

Configuring External Authentication with LDAP and SAML

Date published: 2020-02-28

Date modified:

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 in the center.

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

Configuring External Authentication with LDAP and SAML.....	4
User Sign up Process.....	4
Configuring LDAP/Active Directory Authentication.....	4
LDAP General Settings.....	4
LDAP Over SSL (LDAPS).....	5
LDAP Group Settings.....	6
How Login Works with LDAP Group Settings Enabled.....	7
Test LDAP Configuration.....	7
Configuring SAML Authentication.....	7
Configuration Options.....	8
How Login Works with SAML Group Settings Enabled.....	10
Debug Login URL.....	10

Configuring External Authentication with LDAP and SAML

Cloudera Data Science Workbench supports user authentication against its internal local database, and against external services such as Active Directory, OpenLDAP-compatible directory services, and SAML 2.0 Identity Providers.

By default, Cloudera Data Science Workbench performs user authentication against its internal local database. This topic describes the sign up process for the first user, how to configure authentication using LDAP, Active Directory or SAML 2.0, and an optional workaround that allows site administrators to bypass external authentication by logging in using the local database in case of misconfiguration.

User Sign up Process

The first time you visit the Cloudera Data Science Workbench web console, the first account that you sign up with is a local administrator account.

If in the future you intend to use external services for authentication, Cloudera recommends you use exclusive username & email combinations, rather than site administrators' work email addresses, for both the first site administrator account, and any other local accounts created before switching to external authentication. If the username/email combinations are not unique, an email address might end up being associated with different usernames, one for the external authentication service provider and one for a local Cloudera Data Science Workbench account. This will prevent the user from logging into Cloudera Data Science Workbench with their credentials for the external authentication service.

The link to the signup page is only visible on the login page when the authentication type is set to local. When you enable external services for authentication, signing up through the local database is disabled, and user accounts are automatically created upon their first successful login.

Optionally, site administrators can use a Require invitation to sign up flag under the **Admin Settings** tab to require invitation tokens for account creation. When this flag is enabled, only users that are invited by site administrators can login to create an account, regardless of the authentication type.



Important: If you forget the original credentials, or make a mistake with LDAP or SAML configuration, you can use the workaround described in [Debug Login URL](#).

When you switch to using external authentication methods such as LDAP or SAML 2.0, user accounts will be automatically created upon each user's first successful login. Cloudera Data Science Workbench will extract user attributes such as username, email address and full name from the authentication responses received from the LDAP server or SAML 2.0 Identity Provider and use them to create the user accounts.

Configuring LDAP/Active Directory Authentication

Cloudera Data Science Workbench supports both search bind and direct bind operations to authenticate against an LDAP or Active Directory directory service.

The search bind authentication mechanism performs an `ldapsrch` against the directory service, and binds using the found [Distinguished Name \(DN\)](#) and password provided. The direct bind authentication mechanism binds to the LDAP server using a username and password provided at login.

You can configure Cloudera Data Science Workbench to use external authentication methods by clicking the **Admin** link on the left sidebar and selecting the **Security** tab. Select **LDAP** from the list to start configuring LDAP properties.

LDAP General Settings

Describes each of the LDAP general settings.



Note: If you will be creating a Team with an associated LDAP group, you must specify an LDAP Bind DN and LDAP Bind Password.

- **LDAP Server URI:** Required. The URI of the LDAP/Active Directory server against which Cloudera Data Science Workbench should authenticate. For example, `ldaps://ldap.company.com:636`.
- **Use Direct Bind:** If checked, the username and password provided at login are used with the LDAP Username Pattern for binding to the LDAP server. If unchecked, Cloudera Data Science Workbench uses the search bind mechanism and two configurations, LDAP Bind DN and LDAP Bind Password, are required to perform the `ldapssearch` against the LDAP server.
- **LDAP Bind DN:** Required when using search bind. The DN to bind to for performing `ldapssearch`. For example, `cn=admin,dc=company,dc=com`.
- **LDAP Bind Password:** Required when using search bind. This is the password for the LDAP Bind DN.
- **LDAP Search Base:** Required. The base DN from which to search for the provided LDAP credentials. For example, `ou=Engineering,dc=company,dc=com`.
- **LDAP User Filter:** Required. The [LDAP filter](#) for searching for users. For example, `(&(sAMAccountName={0})(objectclass=person))`. The `{0}` placeholder will be replaced with the username provided at login.
- **LDAP User Username Attribute:** Required. The case-sensitive username attribute of the LDAP directory service. This is used by Cloudera Data Science Workbench to perform the bind operation and extract the username from the response. Common values are `uid`, `sAMAccountName`, or `userPrincipalName`.

