

Data Hub

## Cluster Planning

Date published: 2019-12-17

Date modified: 2023-06-27

# CLOUDERA

<https://docs.cloudera.com/>

# 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

**Cluster planning checklist..... 4**

## Cluster planning checklist

When creating a production cluster, refer to this checklist in order to ensure that you've performed all prerequisite steps required for cluster deployment.

The following table includes a list of available cluster options and prerequisites required for using each option. Depending on where you are at, you can use this checklist as a starting point for planning your cluster and understanding cloud provider requirements related to specific options.

Option	Description
General Settings	
Tags	<p>You can define tags that will be applied to your cluster-related resources (such as VMs) on your cloud provider account.</p> <p>If you would like to use tags, make sure to prepare a list of tags that meet your organization's requirements as well as the cloud provider's requirements. For more information, refer to <a href="#">Tags</a>.</p>
Image catalog	
Image catalog	<p>When creating a DataHub cluster, you can use default prewarmed images, or you can customize a default image. If you would like to use custom images, you should prepare the images and an image catalog, and then register the image catalog prior to cluster creation.</p>
Network and availability	
Subnet	<p>Your cluster is automatically provisioned within the network selected during environment creation. There are two possibilities as far as subnets are concerned:</p> <ul style="list-style-type: none"> <li>• If your network includes a single subnet, your cluster is automatically provisioned into that subnet.</li> <li>• During cluster creation you can select the subnet(s) in which your cluster should be provisioned.</li> <li>• If you would like to deploy your Data Hub in multiple availability zones, you should select multiple subnets.</li> </ul>
Hardware and compute	
EC2 instances	<p>Prior to creating a cluster, determine which instance type you would like to use for each host group. Data Hub supports all instance types that have more than 16 GB RAM. If using cluster definitions, default instance types are suggested. Similarly, for custom deployment, there are defaults provided.</p>
Storage options	<p>Prior to creating a cluster, determine the storage type, a number of storage volumes attached per instance, and the volume size that you would like to use for each host group.</p> <p>Supported storage types vary by region, including:</p> <ul style="list-style-type: none"> <li>• Ephemeral</li> <li>• Magnetic</li> <li>• General purpose</li> <li>• SSD</li> </ul> <p>For more information about these options refer to <a href="#">Amazon EC2 Instance Store</a> in AWS documentation.</p> <p>Note that attaching custom disks is not supported.</p>

Option	Description
Encryption	<p>AWS</p> <p>By default, Data Hubs use the same default key from Amazon's KMS or CMK as the parent environment but you have an option to pass a different CMK during Data Hub creation.</p> <p>For more information about configuring encryption on an environment level, refer to <a href="#">Adding a customer managed encryption key to a CDP environment running on AWS</a>. For more information about configuring a CMK for a specific Data Hub, refer to <a href="#">EBS Encryption on AWS</a>.</p> <p>Azure</p> <p>By default, local Data Hub disks attached to Azure VMs and the PostgreSQL server instance used by the Data Lake and Data Hubs are encrypted with server-side encryption (SSE) using Platform Managed Keys (PMK), but you can optionally configure SSE with Customer Managed Keys (CMK). This can only be configured on an environment level. For more information, refer to <a href="#">Adding a customer managed encryption key for Azure</a>.</p>
Cluster extensions	
Recipes	<p>You can optionally create and run scripts (called "recipes") that perform specific tasks on all nodes of a given host group. If you would like to use recipes, you should prepare them prior to creating a cluster. For more information, refer to <a href="#">Recipes</a>.</p>
Custom properties	<p>When creating a cluster, it is possible to include a list of custom properties and set them to specific values. If you would like to set custom properties, prior to cluster creation you should prepare a JSON file listing these properties. For more information, refer to <a href="#">Custom Properties</a>.</p>