

Resource Planning

Date published: 2020-08-17

Date modified: 2022-11-18



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

Standard resource mode requirements.....	4
Low resource mode requirements.....	4

Standard resource mode requirements

Review the requirements needed to get started with the Cloudera Data Warehouse (CDW) service on Red Hat OpenShift 4.6 and Embedded Container Service (ECS) clusters using the standard resource mode.

- CDP Cloudera Manager must be installed and running.
- CDP Private Cloud must be installed and running. See [Private Cloud Installation Overview and Requirements](#) for more details.
- An environment must have been registered with Management Console on the private cloud. See [CDP Private Cloud Environments](#) for more details.
- In addition to the general requirements, CDW also has the following minimum memory, storage, and hardware requirements for each worker node using the standard resource mode:

Depending on the number of executors you want to run on each physical node, the per-node requirements change proportionally. For example, if you are running 3 executor pods per physical node, you require 384 GB of memory and approximately 1.8 TB of locally attached SSD/NVMe storage.



Important:

When you add memory and storage, it is very important that you add it in the increments stated:

- increments of 128 GB of memory
- increments of 600 GB of locally attached SSD/NVMe storage

If you add memory or storage that is not in the above increments, the memory and storage that exceeds these increments is not used for executor pods. Instead, the extra memory and storage can be used by other pods that require fewer resources.

For example, if you add 200 GB of memory, only 128 GB is used by the executor pods. If you add 2 TB of locally attached storage, only 1.8 TB is used by the executor pods.

Security requirements

The CDW service requires the "cluster-admin" role on the OpenShift and ECS cluster in order to install correctly. The "cluster-admin" role enables namespace creation and the use of the OpenShift Local Storage Operator for local storage.

Related Information

[Hyper-Threading](#)

Low resource mode requirements

Review the requirements needed to run Cloudera Data Warehouse (CDW) service on Red Hat OpenShift 4.6 and Embedded Container Service (ECS) clusters using the low resource mode. This mode reduces the minimum amount of hardware needed.



Important: Lowering the minimum hardware requirement reduces the up-front investment to deploy CDW on OpenShift pods, but it does impact performance. Cloudera recommends that you use the Low Resource Mode option for proof of concept (POC) purposes only. This feature is not recommended for production deployment.

- CDP Cloudera Manager must be installed and running.
- CDP Private Cloud must be installed and running. See [Private Cloud Installation Overview and Requirements](#) for more details.
- An environment must have been registered with Management Console on the private cloud. See [CDP Private Cloud Environments](#) for more details.

- In addition to the general requirements, CDW also has the following minimum memory, storage, and hardware requirements for each worker node using the standard resource mode:

Component	Low resource mode deployment
Nodes	4
CPU	4
Memory	48 GB
Storage	3 x 100 GB (SATA) or 2 x 200 GB (SATA)
Network Bandwidth	1 GB/s guaranteed bandwidth to every CDP Private Cloud Base node



Important: When you add memory and storage for low resource mode, it is very important that you add it in the increments stated in the above table:

- increments of 48 GB of memory
- increments of at least 100 GB or 200 GB of SATA storage

If you add memory or storage that is not in the above increments, the memory and storage that exceeds these increments is not used for executor pods. Instead, the extra memory and storage can be used by other pods that require fewer resources.

Virtual Warehouse low resource mode resource requirements

The following requirements are in addition to the low resource mode requirements listed in the previous section.

Table 1: Impala Virtual Warehouse low resource mode requirements

Component	vCPU	Memory	Local Storage	Number of pods in XSMALL Virtual Warehouse
Coordinator (2)	2 x 0.4	2 x 24 GB	2 x 100 GB	2
Executor (2)	2 x 3	2 x 24 GB	2 x 100 GB	2
Statestore	0.1	512 MB	--	1
Catalogd	0.4	16 GB	--	1
Auto-scaler	0.1	1 GB	--	1
Hue (backend)	0.5	8 GB	--	1
Hue (frontend)	--	--	--	1
Total for XSMALL Virtual Warehouse	8 (7.9)	121.5 GB	400 GB - 3 volumes	--

Impala Admission Control Configuration

- Maximum concurrent queries per executor: 4
- Maximum query memory limit: 8 GB

Table 2: Hive Virtual Warehouse low resource mode requirements

Component	vCPU	Memory	Local Storage	Number of pods in XSMALL Virtual Warehouse
Coordinator (2)	2 x 1	2 x 4 GB	2 x 100 GB	2
Executor (2)	2 x 4	2 x 48 GB (16 GB heap; 32 GB off-heap)	2 x 100 GB	2
HiveServer2	1	16 GB	--	1

Component	vCPU	Memory	Local Storage	Number of pods in XSMALL Virtual Warehouse
DAS	0.5	4 GB	--	1
Hue (backend)	0.5	8 GB	--	1
Hue (frontend)	--	--	--	1
Standalone compute operator	0.1	100 MB (.1 GB)	--	--
Standalone query executor (separate)	Same as executor	Same as executor	Same as executor	--
Data Visualization (optional)	0.5	8 GB	--	1
Total for XSMALL Virtual Warehouse	21 (20.6)	237 GB (236.1)	400 GB - 4 volumes	--

Database Catalog low resource mode requirements

The metastore and the DAS event processor each use 2 CPUs and 8 GB of memory.

Security Requirements

The CDW service requires the "cluster-admin" role on the OpenShift and ECS cluster in order to install correctly. The "cluster-admin" role enables namespace creation and the use of the OpenShift Local Storage Operator for local storage.