

# CDP Private Cloud Data Services Overview

Date published: 2020-12-16

Date modified: 2023-1-24

The Cloudera logo is displayed in a bold, orange, sans-serif font. The word "CLOUDERA" is written in all caps, with the letter 'E' stylized as a horizontal bar with a small triangle on its right side.

# 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

**CDP Private Cloud Data Services overview..... 4**  
 CDP Private Cloud Data Services Components..... 4  
 Benefits of CDP Private Cloud Data Services..... 5

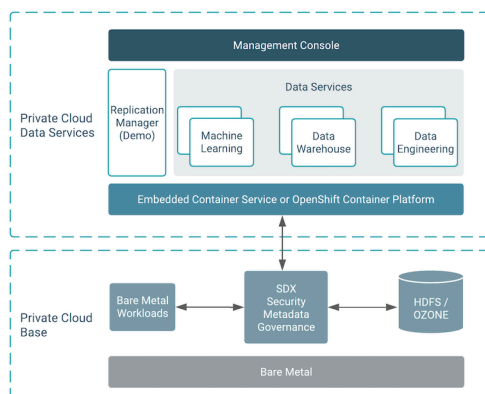
## CDP Private Cloud Data Services overview

CDP Private Cloud Data Services works on top of CDP Private Cloud Base and is the on-premises offering of CDP that brings many of the benefits of the public cloud deployments to on-premises CDP deployments. CDP Private Cloud Data Services lets you deploy and use the Cloudera Data Warehouse (CDW), Cloudera Machine Learning (CML), and Cloudera Data Engineering (CDE) Data Services.

CDP Private Cloud disaggregates compute and storage, which allows independent scaling of compute and storage clusters. The Data Services provide containerized analytic applications that scale dynamically and can be upgraded independently. Through the use of containers deployed on Kubernetes, CDP Private Cloud Data Services brings both agility and predictable performance to analytic applications. CDP Private Cloud Data Services inherits unified security, governance, and metadata management through Cloudera Shared Data Experience (SDX), which is available on the CDP Private Cloud Base cluster.

CDP Private Cloud Data Services users can rapidly provision and deploy services such as Cloudera Data Warehousing, Cloudera Machine Learning, and Cloudera Data Engineering through the Management Console, and easily scale them up or down as required.

A CDP Private Cloud Data Services deployment requires a Private Cloud Base cluster, along with container-based clusters that run the Data Services. You can either use a dedicated RedHat OpenShift container cluster or deploy an Embedded Container Service (ECS) container cluster.



You can install CDP Private Cloud Base on virtual machines or bare-metal hardware. CDP Private Cloud Base provides the following components and services that are used by CDP Private Cloud Data Services:

- SDX Data Lake cluster for security, metadata, and governance
- HDFS or Ozone for storage
- Cloudera Runtime components such as Ranger, Atlas, and Hive Metastore (HMS)
- Networking infrastructure that supports network traffic between storage and compute environments

## CDP Private Cloud Data Services Components

A CDP Private Cloud Data Services deployment includes an Environment, a Data Lake, the Management Console, and Data Services. Data Services include a Data Warehouse, Machine Learning, Data Engineering, and Replication Manager.

### Environment:

A logical entity that represents the association of your Private Cloud user account with compute resources using which you can provision and manage workloads such as Data Warehouse and Machine Learning.

**Data Lake:**

Creates a protective ring of security and governance around your data. For a CDP Private Cloud deployment, the Data Lake services are hosted on the CDP Private Cloud Base cluster. In addition, the Data Lake services are shared between multiple workloads.

**Management Console:**

A service for administering CDP. As a CDP administrator, you can use Management Console for managing environments, data lakes, environment resources, and users across all Private Cloud Data Services.

**Data Warehouse:**

This data service enables an enterprise to provision a new data warehouse and share a subset of the data with a specific team or department. A Data Warehouse cluster can be created from the Management Console and accessed by end users (data analysts).

**Machine Learning:**

This data service enables teams of data scientists to develop, test, train, and ultimately deploy machine learning models for building predictive applications all on the data under management within the enterprise data cloud. A Machine Learning workspace can be created from the Management Console and accessed by end users (data scientists).

**Data Engineering**

This data service allows you to create, manage, and schedule Apache Spark jobs without the overhead of creating and maintaining Spark clusters. You can define virtual clusters with a range of CPU and memory resources, and the cluster scales up and down as needed to execute your Spark workloads, helping to control your cloud costs.

**Replication Manager**

This data service allows you to copy and migrate HDFS data, Hive external tables, and Ozone data between CDP Private Cloud Base 7.1.8 or higher clusters using Cloudera Manager version 7.7.3 or higher.

## Benefits of CDP Private Cloud Data Services

The three main benefits of CDP Private Cloud Data Services are simplified multitenancy and isolation, simplified deployment of applications, and better utilization of infrastructure.

### Simplified multitenancy and isolation

The containerized deployment of applications in CDP Private Cloud ensures that each application is sufficiently isolated and can run independently from others on the same Kubernetes infrastructure, in order to eliminate resource contention. Such a deployment also helps in independently upgrading applications based on your requirements. In addition, all these applications can share a common Data Lake instance.

### Simplified deployment of applications

CDP Private Cloud ensures a much faster deployment of applications with a shared Data Lake compared to monolithic clusters where separate copies of security and governance data would be required for each separate application. In situations where you need to provision applications on an arbitrary basis, for example, to deploy test applications or to allow for self-service, transient workloads without administrative or operations overhead, CDP Private Cloud enables you to rapidly perform such deployments.

### Better utilization of infrastructure

Similar to CDP Public Cloud, CDP Private Cloud enables you to provision resources in real time when deploying applications. In addition, the ability to scale or suspend applications on a need basis in CDP Private Cloud ensures that your on-premises infrastructure is utilized optimally.