

Custom images and catalogs

Date published: 2022-01-31

Date modified:

CLOUDERA

Legal Notice

© Cloudera Inc. 2024. All rights reserved.

The documentation is and contains Cloudera proprietary information protected by copyright and other intellectual property rights. No license under copyright or any other intellectual property right is granted herein.

Unless otherwise noted, scripts and sample code are licensed under the Apache License, Version 2.0.

Copyright information for Cloudera software may be found within the documentation accompanying each component in a particular release.

Cloudera software includes software from various open source or other third party projects, and may be released under the Apache Software License 2.0 (“ASLv2”), the Affero General Public License version 3 (AGPLv3), or other license terms. Other software included may be released under the terms of alternative open source licenses. Please review the license and notice files accompanying the software for additional licensing information.

Please visit the Cloudera software product page for more information on Cloudera software. For more information on Cloudera support services, please visit either the Support or Sales page. Feel free to contact us directly to discuss your specific needs.

Cloudera reserves the right to change any products at any time, and without notice. Cloudera assumes no responsibility nor liability arising from the use of products, except as expressly agreed to in writing by Cloudera.

Cloudera, Cloudera Altus, HUE, Impala, Cloudera Impala, and other Cloudera marks are registered or unregistered trademarks in the United States and other countries. All other trademarks are the property of their respective owners.

Disclaimer: EXCEPT AS EXPRESSLY PROVIDED IN A WRITTEN AGREEMENT WITH CLOUDERA, CLOUDERA DOES NOT MAKE NOR GIVE ANY REPRESENTATION, WARRANTY, NOR COVENANT OF ANY KIND, WHETHER EXPRESS OR IMPLIED, IN CONNECTION WITH CLOUDERA TECHNOLOGY OR RELATED SUPPORT PROVIDED IN CONNECTION THEREWITH. CLOUDERA DOES NOT WARRANT THAT CLOUDERA PRODUCTS NOR SOFTWARE WILL OPERATE UNINTERRUPTED NOR THAT IT WILL BE FREE FROM DEFECTS NOR ERRORS, THAT IT WILL PROTECT YOUR DATA FROM LOSS, CORRUPTION NOR UNAVAILABILITY, NOR THAT IT WILL MEET ALL OF CUSTOMER’S BUSINESS REQUIREMENTS. WITHOUT LIMITING THE FOREGOING, AND TO THE MAXIMUM EXTENT PERMITTED BY APPLICABLE LAW, CLOUDERA EXPRESSLY DISCLAIMS ANY AND ALL IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO IMPLIED WARRANTIES OF MERCHANTABILITY, QUALITY, NON-INFRINGEMENT, TITLE, AND FITNESS FOR A PARTICULAR PURPOSE AND ANY REPRESENTATION, WARRANTY, OR COVENANT BASED ON COURSE OF DEALING OR USAGE IN TRADE.

Contents

Custom images and image catalogs.....	4
Creating a custom image and image catalog with the CDP CLI.....	5
Switching image catalogs.....	7

Custom images and image catalogs

If necessary, you can use a custom Runtime or FreeIPA image for compliance or security reasons. You can then use the CDP CLI to register a custom image catalog and set the custom image within the custom image catalog. Later, you can use this custom image to create a Data Lake/Data Hub cluster or environment with a custom FreeIPA image.

Overview

A custom image should inherit most of its attributes from its source image, which is a default image that you select from the cdp-default image catalog.

The typical method of creating a Data Lake or Data Hub picks up the latest pre-warmed image from the cdp-default image catalog for the specified version of Runtime. These default images are pre-warmed VM images that contain a base URL to the default parcels in the Cloudera archive, amongst other configurations. If the default pre-warmed images do not suit your business needs, you can specify that the Data Lake/Data Hub or the environment (in the case of FreeIPA) uses a custom image instead.

What is a custom image?

A custom image is an entry in a custom image catalog that inherits most of its attributes from a source (default) image.

Custom image entries have:

- An image type: Runtime [which includes Data Hub and Data Lake images] or FreeIPA
- A source image ID that points to an image in the cdp-default image catalog
- A timestamp of creation
- An option to specify a VM region and image reference (such as an AMI ID) if you are overriding the source image with a custom VM image
- An option to override the parcel base URL

Why use a custom image?

You might require a custom image for compliance or security reasons (a “hardened” image), or to have your own packages pre-installed on the image, for example monitoring tools or software. You might also want to specify a custom image if you need to use a default image with a specific Runtime maintenance version applied, rather than simply specifying the latest major Runtime version.



Note: When customizing VM images, note that certain customizations (for example, CIS hardening rules) may not be compatible with CDP.

What can you customize?

In a custom image entry, you can override the VM images themselves with your own custom images that are sufficiently hardened. Importantly, you should only customize a default image from the cdp-default catalog as opposed to creating one from scratch. You can also override the default parcel base URL (at archive.cloudera.com) with your own host site.

What is a custom image catalog?

A custom image catalog is simply a catalog that holds custom images. A custom image catalog can contain a single or multiple custom image entries.

