

## Apache Knox Authentication

Date published: 2021-02-29

Date modified: 2021-03-91



# Legal Notice

© Cloudera Inc. 2025. 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

|                                                            |              |
|------------------------------------------------------------|--------------|
| <b>Apache Knox Overview.....</b>                           | <b>4</b>     |
| Securing Access to Hadoop Cluster: Apache Knox.....        | 4            |
| Apache Knox Gateway Overview.....                          | 4            |
| Knox Supported Services Matrix.....                        | 5            |
| <br><b>Proxy Cloudera Manager through Apache Knox.....</b> | <br><b>6</b> |
| <br><b>Installing Apache Knox.....</b>                     | <br><b>7</b> |
| Apache Knox Install Role Parameters.....                   | 9            |

# Apache Knox Overview

## Securing Access to Hadoop Cluster: Apache Knox

The Apache Knox Gateway (“Knox”) is a system to extend the reach of Apache™ Hadoop® services to users outside of a Hadoop cluster without reducing Hadoop Security. Knox also simplifies Hadoop security for users who access the cluster data and execute jobs. The Knox Gateway is designed as a reverse proxy.

Establishing user identity with strong authentication is the basis for secure access in Hadoop. Users need to reliably identify themselves and then have that identity propagated throughout the Hadoop cluster to access cluster resources.

### Layers of Defense for a CDP Private Cloud Base Cluster

- Authentication: Kerberos

Cloudera uses Kerberos for authentication. Kerberos is an industry standard used to authenticate users and resources within a Hadoop cluster. CDP also includes Cloudera Manager, which simplifies Kerberos setup, configuration, and maintenance.

- Perimeter Level Security: Apache Knox

Apache Knox Gateway is used to help ensure perimeter security for Cloudera customers. With Knox, enterprises can confidently extend the Hadoop REST API to new users without Kerberos complexities, while also maintaining compliance with enterprise security policies. Knox provides a central gateway for Hadoop REST APIs that have varying degrees of authorization, authentication, SSL, and SSO capabilities to enable a single access point for Hadoop.

- Authorization: Ranger

OS Security: Data Encryption and HDFS

## Apache Knox Gateway Overview

A conceptual overview of the Apache Knox Gateway, a reverse proxy.

### Overview

Knox integrates with Identity Management and SSO systems used in enterprises and allows identity from these systems be used for access to Hadoop clusters.

Knox Gateway provides security for multiple Hadoop clusters, with these advantages:

- Simplifies access: Extends Hadoop’s REST/HTTP services by encapsulating Kerberos to within the Cluster.
- Enhances security: Exposes Hadoop’s REST/HTTP services without revealing network details, providing SSL out of the box.
- Centralized control: Enforces REST API security centrally, routing requests to multiple Hadoop clusters.
- Enterprise integration: Supports LDAP, Active Directory, SSO, SAML and other authentication systems.

### Typical Security Flow: Firewall, Routed Through Knox Gateway

Knox can be used with both unsecured Hadoop clusters, and Kerberos secured clusters. In an enterprise solution that employs Kerberos secured clusters, the Apache Knox Gateway provides an enterprise security solution that:

- Integrates well with enterprise identity management solutions
- Protects the details of the Hadoop cluster deployment (hosts and ports are hidden from end users)
- Simplifies the number of services with which a client needs to interact

## Knox Gateway Deployment Architecture

Users who access Hadoop externally do so either through Knox, via the Apache REST API, or through the Hadoop CLI tools.

## Knox Supported Services Matrix

A support matrix showing which services Apache Knox supports for Proxy and SSO, for both Kerberized and Non-Kerberized clusters.

