Machine Learning

Securing Cloudera Machine Learning

Date published: 2020-07-16
Date modified: 2024-05-30

https://docs.cloudera.com/

Legal Notice

Contents

Configuring HTTP Headers for Cloudera Machine Learning..... 4
Enable HTTP Security Headers..... 5
Enable HTTP Strict Transport Security (HSTS)..... 5
Enable Cross-Origin Resource Sharing (CORS)..... 5

SSH Keys..... 5
Personal Key..... 6
Team Key..... 6
Adding an SSH Key to GitHub..... 6
Creating an SSH Tunnel..... 6

Hadoop Authentication for ML Workspaces..... 7

CML and outbound network access..... 7

Configuring HTTP Headers for Cloudera Machine Learning

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

Required Role: Site Administrator

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

Important: Any changes to the following properties require a full restart of Cloudera Machine Learning. To do so, run cdsw restart on the master host.

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 Machine Learning, 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 Machine Learning 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 Machine Learning 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 Machine Learning 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 Machine Learning will inform your browser that it should never load the site using HTTP. Additionally, all attempts to access Cloudera Machine Learning 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 Machine Learning. Then, load the Cloudera Machine Learning 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 Machine Learning 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 Machine Learning will inform your browser that it should never load the site using HTTP. Additionally, all attempts to access Cloudera Machine Learning 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 Machine Learning. Then, load the Cloudera Machine Learning 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 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 Machine Learning, 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 Machine Learning 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 Machine Learning API through browsers.

SSH Keys

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

Personal Key

Cloudera Machine Learning 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 #unique_9.

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 Machine Learning users, each Cloudera Machine Learning 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 Machine Learning 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 Machine Learning.
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 Machine Learning 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 Machine Learning 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 Machine Learning 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 ML Workspaces

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

About this task

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

Procedure

1. Navigate to your ML workspace.
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 Machine Learning uses your stored credentials to ensure you are secure when running workloads.

CML and outbound network access

Cloudera Machine Learning expects access to certain external networks. See the related information *Configuring proxy hosts for CML workspace connections* for further information.

Note: The outbound network access destinations listed in *Configuring proxy hosts for CML workspace connections* are only the minimal set required for CDP 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
Configuring proxy hosts for CML workspace connections