Cloudera AI

Securing Cloudera AI

Date published: 2020-07-16
Date modified: 2025-10-31

CLOUDERA

Legal Notice

Contents

Configuring external authentication with LDAP and SAML

⚠ **Important:** Cloudera recommends you leverage Single Sign-On for users via the Cloudera Management Console. For instructions on how to configure this, see Configuring User Access to Cloudera AI. If you cannot do this, we recommend contacting Cloudera Support before attempting to use the LDAP or SAML instructions provided in this section.

Cloudera AI 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 AI performs user authentication against its internal local database. This topic describes the signup 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.

Configuring LDAP/Active Directory authentication

This topic describes how to set up LDAP authentication for a workbench.

⚠ **Important:** This is not the recommended method to set up LDAP authentication. Cloudera recommends you use the Cloudera Management Console to set this up: Configuring User Access to Cloudera AI.

Cloudera AI 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 ldapsearch 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 AI 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

Lists the general settings required to configure LDAP authentication.

- LDAP Server URI: Required. The URI of the LDAP/Active Directory server against which Cloudera AI shall 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 AI uses the search bind mechanism and two configurations, LDAP Bind DN and LDAP Bind Password, are required to perform the ldapsearch against the LDAP server.
- LDAP Bind DN: Required when using search bind. The DN to bind to for performing ldapsearch. 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 AI to perform the bind operation and extract the username from the response. Common values are uid, sAMAccountName, or userPrincipalName.

General Settings

LDAP Server URI *

ldaps://ldap.company.com

☐ **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 *

CN=Test1 Person,OU=People,DC=company,DC=com

Update LDAP Bind Password

LDAP User Search Base *

OU=People,DC=company,DC=com

LDAP User Search Filter

objectClass=person

LDAP User Username Attribute *

sAMAccountName

When you select Use Direct Bind, Cloudera AI 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 AI 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 AI then validates the user-provided password against the DN found as a result of the search.

LDAP group settings

In addition to the general LDAP settings, you can use group settings to restrict the access to Cloudera AI to certain groups in LDAP.

- LDAP Group Search Base: The base distinguished name (DN) where Cloudera AI will search for groups.
- LDAP Group Search Filter: The LDAP filter that Cloudera AI 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 AI will automatically substitute the {0} placeholder for the DN of the authenticated user.

- LDAP User Groups: A list of LDAP groups whose users have access to Cloudera AI. 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 AI.

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

- LDAP Full Administrator Groups: A list of LDAP groups whose users are automatically granted the site administrator role on Cloudera AI.

  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 AI to members of a group whose DN is:

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

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

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

  Add the CNs of both groups to the following settings in Cloudera AI:
  - LDAP User Groups: MLUsers
  - LDAP Full Administrator Groups: MLAdmins

Test LDAP Configuration

Use the **Test LDAP Configuration** form to test your settings.

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

This topic describes how to set up SAML for Single Sign-on authentication for a workbench.

⚠️ **Important:** This is not the recommended method to set up SSO. Cloudera recommends you use the Cloudera Management Console to set this up: Configuring User Access to Cloudera AI.

Cloudera AI 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.

📝 **Note:** The user sync feature only works with the SAML IDP provided by the control plane. If a custom SAML IDP is provided then customer has to make sure to turn usersync off. Otherwise, there is a risk that users will be deactivated and therefore causing cron jobs scheduled by users that are deactivated to fail.

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 AI operates as a SP.

Cloudera AI 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 AI 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 AI uses HTTP Redirect Binding for authentication requests and expects to receive responses from HTTP POST Binding. Note: When visiting these pages, view the Page Source in the browser to see the full XML.

When Cloudera AI 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 AI 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

List of properties to configure SAML authentication and authorization in Cloudera AI.

Cloudera AI settings

- Entity ID: Required. A globally unique name for Cloudera AI as a Service Provider. This is typically the URI.
- NameID Format: Optional. The name identifier format for both Cloudera AI 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 AI. You can upload the private key used for both signing authentication requests sent to Identity Provider and decrypting assertions received from the Identity Provider.
- Cloudera AI Certificate for Signature Validation: Required if the Cloudera AI 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 AI. The uploaded certificate is made available at the http://cdsw.*COMPANY*.com/api/v1/saml/metadata endpoint.

SAML assertion decryption

Cloudera AI uses the following properties to support SAML assertion encryption & decryption.