**Table 1: Knox Supported Components**

| Component                                | UI Proxy (with SSO) | API Proxy |
|------------------------------------------|---------------------|-----------|
| Atlas API                                | #                   | #         |
| Atlas UI                                 | #                   | #         |
| Beacon                                   |                     |           |
| Cloudera Manager API                     | #                   | #         |
| Cloudera Manager UI                      | #                   |           |
| Data Analytics Studio (DAS)              | #                   |           |
| Druid                                    |                     |           |
| Falcon                                   |                     |           |
| Flink                                    |                     |           |
| HBase REST API(aka WebHBase & Stargate)  |                     | #         |
| HBase UI                                 | #                   |           |
| HDFS UI                                  | #                   |           |
| HiveServer2 HTTP JDBC API (HS2 via HTTP) |                     | #         |
| HiveServer2 LLAP JDBC API                |                     |           |
| HiveServer2 LLAP UI                      |                     |           |
| HiveServer2 UI                           |                     |           |
| Hue                                      | #                   |           |
| Impala HTTP JDBC API                     |                     | #         |
| Impala UI                                | #                   |           |
| JobHistory UI                            | #                   |           |
| JobTracker                               |                     | #         |
| Kudu UI                                  | #                   |           |
| Livy API + UI                            | #                   | #         |
| LogSearch                                |                     |           |
| NameNode                                 | #                   | #         |
| NiFi                                     | #                   | #         |
| NiFi Registry                            | #                   | #         |
| Oozie API                                | #                   | #         |
| Oozie UI                                 | #                   |           |
| Phoenix (aka Avatica)                    |                     | #         |

| Component                           | UI Proxy (with SSO) | API Proxy |
|-------------------------------------|---------------------|-----------|
| Profiler                            | #                   |           |
| Ranger API                          | #                   | #         |
| Ranger UI                           | #                   |           |
| ResourceManager API                 | #                   | #         |
| Schema Registry API + UI            | #                   | #         |
| Streams Messaging Manager (SMM) API | #                   | #         |
| Streams Messaging Manager (SMM) UI  | #                   |           |
| Solr                                | #                   | #         |
| Spark3History UI                    | #                   |           |
| SparkHistory UI                     | #                   |           |
| Storm                               |                     |           |
| Storm LogViewer                     |                     |           |
| Superset                            |                     |           |
| WebHCat                             |                     |           |
| WebHDFS                             |                     | #         |
| YARN UI                             | #                   |           |
| YARN UI V2                          | #                   |           |
| Zeppelin UI                         | #                   |           |
| Zeppelin WS                         | #                   |           |

**Note:**

APIs, UIs, and SSO in the Apache Knox project that are not listed above are considered Community Features.

Community Features are developed and tested by the Apache Knox community but are not officially supported by Cloudera. These features are excluded for a variety of reasons, including insufficient reliability or incomplete test case coverage, declaration of non-production readiness by the community at large, and feature deviation from Cloudera best practices. Do not use these features in your production environments.

## Proxy Cloudera Manager through Apache Knox

In order to have Cloudera Manager proxied through Knox, there are some steps you must complete.

### Procedure

- Set the value for `frontend_url`: Cloudera Manager Administration Settings Cloudera Manager Frontend URL :
  - Non-HA value: `https://$Knox_host:$knox_port`
  - HA value: `https://$Knox_loadbalancer_host:$Knox_loadbalancer_port`
- Set allowed groups, hosts, and users for Knox Proxy: Cloudera Manager Administration Settings External Authentication :
  - Allowed Groups for Knox Proxy: \*
  - Allowed Hosts for Knox Proxy: \*
  - Allowed Users for Knox Proxy: \*

3. Enable Kerberos/SPNEGO authentication for the Admin Console and API: Cloudera Manager Administration Settings External Authentication Enable SPNEGO/Kerberos Authentication for the Admin Console and API: : true
4. From Cloudera Manager Administration Settings External Authentication , set Knox Proxy Principal: knox.

#### What to do next

External authentication must be set up correctly. Cloudera Manager must be configured to use LDAP, following the standard procedure for setting up LDAP. This LDAP server should be the same LDAP that populates local users on Knox hosts (if using PAM authentication with Knox), or the same LDAP that Knox is configured to use (if using LDAP authentication with Knox).

## Installing Apache Knox

This document provides instructions on how to install Apache Knox using the installation process.

