Registry (ECR)

To use this feature, you must obtain the CDP_CUSTOM_REPO entitlement.

Cloudera Data Warehouse architecture

The following diagram shows the Cloudera Data Warehouse architecture using a custom image repository based on ACR or ECR.
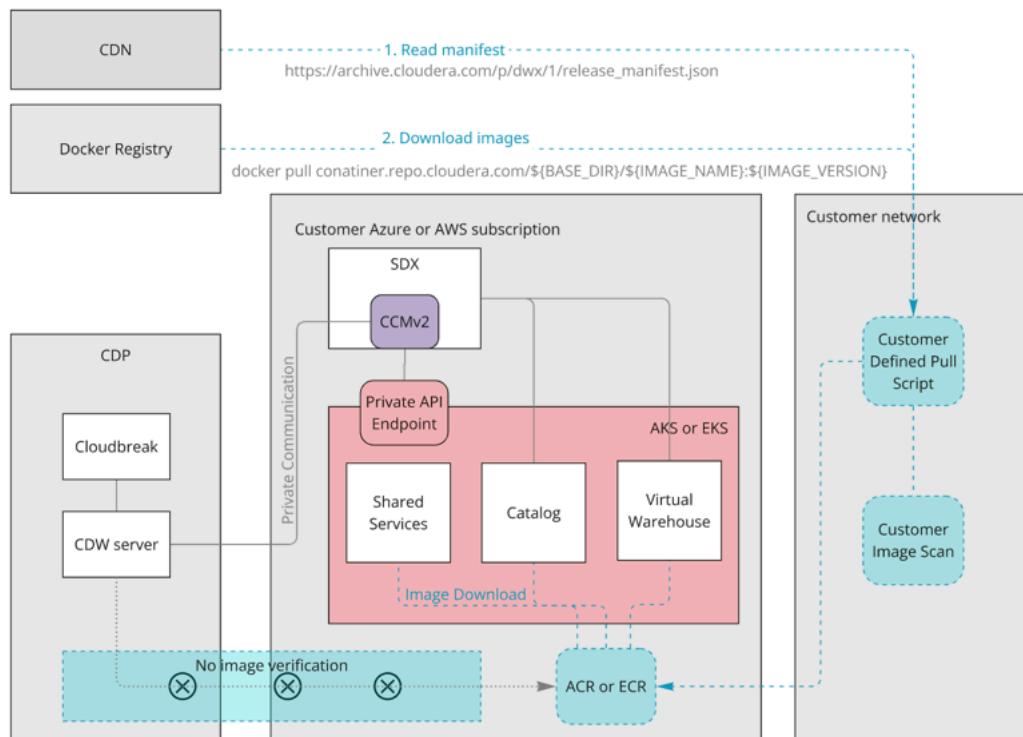
When you set up a custom repository, you must read the manifest.json file using your paywall credentials from Cloudera to obtain a list of images. Cloudera provides an image list with each release. You must pull images each time a new release of Cloudera Data Warehouse occurs. Cloudera recommends that you create an automated way to pull these images.

Image version verification is not available in private registries, the server starts the new images without verification. If the images are not present in the registry, provisioning fails.

Limitations

Learn about the limitations of using custom image repositories in Cloudera.

- Image namespaces must not include arbitrary paths, such as .../mycompany/myhierarchy/hive:… Only the following namespace format is allowed:

```
container.repo.cloudera.com/cloudera/hive:…
```

Downloading images

Learn how to download Cloudera Data Warehouse container images from the registry. You can run a security scan and approve (or block) each image before copying the image into your repository.

Before you begin
Obtain the following entitlements:

- Salesforce entitlement
- CDP_CUSTOM_REPO entitlement

Procedure

1. Login to Cloudera container registry (https://container.repo.cloudera.com) using your Salesforce entitlement token and password.
   For example:

   ```
   $ docker login container.repo.cloudera.com -u 634ae2db-f926-4530-842c-95
   d3401f37e5
   ```

2. Download container images using the docker pull command. For example:

   ```
   docker pull container.repo.cloudera.com/cloudera/hive:2025.0.19.0-123
   ```

   For example:

   ```
   Pulling from cloudera/hive
   Digest:sha256:29c635b8612c770c8089d4ab134c10e599dcf61553dfa4147975d6af33c
   835a9
   Status: Downloaded newer image for container.repo.cloudera.com/cloudera/hi
   ve:2025.0.19.0-123
   ```

3. It is recommended that you set up a script that periodically pulls all the images from the registry. Use the following example script to automate this process:

   ```
   for CDW_IMAGE in $(curl -L -s -u ${PAYWALL_USER}:${PAYWALL_PASSWORD} htt
   ps://archive.cloudera.com/p/dwx/1/release_manifest.json | jq -r '.images
   [].paths[] | "\(.path),\(.version)"')
   do
       BASE_DIR=$(dirname $(echo ${CDW_IMAGE} | sed 's|,| |' | awk '{print
    $1}'))
       IMAGE_NAME=$(basename $(echo ${CDW_IMAGE} | sed 's|,| |' | awk '{print
    $1}'))
       IMAGE_VERSION=$(echo ${CDW_IMAGE} | sed 's|,| |' | awk '{print $NF}')
       skopeo copy -all container.repo.cloudera.com/${BASE_DIR}/${IMAGE_NAM
   E}:${IMAGE_VERSION}
       ${DESTINATION_REGISTRY}/${BASE_DIR}/${IMAGE_NAME}:${IMAGE_VERSION}
   done
   ```

Setting up a custom repository

When you activate the environment from Cloudera Data Warehouse, you can choose to use a custom repository.

Procedure

1. Log in to the Cloudera Data Warehouse service as DWAdmin.
2. Go to the Environments tab, locate the environment you want to activate, and click Activate.

   The **Activate Environment** modal is displayed.
3. If you have an ECR repository, select the Use Custom ECR repository option and specify the URL of the ECR repository with https:// as a prefix.

   ☑ Use Custom ECR repository ⓘ

   ```
   Enter custom ECR repository URL
   ```

4. Alternatively, if you have an ACR repository, select the Use Custom repository option, choose the acr reigstry type, and specify the URL of the ACR repository with https:// as a prefix.

☑ Use Custom repository ⓘ

```
Enter Repository URL
```

Registry Type:

```
acr                                                          ⌄
```

```
Enter Repository URL
```

5. Click ACTIVATE to activate the environment.

Copying images to custom ECR repository

While copying images from the Cloudera hosted repository to your custom ECR repository, ensure that the metadata (SHA) of the copied images remains unchanged. Cloudera recommends using third-party tools, such as 'skopeo', to copy images while preserving the SHA.

About this task

Due to security considerations, Cloudera images are sensitive and require their hash values to remain unchanged when transferring images between repositories. Ensure that images are copied while retaining the image manifests and hash values (SHA). If the image SHA in the custom ECR repository differs from the SHA in the Cloudera hosted repository, you may encounter issues while activating the Cloudera Data Warehouse clusters.

To copy images between repositories while preserving image metadata, you may use third-party tools such as 'skopeo'. For more information on installing or building skopeo, refer to the *skopeo documentation*.

Note: This change applies only to Cloudera Data Warehouse 1.10.1 and higher versions.

Procedure

1. Run the following command to copy images from the source to your custom ECR repository:

```
$ skopeo copy --all [***SOURCE-IMAGE***] [***DESTINATION-IMAGE***]
```

2. Run the following Docker command on both the Cloudera hosted repository and your custom ECR repository to verify that the RepoDigests field remains unchanged after the transfer:

Note: The following example uses the calico-cni:v3.28.2-r2-202501062350 image. You can follow this example to verify all images that are copied from the Cloudera repository.

```
docker inspect [***CLOUDERA REPOSITORY***]/[***IMAGE-PATH***]/hardened/c
alico-cni:v3.28.2-r2-202501062350

Output:
[
    {
        "Id": "sha256:6282eb96238744a4f44b3086c5d16f8214c58969f3938b212d1
329c3da65bd09",
        "RepoTags": [
            "[***CLOUDERA REPOSITORY***]/[***IMAGE-PATH***]/hardened/calic
o-cni:v3.28.2-r2-202501062350"
        ],
        "RepoDigests": [
```

```
                "[***CLOUDERA REPOSITORY***]/[***IMAGE-PATH***]/hardened/cali
co-cni@sha256:99c0db8740c07316b709e0bf6ea9f30179a78481ee55d81da6bdc3268a
05bf6a"
        ],
        ....
        ....
```

Related Information
Installing skopeo

Handling Cloudera Data Warehouse upgrades

About this task

You must update your custom repository to contain New Cloudera Data Warehouse images when available before you upgrade your Virtual Warehouse or Database Catalog. The Virtual Warehouse UI indicates when an upgrade is available. If the new images are not available in your custom repository the upgrade fails. Currently there is no rollback option available.