Custom image catalogs have:

- A name. The name is a unique identifier and is used to refer to the catalog during environment, Data Lake, and Data Hub creation; as well as during catalog operations like creating an image.

- A description.
- An owner. The owner is the user who runs the command to create the catalog.

What is the process for creating a custom image and catalog?

- If you are replacing the VM images in a custom image entry with a customized version, you should first prepare the image by modifying an official Cloudera default image, which you can find under Shared Resources > Image Catalogs > cdp-default.
- Select a source image from the cdp-default image catalog to be the source of customization. When you run the CLI command to find a default image, you specify the Runtime version, provider, image type, or a combination of the three.
- Create a custom image catalog, or identify an existing catalog where you want to save the custom image entry.
- Apply the necessary changes to the custom image entry, like the override AMI IDs with the new, customized AMIs; or add a new parcel base URL using the `--base-parcel-url` command when you set the custom image.
- You can then create an environment, Data Lake, or Data Hub, based on custom catalogs via the CDP CLI.

Creating a custom image and image catalog with the CDP CLI

You can create a custom Runtime or FreeIPA image and image catalog through the CDP CLI.

Before you begin

If you are replacing the VM images in a custom image entry with a customized version, you must first prepare the image by modifying an official Cloudera default image, which you can find under Shared ResourcesImage Catalogscdp-default. Take note of the image reference, such as the AMI ID.

Procedure

1. To find a source image from the cdp-default catalog that you want to use as the source of your custom image entry, run the following command:

```
cdp imagecatalog find-default-image --provider <cloud-provider> --image-type <image-type> --runtime-version <Cloudera-Runtime-version>
```

For example:

```
cdp imagecatalog find-default-image --provider AWS --image-type runtime --runtime-version 7.2.12
```

2. A custom image requires a custom catalog. If you haven't yet created a custom catalog, or if you want to create a new one for a new custom image, run the following command:

```
cdp imagecatalog create-custom-catalog --catalog-name <unique-catalog-name> --description <catalog-description>
```

For example:

```
cdp imagecatalog create-custom-catalog --catalog-name my-custom-catalog --description test-catalog
```

3. Within the custom image catalog that you created (or an older custom catalog that you want to use), create a custom image entry with the selected source image marked as its source. Providing your own VM images is optional. Run the following command:

```
cdp imagecatalog set-<image-type>-image --catalog-name <name-of-the-custom-catalog> --vm-images region=<region-of-customized-image>,imageReference=<cloud-provider-specific-ID-of-a-customized-image> --source-image-id <cdp-image-ID-of-source-image>
```

For example:

```
cdp imagecatalog set-runtime-image --catalog-name my-custom-catalog --vm-images region=eu-central-1,imageReference=ami-7torotmqi6q7438y --source-image-id 8t4y9853-12b6-3n6z-75dh-tx775k4c793w
```

4. You can then apply the necessary changes to the custom image entry, like overriding AMI IDs with new, customized ones, or adding a new parcel base URL.

Command	Description
cdp imagecatalog find-default-image	Finds the default images in the cdp-default image catalog for a specified version of Runtime. Parameters: <ul style="list-style-type: none"> • --provider • --image-type • --runtime-version You can provide any combination of these parameters.
cdp imagecatalog create-custom-catalog	Creates a custom catalog. Parameters: <ul style="list-style-type: none"> • --catalog-name • --description --catalog-name is required.
cdp imagecatalog set-runtime-image or cdp imagecatalog set-freeipa-image	Creates a custom image entry (either Runtime [Data Hub/Data Lake] or FreeIPA) within the specified catalog. Parameters: <ul style="list-style-type: none"> • --catalog-name • --vm-images [region,imageReference] • --source-image-id • --image-id • --base-parcel-url --catalog-name and --source-image-id are required.

Switching image catalogs

You can switch the image catalog of an already existing Data Hub, Data Lake, or FreeIPA cluster. You may want to switch the image catalog for a cluster in order to restrict which Runtime version can be upgraded to, or in order to move to custom images for an existing cluster.

Use the following CDP CLI commands to switch the image catalog for an existing cluster:

- FreeIPA:

```
cdp environments set-catalog --environment $ENVIRONMENT_NAME --catalog $
CATALOG
```

Parameter	Description
--environment	Name or CRN of the environment that holds the FreeIPA installation.
--catalog	URL of the FreeIPA catalog to be used.

- Data Hub:

```
cdp datahub set-catalog --cluster $DATAHUB_NAME --catalog-name $CATALOG
```

Parameter	Description
--cluster	Name or CRN of the Data Hub for which you want to use the new image catalog.
--catalog-name	Name of the image catalog to be used. The image catalog must be a JSON based catalog, and switching is only possible from catalogs that are JSON based.

- Data Lake:

```
cdp datalake set-catalog --datalake $DATALAKE_NAME --catalog-name $CATALOG
```

Parameter	Description
--datalake	Name of the Data Lake for which you want to use the new image catalog.
--catalog-name	Name of the image catalog to be used. The image catalog must be a JSON based catalog, and switching is only possible from catalogs that are JSON based.