#### About this task

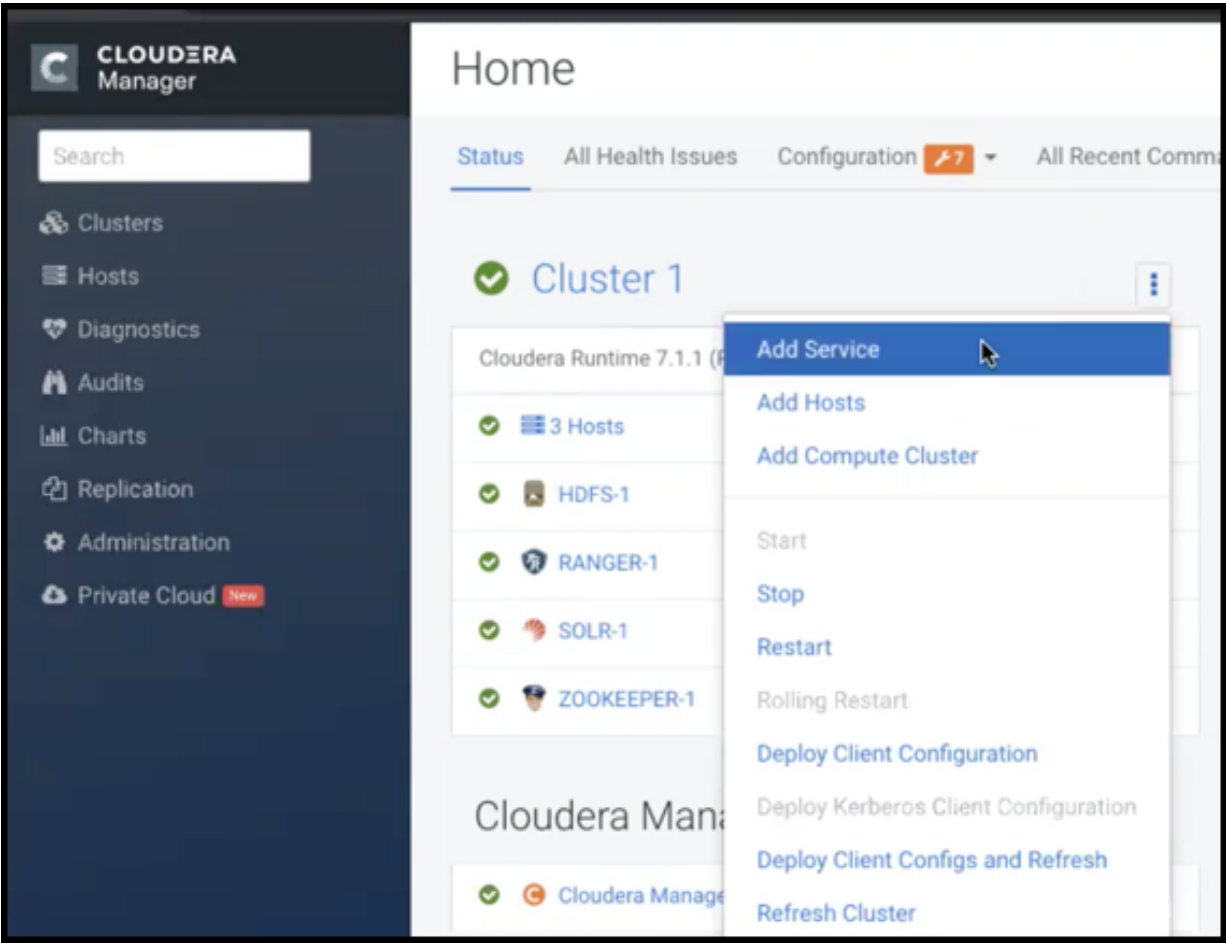
Apache Knox is an application gateway for interacting with the REST APIs and UIs. The Knox Gateway provides a single access point for all REST and HTTP interactions in your Cloudera Data Platform cluster.

#### Before you begin

When installing Knox, you must have Kerberos enabled on your cluster.