General Settings

LDAP Server URI *

☐ **Use Direct Bind**
By default, Cloudera Data Science Workbench searches for the users by binding to the provided **LDAP Bind DN** and **LDAP Bind Password**. When checked, Cloudera Data Science Workbench will attempt to search for the user by binding to the user-provided username and password. In such case, please make sure the users have permissions to search for themselves in the LDAP server.

LDAP Bind DN *

[Update LDAP Bind Password](#)

LDAP User Search Base *

LDAP User Search Filter

LDAP User Username Attribute *

When you select **Use Direct Bind**, Cloudera Data Science Workbench performs a direct bind to the LDAP server using the LDAP Username Pattern with the credentials provided on login (not LDAP Bind DN and LDAP Bind Password).

By default, Cloudera Data Science Workbench performs an LDAP search using the bind DN and credentials specified for the LDAP Bind DN and LDAP Bind Password configurations. It searches the subtree, starting from the base DN specified for the LDAP Search Base field, for an entry whose attribute specified in LDAP User Username Attribute, has the same value as the username provided on login. Cloudera Data Science Workbench then validates the user-provided password against the DN found as a result of the search.

LDAP Over SSL (LDAPS)

To support secure communication between Cloudera Data Science Workbench and the LDAP/Active Directory server, Cloudera Data Science Workbench might require a CA certificate to be able to validate the identity of the LDAP/Active Directory service.

- **CA Certificate:** If the certificate of your LDAP/Active Directory service was signed by a trusted or commercial Certificate Authority (CA), it is not necessary to upload the CA certificate here. However, if your LDAP/Active Directory certificate was signed by a self-signed CA, you must upload the self-signed CA certificate to Cloudera Data Science Workbench in order to use LDAP over SSL (LDAPS).

LDAP Group Settings

In addition to the general LDAP settings, you can use the following group settings to restrict the access to Cloudera Data Science Workbench to certain groups in LDAP. This will sync an LDAP group with a team to easily manage the membership for that team.



Note: If you will be creating a Team with an associated LDAP group, you must specify an LDAP User Group or an LDAP Full Administrator Group.



Note: Only site administrators can modify member roles when an LDAP group is first created. The site administrator can assign other members with administrator privileges after group creation.

- **LDAP Group Search Base:** The base distinguished name (DN) where Cloudera Data Science Workbench will search for groups.
- **LDAP Group Search Filter:** The LDAP filter that Cloudera Data Science Workbench will use to determine whether a user is affiliated to a group.

A group object in LDAP or Active Directory typically has one or more member attributes that stores the DNs of users in the group. If LDAP Group Search Filter is set to `member={0}`, Cloudera Data Science Workbench will automatically substitute the `{0}` placeholder for the DN of the authenticated user.

- **LDAP User DN attribute override:** The LDAP user object in the Active directory or LDAP services typically uses the DN attribute that contains the distinguished names for the user. Because this attribute sometimes differs, you might need to override it with a custom value (for example, `memberUid`).
- **LDAP User Groups:** A list of LDAP groups whose users have access to Cloudera Data Science Workbench. When this property is set, only users that successfully authenticate themselves AND are affiliated to at least one of the groups listed here, will be able to access Cloudera Data Science Workbench.

If this property is left empty, all users that can successfully authenticate themselves to LDAP will be able to access Cloudera Data Science Workbench.