- Cloudera AI 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 AI

- Cloudera AI 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 Signing Certificate: Optional. Administrators can upload the X.509 certificate of the Identity Provider for Cloudera AI to validate the incoming SAML responses.

  Cloudera AI extracts the Identity Provider SSO URL and Identity Provider Signing Certificate information from the uploaded Identity Provider Metadata file. Cloudera AI 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 AI to establish secure communication with the Identity Provider.

Authorization

When you're using SAML 2.0 authentication, you can use the following properties to restrict the access to Cloudera AI 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 AI 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 AI. 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 AI.

  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 AI to that group.

  If this property is left empty, all users that can successfully authenticate themselves will be able to access Cloudera AI.
- SAML Full Administrator Groups: A list of groups whose users are automatically granted the site administrator role on Cloudera AI.

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

Configuring HTTP Headers for Cloudera AI

This topic explains how to customize the HTTP headers that are accepted by Cloudera AI.

Required Role: Site Administrator

These properties are available under the site administrator panel at Admin Security .

Enable Cross-Origin Resource Sharing (CORS)

Most modern browsers implement the Same-Origin Policy, which restricts how a document or a script loaded from one origin can interact with a resource from another origin. When the Enable cross-origin resource sharing property is enabled on Cloudera AI, web servers will include the Access-Control-Allow-Origin: * HTTP header in their HTTP responses. This gives web applications on different domains permission to access the Cloudera AI API through browsers.

This property is disabled by default .

If this property is disabled, web applications from different domains will not be able to programmatically communicate with the Cloudera AI API through browsers.

Enable HTTP security headers

When Enable HTTP security headers is enabled, the following HTTP headers will be included in HTTP responses from servers:

- X-XSS-Protection
- X-DNS-Prefetch-Control
- X-Frame-Options
- X-Download-Options
- X-Content-Type-Options

This property is enabled by default .

Disabling this property could leave your Cloudera AI deployment vulnerable to clickjacking, cross-site scripting (XSS), or any other injection attacks.

Enable HTTP Strict Transport Security (HSTS)

Note: Without TLS/SSL enabled, configuring this property will have no effect on your browser.

When both TLS/SSL and this property (Enable HTTP Strict Transport Security (HSTS)) are enabled, Cloudera AI will inform your browser that it should never load the site using HTTP. Additionally, all attempts to access Cloudera AI using HTTP will automatically be converted to HTTPS.

This property is disabled by default .

If you ever need to downgrade to back to HTTP, use the following sequence of steps: First, deactivate this checkbox to disable HSTS and restart Cloudera AI. Then, load the Cloudera AI web application in each browser to clear the respective browser's HSTS setting. Finally, disable TLS/SSL across the cluster. Following this sequence should help avoid a situation where users get locked out of their accounts due to browser caching.

Enable HTTP security headers

When Enable HTTP security headers is enabled, the following HTTP headers will be included in HTTP responses from servers:

- X-XSS-Protection
- X-DNS-Prefetch-Control
- X-Frame-Options
- X-Download-Options
- X-Content-Type-Options

This property is enabled by default .

Disabling this property could leave your Cloudera AI deployment vulnerable to clickjacking, cross-site scripting (XSS), or any other injection attacks.

Enable HTTP Strict Transport Security (HSTS)

Note: Without TLS/SSL enabled, configuring this property will have no effect on your browser.

When both TLS/SSL and this property (Enable HTTP Strict Transport Security (HSTS)) are enabled, Cloudera AI will inform your browser that it should never load the site using HTTP. Additionally, all attempts to access Cloudera AI using HTTP will automatically be converted to HTTPS.

This property is disabled by default .

If you ever need to downgrade to back to HTTP, use the following sequence of steps: First, deactivate this checkbox to disable HSTS and restart Cloudera AI. Then, load the Cloudera AI web application in each browser to clear the respective browser's HSTS setting. Finally, disable TLS/SSL across the cluster. Following this sequence shall help tto avoid a situation where users get locked out of their accounts due to browser caching.

Enable Cross-Origin Resource Sharing (CORS)

Most modern browsers implement the Same-Origin Policy, which restricts how a document or a script loaded from one origin can interact with a resource from another origin. When the Enable cross-origin resource sharing property is enabled on Cloudera AI, web servers will include the Access-Control-Allow-Origin: * HTTP header in their HTTP responses. This gives web applications on different domains permission to access the Cloudera AI API through browsers.