Procedure

- 1. From your Cloudera Manager homepage, go to Status tab \$Cluster Name ... Add Service



- 2. From the list of services, select Knox and click Continue.
- 3. On the **Select Dependencies** page, choose the dependencies you want Knox to set up:

|                                      |                                                                                                                                  |
|--------------------------------------|----------------------------------------------------------------------------------------------------------------------------------|
| <b>HDFS, Ranger, Solr, Zookeeper</b> | For users that require Apache Ranger for authorization. HDFS with Ranger. HDFS depends on Zookeeper, and Ranger depends on Solr. |
| <b>HDFS, Zookeeper</b>               | HDFS depends on Zookeeper.                                                                                                       |
| <b>No optional dependencies</b>      | For users that do not wish to have Knox integrate with HDFS or Ranger.                                                           |

- 4. On the **Assign Roles** page, select role assignments for your dependencies and click Continue:

| Knox service roles | Description                                                                                                                                                                                                              | Required? |
|--------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------|
| Knox Gateway       | If Knox is installed, at least one instance of this role should be installed. This role represents the Knox Gateway which provides a single access point for all REST and HTTP interactions with Apache Hadoop clusters. | Required  |

| Knox service roles | Description                                                                                                                                                                                                                                                                                       | Required? |
|--------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------|
| KnoxIDBroker*      | It is strongly recommended that this role is installed on its own dedicated host. As its name suggests this role will allow you to take advantage of Knox's Identity Broker capabilities, an identity federation solution that exchanges cluster authentication for temporary cloud credentials.* | Optional* |
| Gateway            | This role comes with the CSD framework. The gateway structure is used to describe the client configuration of the service on each host where the gateway role is installed.                                                                                                                       | Optional  |

\* Note: KnoxIDBroker appears in the Assign Roles page, but it is not currently supported in CDP Private Cloud.

5. On the **Review Changes** page, most of the default values are acceptable, but you must Enable Kerberos Authentication and supply the Knox Master Secret. There are additional parameters you can specify or change, listed in "Knox Install Role Parameters".
  - a) Click Enable Kerberos Authentication  
Kerberos is required where Knox is enabled.
  - b) Supply the Knox Master Secret, e.g. knoxsecret.
  - c) Click Continue.
6. The **Command Details** page shows the status of your operation. After completion, your system admin can view logs for your installation under stdout.

## Apache Knox Install Role Parameters

Reference information on all the parameters available for Knox service roles.

### Service-level parameters

**Table 2: Required service-level parameters**

| Name                                    | In Wizard | Type    | Default Value                  |
|-----------------------------------------|-----------|---------|--------------------------------|
| kerberos.auth.enabled*                  | Yes       | Boolean | false                          |
| ranger_knox_plugin_hdfs_audit_directory | No        | Text    | \${ranger_base_audit_url}/knox |
| autorestart_on_stop                     | No        | Boolean | false                          |
| knox_pam_realm_service                  | No        | Text    | login                          |
| save_alias_command_input_password       | No        | Text    | -                              |

### Knox Gateway role parameters

**Table 3: Required parameters for Knox Gateway role**

| Name                  | In Wizard | Type     | Default Value              |
|-----------------------|-----------|----------|----------------------------|
| gateway_master_secret | Yes       | Password | -                          |
| gateway_conf_dir      | Yes       | Path     | /var/lib/knox/gateway/conf |
| gateway_data_dir      | Yes       | Path     | /var/lib/knox/gateway/data |
| gateway_port          | No        | Port     | 8443                       |
| gateway_path          | No        | Text     | gateway                    |

| Name                                                  | In Wizard | Type   | Default Value                             |
|-------------------------------------------------------|-----------|--------|-------------------------------------------|
| gateway_heap_size                                     | No        | Memory | 1 GB (min = 256 MB; soft min = 512 MB)    |
| gateway_ranger_knox_plugin_conf_path                  | No        | Path   | /var/lib/knox/ranger-knox-plugin          |
| gateway_ranger_knox_plugin_policy_cache_directory     | No        | Path   | /var/lib/ranger/knox/gateway/policy-cache |
| gateway_ranger_knox_plugin_hdfs_audit_spool_directory | No        | Path   | /var/log/knox/gateway/audit/hdfs/spool    |
| gateway_ranger_knox_plugin_solr_audit_spool_directory | No        | Path   | /var/log/knox/gateway/audit/solr/spool    |