- **LDAP Business Groups:** Provides a list of case insensitive LDAP group names(CN). If this list is provided, only users that are members of at least one of the groups in the list will be allowed to log into Cloudera Machine Learning as Business User. If this property is left empty, no LDAP users will be able to log into Cloudera Machine Learning as Business User. For example, if there is a group called `CN=CDSWBusinessUsers,OU=Groups,DC=company,DC=com`, add the group name (CN) `CDSWBusinessUsers` to the LDAP User Groups list to allow members of that group to log in to Cloudera Machine Learning as Business User.
- **LDAP Full Administrator Groups:** A list of LDAP groups whose users are automatically granted the site administrator role on Cloudera Data Science Workbench.

When the LDAP Full Administrator Groups field is used, only users that belong to at least one group specified in the LDAP Full Administrator Groups are granted Admin privilege upon successful login. This means that you

cannot manually grant users Admin permissions if they are not part of one of the groups listed in LDAP Full Administrator Groups. If you do, their Admin access will be revoked when CDSW syncs with the LDAP server.

The groups listed under LDAP Full Administrator Groups do not need to be listed again under the LDAP User Groups property.

Figure 1: Example

If you want to restrict access to Cloudera Data Science Workbench to members of a group whose DN is:

```
CN=CDSWUsers,OU=Groups,DC=company,DC=com
```

And automatically grant site administrator privileges to members of a group whose DN is:

```
CN=CDSWAdmins,OU=Groups,DC=company,DC=com
```

Add the CNs of both groups to the following settings in Cloudera Data Science Workbench:

- LDAP User Groups: CDSWUsers
- LDAP Full Administrator Groups: CDSWAdmins

How Login Works with LDAP Group Settings Enabled

With LDAP Group settings enabled, the login process in Cloudera Data Science Workbench works as follows.

1. Authentication with LDAP

When an unauthenticated user first accesses Cloudera Data Science Workbench, they are sent to the login page where they can login by providing a username and password.

Cloudera Data Science Workbench will search for the user by binding to the LDAP Bind DN and verify the username/password credentials provided by the user.

2. Authorization Check for Access to Cloudera Data Science Workbench

If the user is authenticated successfully, Cloudera Data Science Workbench will then use the LDAP Group Search Filter to search for all groups the user is affiliated to, in the DN provided by LDAP Group Search Base.

The list of LDAP groups the user belongs to is then compared to the pre-authorized lists of groups specified in the LDAP User Groups and LDAP Full Administrator Groups properties.

If there is a match with a group listed under LDAP User Groups, this user will be allowed to access Cloudera Data Science Workbench as a regular user.

If there is a match with a group listed under LDAP Full Administrator Groups, this user will be allowed to access Cloudera Data Science Workbench as a site administrator.

Test LDAP Configuration

You can test your LDAP/Active Directory configuration by entering your username and password in the Test LDAP Configuration section.

This form simulates the user login process and allows you to verify the validity of your LDAP/Active Directory configuration without opening a new window.

Before using this form, make sure you click Update to save the LDAP configuration you want to test.

Configuring SAML Authentication

Cloudera Data Science Workbench supports the [Security Assertion Markup Language \(SAML\)](#) for [Single Sign-on \(SSO\)](#) authentication; in particular, between an identity provider (IDP) and a service provider (SP).

The SAML specification defines three roles: the principal (typically a user), the IDP, and the SP. In the use case addressed by SAML, the principal (user agent) requests a service from the service provider. The service provider

requests and obtains an identity assertion from the IDP. On the basis of this assertion, the SP can make an access control decision—in other words it can decide whether to perform some service for the connected principal.

The primary SAML use case is called web browser single sign-on (SSO). A user with a user agent (usually a web browser) requests a web resource protected by a SAML SP. The SP, wanting to know the identity of the requesting user, issues an authentication request to a SAML IDP through the user agent. In the context of this terminology, Cloudera Data Science Workbench operates as a SP.