This property is disabled by default .

If this property is disabled, web applications from different domains will not be able to programmatically communicate with the Cloudera AI API through browsers.

SSH Keys

This topic describes the different types of SSH keys used by Cloudera AI, and how you can use those keys to authenticate to an external service such as GitHub.

Personal key

Cloudera AI automatically generates an SSH key pair for your user account. You can rotate the key pair and view your public key on your user settings page. It is not possible for anyone to view your private key.

Every console you run has your account's private key loaded into its SSH-agent. Your consoles can use the private key to authenticate to external services, such as GitHub. For instructions, see *Adding an SSH Key to GitHub*.

Related Information
Adding an SSH Key to GitHub

Team key

Team SSH keys provide a useful way to give an entire team access to external resources such as databases or GitHub repositories (as described in the next section).

Like Cloudera AI users, each Cloudera AI team has an associated SSH key. You can access the public key from the team's account settings. Click Account, then select the team from the drop-down menu at the upper right corner of the page.

When you launch a console in a project owned by a team, you can use that team's SSH key from within the console.

Adding an SSH key to GitHub

Cloudera AI creates a public SSH key for each account. You can add this SSH public key to your GitHub account if you want to use password-protected GitHub repositories to create new projects or collaborate on projects.

Procedure

1. Sign in to Cloudera AI.
2. Go to the upper right drop-down menu and switch context to the account whose key you want to add. This could be a individual user account or a team account.
3. On the left sidebar, click User Settings.
4. Go to the Outbound SSH tab and copy the User Public SSH Key.
5. Sign in to your GitHub account and add the Cloudera AI key copied in the previous step to your GitHub account. For instructions, refer the GitHub documentation on Adding a new SSH key to your GitHub account.

Creating an SSH tunnel

You can use your SSH key to connect Cloudera AI to an external database or cluster by creating an SSH tunnel.

About this task

In some environments, external databases and data sources reside behind restrictive firewalls. A common pattern is to provide access to these services using a bastion host with only the SSH port open. Cloudera AI provides a convenient way to connect to such resources using an SSH tunnel.

If you create an SSH tunnel to an external server in one of your projects, then all engines that you run in that project are able to connect securely to a port on that server by connecting to a local port. The encrypted tunnel is completely transparent to the user and code.

Procedure

1. Open the Project Settings page.
2. Open the Tunnels tab.
3. Click New Tunnel.
4. Enter the server IP Address or DNS hostname.
5. Enter your username on the server.
6. Enter the local port that should be proxied, and to which remote port on the server.

What to do next

On the remote server, configure SSH to accept password-less logins using your individual or team SSH key. Often, you can do so by appending the SSH key to the file /home/username/.ssh/authorized_keys.

Hadoop authentication for Cloudera AI Workbenches

Cloudera AI does not assume that your Kerberos principal is always the same as your login information. Therefore, you will need to make sure Cloudera AI knows your Kerberos identity when you sign in.

About this task

This procedure is required if you want to run Spark workloads in an Cloudera AI Workbench. This is also required if connecting Cloudera Data Visualization running in Cloudera AI to an Impala instance using Kerberos for authentication.

Procedure

1. Navigate to your Cloudera AI Workbench.

2. Go to the top-right dropdown menu, click Account settings Hadoop Authentication .

3. To authenticate, either enter your password or click Upload Keytab to upload the keytab file directly.

Results

Once successfully authenticated, Cloudera AI uses your stored credentials to ensure you are secure when running workloads.

Cloudera AI and outbound network access

Cloudera AI expects access to certain external networks. See the related information *Outbound internet access and proxy* for further information.

Note: The outbound network access destinations listed in *Outbound internet access and proxy* are only the minimal set required for Cloudera installation and operation. For environments with limited outbound internet access due to using a firewall or proxy, access to Python or R package repositories such as Python Package Index or CRAN may need to be whitelisted if your use cases require installing packages from those repositories. Alternatively, you may consider creating mirrors of those repositories within your environment.

Related Information
Outbound internet access and proxy

Non-transparent proxy and egress trusted list

Cloudera AI, when used on AWS on cloud, supports non-transparent proxies. Non-transparent proxy enables Cloudera AI to proxy web requests without requiring any particular browser setup.

Egress Trusted List

In normal operation, Cloudera AI requires the ability to reach several external domains. See Outbound internet access and proxy for more information.

Related Information
Outbound internet access and proxy