Table 4: Optional parameters for Knox Gateway role

| Name                                                           | Type       | Default Value                                                                                                                                                             |
|----------------------------------------------------------------|------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| gateway_default_topology_name                                  | Text       | cdp-proxy                                                                                                                                                                 |
| gateway_auto_discovery_enabled                                 | Boolean    | true                                                                                                                                                                      |
| gateway_cluster_configuration_monitor_interval                 | Time       | 60 seconds (minimum = 30 seconds)                                                                                                                                         |
| gateway_auto_discovery_advanced_configuration_monitor_interval | Time       | 10 seconds (minimum = 5 seconds)                                                                                                                                          |
| gateway_cloudera_manager_descriptors_monitor_interval          | Time       | 10 seconds (minimum = 5 seconds)                                                                                                                                          |
| gateway_auto_discovery_cdp_proxy_enabled_*                     | Boolean    | true                                                                                                                                                                      |
| gateway_auto_discovery_cdp_proxy_api_enabled_*                 | Boolean    | true                                                                                                                                                                      |
| gateway_descriptor_cdp_proxy                                   | Text Array | Contains the required properties of cdp-proxy topology                                                                                                                    |
| gateway_descriptor_cdp_proxy_api                               | Text Array | Contains the required properties of cdp-proxy-api topology                                                                                                                |
| gateway_sso_authentication_provider                            | Text Array | Contains the required properties of the authentication provider used by the UIs using the Knox SSO capabilities (Admin UI and Home Page). Defaults to PAM authentication. |
| gateway_api_authentication_provider                            | Text Array | Contains the required properties of the authentication provider used by pre-defined topologies such as admin, metadata or cdp-proxy-api. Defaults to PAM authentication.  |

### Knox IDBroker role parameters



**Note:** Knox IDBroker is not currently supported in CDP Private Cloud.

Table 5: Required parameters for Knox IDBroker role

| Name                   | In Wizard | Type     | Default Value               |
|------------------------|-----------|----------|-----------------------------|
| idbroker_master_secret | Yes       | Password | -                           |
| idbroker_conf_dir      | Yes       | Path     | /var/lib/knox/idbroker/conf |
| idbroker_data_dir      | Yes       | Path     | /var/lib/knox/idbroker/data |
| idbroker_gateway_port  | No        | Port     | 8444                        |
| idbroker_gateway_path  | No        | Text     | gateway                     |

| Name               | In Wizard | Type   | Default Value                          |
|--------------------|-----------|--------|----------------------------------------|
| idbroker_heap_size | No        | Memory | 1 GB (min = 256 MB; soft min = 512 MB) |

**Table 6: Optional parameters for Knox IDBroker role**

| Name                                      | Type       | Default Value                                                                                             |
|-------------------------------------------|------------|-----------------------------------------------------------------------------------------------------------|
| idbroker_aws_user_mapping                 | Text       | -                                                                                                         |
| idbroker_aws_group_mapping                | Text       | -                                                                                                         |
| idbroker_aws_user_default_group_mapping   | Text       | -                                                                                                         |
| idbroker_aws_credentials_key              | Password   | -                                                                                                         |
| idbroker_aws_credentials_secret           | Password   | -                                                                                                         |
| idbroker_gcp_user_mapping                 | Text       | -                                                                                                         |
| idbroker_gcp_group_mapping                | Text       | -                                                                                                         |
| idbroker_gcp_user_default_group_mapping   | Text       | -                                                                                                         |
| idbroker_gcp_credential_key               | Password   | -                                                                                                         |
| idbroker_gcp_credential_secret            | Password   | -                                                                                                         |
| idbroker_azure_user_mapping               | Text       | -                                                                                                         |
| idbroker_azure_group_mapping              | Text       | -                                                                                                         |
| idbroker_azure_user_default_group_mapping | Text       | -                                                                                                         |
| idbroker_azure_adls2_tenant_name          | Text       | -                                                                                                         |
| idbroker_azure_vm_assumer_identity        | Text       | -                                                                                                         |
| idbroker_relaodable_refresh_interval_ms   | Time       | 10 seconds (minimum = 1 second)                                                                           |
| idbroker_kerberos_dt_proxyuser_block      | Text Array | A comma-separated list of proxy user configuration used in Knox's dt topology in case Kerberos is enabled |
| idbroker_knox_token_ttl_ms                | Time       | 1 hour (minimum = 1 second)                                                                               |