Cloudera Data Science Workbench supports both SP- and IDP-initiated SAML 2.0-based SSO. Its [Assertion Consumer Service \(ACS\)](#) API endpoint is for consuming assertions received from the Identity Provider. If your Cloudera Data Science Workbench domain root were `cdsw.company.com`, then this endpoint would be available at `http://cdsw.company.com/api/v1/saml/acs`. SAML 2.0 metadata is available at `http://cdsw.company.com/api/v1/saml/metadata` for IDP-initiated SSO. Cloudera Data Science Workbench uses [HTTP Redirect Binding](#) for authentication requests and expects to receive responses from [HTTP POST Binding](#).

When Cloudera Data Science Workbench receives the SAML responses from the Identity Provider, it expects to see at least the following user attributes in the SAML responses:

- The unique identifier or username. Valid attributes are:
 - `uid`
 - `urn:oid:0.9.2342.19200300.100.1.1`
- The email address. Valid attributes are:
 - `mail`
 - `email`
 - `urn:oid:0.9.2342.19200300.100.1.3`
- The common name or full name of the user. Valid attributes are:
 - `cn`
 - `urn:oid:2.5.4.3`

In the absence of the `cn` attribute, Cloudera Data Science Workbench will attempt to use the following user attributes, if they exist, as the full name of the user:

- The first name of the user. Valid attributes are:
 - `givenName`
 - `urn:oid:2.5.4.42`
- The last name of the user. Valid attributes are:
 - `sn`
 - `urn:oid:2.5.4.4`

Configuration Options

Use the following properties to configure SAML authentication and authorization in Cloudera Data Science Workbench.

For an overview of the login process, see [How Login Works with SAML Group Settings Enabled](#).

Cloudera Data Science Workbench Settings

- **Entity ID:** Required. A globally unique name for Cloudera Data Science Workbench as a Service Provider. This is typically the URI.
- **NameID Format:** Optional. The name identifier format for both Cloudera Data Science Workbench and Identity Provider to communicate with each other regarding a user. Default: `urn:oasis:names:tc:SAML:1.1:nameid-format:emailAddress`.
- **Authentication Context:** Optional. [SAML authentication context](#) classes are URIs that specify authentication methods used in SAML authentication requests and authentication statements. Default: `urn:oasis:names:tc:SAML:2.0:ac:classes:PasswordProtectedTransport`.

Signing SAML Authentication Requests

- CDSW Private Key for Signing Authentication Requests: Optional. If you upload a private key, you must upload a corresponding certificate as well so that the Identity Provider can use the certificate to verify the authentication requests sent by Cloudera Data Science Workbench. You can upload the private key used for both signing authentication requests sent to Identity Provider and decrypting assertions received from the Identity Provider.
- CDSW Certificate for Signature Validation: Required if the Cloudera Data Science Workbench Private Key is set, otherwise optional. You can upload a certificate in the [PEM format](#) for the Identity Provider to [verify the authenticity](#) of the authentication requests generated by Cloudera Data Science Workbench. The uploaded certificate is made available at the <http://cdsw.company.com/api/v1/saml/metadata> endpoint.

SAML Assertion Decryption

Cloudera Data Science Workbench uses the following properties to support SAML assertion encryption & decryption.

- CDSW Certificate for Encrypting SAML Assertions - Must be configured on the Identity Provider so that Identity Provider can use it for encrypting SAML assertions for Cloudera Data Science Workbench
- CDSW Private Key for Decrypting SAML Assertions - Used to decrypt the encrypted SAML assertions.

Identity Provider

- Identity Provider SSO URL: Required. The entry point of the Identity Provider in the form of URI.
- Identity Provider Logout URL: Optional. When this URL is provided, and the Enable SAML Logout checkbox is enabled, a user clicking the Sign Out button on CDSW will also be logged out of the identity provider.
- Identity Provider Signing Certificate: Optional. Administrators can upload the [X.509](#) certificate of the Identity Provider for Cloudera Data Science Workbench to validate the incoming SAML responses.

Cloudera Data Science Workbench extracts the Identity Provider SSO URL and Identity Provider Signing Certificate information from the uploaded Identity Provider Metadata file. Cloudera Data Science Workbench also expects all Identity Provider metadata to be defined in a `<md:EntityDescriptor>` XML element with the namespace "urn:oasis:names:tc:SAML:2.0:metadata", as defined in the [SAMLMeta-xsd schema](#).

