

CDP Private Cloud Experiences Security Overview

Date published:

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

CDP user management.....4

Handling of sensitive data in CDP..... 4

Secure in-bound communication.....4

Data Lake Security.....5

CDP user management

The Cloudera Data Platform (CDP) Private Cloud Management Console includes a user management system that allows you to integrate your LDAP identity provider and manage user access to CDP resources.

When CDP Private Cloud is installed, a CDP account administrator user is created. A CDP account administrator has all privileges and can perform any task in CDP. Administrators can create other administrators by assigning the EnvironmentAdmin role to users. CDP users with the EnvironmentAdmin role can also register environments and create Data Lake clusters.

The CDP Private Cloud Management Console also enables account administrators to federate access to CDP by configuring an external LDAP identity provider. CDP users can include users synched with an external LDAP identity provider, or machine users. Machine users can be assigned roles and resource roles, but cannot log in to the web console.

Related Information

[Managing user access and authorization](#)

Handling of sensitive data in CDP

CDP uses [Vault](#) to encrypt sensitive data (such as tokens, passwords, and encryption keys).

The CDP Private Cloud installer can install Vault, but typically this is a pre-existing customer-managed external Vault deployment.

- For more information on how to install an external HashiCorp Vault, see [Install Vault](#).

Vault install notes:

- Supported Vault version: 1.4.0
- Secrets engine: kv-v2
- Auth type: kubernetes
- For more information on how to configure an external HashiCorp Vault for CDP Private Cloud, see [External Vault Requirements](#).

Secure in-bound communication

CDP uses [Vault](#) to encrypt sensitive data (such as tokens, passwords, certificates, and encryption keys). The CDP Private Cloud installer can install Vault, but typically this is a pre-existing customer-managed Vault deployment.

Data Warehouse communication endpoints

The Data Warehouse service runs on top of a Kubernetes cluster and does not include a Cloudera Manager instance.

Primary command and control communication goes to the Kubernetes API server. This endpoint is specific to a particular Kubernetes cluster. The Data Warehouse service does not make connections to endpoints in the cluster.

Machine Learning communication endpoints

In terms of communication, a Machine Learning Workspace looks very similar to a Data Warehouse workspace in that it is also a Kubernetes cluster, although the contents differ.

Primary command and control communication goes to the Kubernetes API server. This endpoint is specific to a particular Kubernetes cluster. The Machine Learning service does not make connections to endpoints in the cluster.

Data Lake Security

CDP Private Cloud security and governance are managed by Apache Ranger and Apache Atlas.

A Data Lake refers to the shared security and governance services in a CDP Private Cloud Base cluster linked to a CDP Private Cloud environment, and managed by Cloudera Manager. This set of shared services is also referred to as SDX (Shared Data eXperience).

Data Lake services

Data Lake services are managed by Cloudera Manager, and can include the following services:

- Hive MetaStore (HMS) -- table metadata
- Apache Ranger -- fine-grained authorization policies, auditing
- Apache Atlas -- metadata management and governance: lineage, analytics, attributes

Security in all workload clusters created in an environment is managed by these shared security and governance services.

Links to the Atlas and Ranger web UIs are provided on each environment home page. A link to the Cloudera Manager instance provides access to configuration settings.

Apache Ranger

Apache Ranger manages access control through a user interface that ensures consistent policy administration in CDP clusters.

Security administrators can define security policies at the database, table, column, and file levels, and can administer permissions for groups or individual users. Rules based on dynamic conditions such as time or geolocation can also be added to an existing policy rule. Ranger security zones enable you to organize service resources into multiple security zones.

Ranger also provides a centralized framework for collecting access audit history and reporting data, including filtering on various parameters.

Apache Atlas

Apache Atlas provides a set of metadata management and governance services that enable you to manage CDP cluster assets.

- Search and Proscriptive Lineage – facilitates pre-defined and ad hoc exploration of data and metadata, while maintaining a history of data sources and how specific data was generated.
- Ranger plugin for metadata-driven data access control.
- Flexible modeling of both business and operational data.
- Data Classification – helps you understand the nature of the data within Hadoop and classify it based on external and internal sources.