For on-premises deployments, you must provide a certificate and private key, generated and signed with your trusted Certificate Authority, for Cloudera Data Science Workbench to establish secure communication with the Identity Provider.

- Enable SAML Logout: Optional. When this checkbox is enabled, and the Identity Provider Logout URL is provided, a user clicking the Sign Out button on CDSW will also be logged out of the identity provider. As a result of this, the user might also be logged out from any other services that they authenticate to using the same identity provider. For this feature to work, the identity provider must support Single Logout Service with HTTP-Redirect binding.

Authorization

When you're using SAML 2.0 authentication, you can use the following properties to restrict the access to Cloudera Data Science Workbench to certain groups of users:

- SAML Attribute Identifier for User Role: The Object Identifier (OID) of the user attribute that will be provided by your identity provider for identifying a user's role/affiliation. You can use this field in combination with the following SAML User Groups property to restrict access to Cloudera Data Science Workbench to only members of certain groups.

For example, if your identity provider returns the `OrganizationalUnitName` user attribute, you would specify the OID of the `OrganizationalUnitName`, which is `urn:oid:2.5.4.11`, as the value for this property.

- **SAML User Groups:** A list of groups whose users have access to Cloudera Data Science Workbench. When this property is set, only users that are successfully authenticated AND are affiliated to at least one of the groups listed here, will be able to access Cloudera Data Science Workbench.

For example, if your identity provider returns the `OrganizationalUnitName` user attribute, add the value of this attribute to the SAML User Groups list to restrict access to Cloudera Data Science Workbench to that group.

If this property is left empty, all users that can successfully authenticate themselves will be able to access Cloudera Data Science Workbench.

- **SAML Full Administrator Groups:** A list of groups whose users are automatically granted the site administrator role on Cloudera Data Science Workbench.

The groups listed under SAML Full Administrator Groups do need to be listed again under the SAML User Groups property.

How Login Works with SAML Group Settings Enabled

With SAML Group settings enabled, the login process in Cloudera Data Science Workbench works as follows.

1. Authentication by Identity Provider

When an unauthenticated user accesses Cloudera Data Science Workbench, they are first sent to the identity provider's login page, where the user can login as usual.

Once successfully authenticated by the identity provider, the user is sent back to Cloudera Data Science Workbench along with a SAML assertion that includes, amongst other things, a list of the user's attributes.

2. Authorization Check for Access to Cloudera Data Science Workbench

Cloudera Data Science Workbench will attempt to look up the value of the SAML Attribute Identifier for User Role in the SAML assertion and check to see whether that value, which could be one or more group names, exists in the SAML User Groups and SAML Full Administrator Groups whitelists.

If there is a match with a group listed under SAML User Groups, this user will be allowed to access Cloudera Data Science Workbench as a regular user.

If there is a match with a group listed under SAML Full Administrator Groups, this user will be allowed to access Cloudera Data Science Workbench as a site administrator.



Note: Ensure that your Identity Provider is configured to return the SAML groups as an XML list, with one group per list element, and not as a single comma separated string.

Debug Login URL

When using external authentication, such as LDAP, Active Directory or SAML 2.0, even a small mistake in authentication configurations in either Cloudera Data Science Workbench or the Identity Provider could potentially block all users from logging in.

Cloudera Data Science Workbench provides an optional fallback debug login URL for site administrators to log in against the local database with their username/password created during the sign up process before changing the external authentication method. The debug login URL is `http://cdsw.company.com/login?debug=1`. If you do not remember the original password, you can reset it by going directly to `http://cdsw.company.com/forgot-password`. When configured to use external authentication, the link to the forgot password page is disabled on the login page for security reasons.

Disabling the Debug Login Route

Optionally, the debug login route can be disabled to prevent users from accessing Cloudera Data Science Workbench via local database when using external authentication. In case of external authentication failures, when the debug login route is disabled, root access to the master host is required to re-enable the debug login route.

Contact Cloudera Support for more information.