Cloudera Data Science Workbench

Proxy Configuration

Date published: 2020-02-28
Date modified: 2021-11-30

https://docs.cloudera.com/

Legal Notice

Contents

Configuring Cloudera Data Science Workbench Deployments Behind a Proxy

If your deployment is behind an HTTP or HTTPS proxy, you must configure the hostname of the proxy you are using in Cloudera Data Science Workbench.

Procedure

1. Configure the hostname of the proxy you are using in Cloudera Data Science Workbench as follows:

   ```
   HTTP_PROXY="<http://proxy_host>:<proxy-port>"
   HTTPS_PROXY="<http://proxy_host>:<proxy_port>"
   ```

2. Depending on your deployment, use one of the following methods to configure the proxy in Cloudera Science Workbench:
   - CSD - Set the HTTP Proxy or HTTPS Proxy properties in the Cloudera Manager's CDSW service.
   - RPM - Set the HTTP_PROXY or HTTPS_PROXY properties in /etc/cdsw/config/cdsw.conf on all Cloudera Data Science Workbench gateway hosts.
3. If you are using an intermediate proxy such as Cntlm to handle NTLM authentication, add the Cntlm proxy address to these fields.

   ```
   HTTP_PROXY="http://localhost:3128"
   HTTPS_PROXY="http://localhost:3128"
   ```

Supporing a TLS-Enabled Proxy Server

If the proxy server uses TLS encryption to handle connection requests, you will need to add the proxy's root CA certificate to your host's store of trusted certificates. This is because proxy servers typically sign their server certificate with their own root certificate. Therefore, any connection attempts will fail until the Cloudera Data Science Workbench host trusts the proxy's root CA certificate. If you do not have access to your proxy's root certificate, contact your Network / IT administrator.

About this task

To enable trust, perform the following steps on the master and worker hosts.

Procedure

1. Copy the proxy's root certificate to the trusted CA certificate store (ca-trust) on the Cloudera Data Science Workbench host.

   ```
   cp /tmp/<proxy-root-certificate>.crt /etc/pki/ca-trust/source/anchors/
   ```

2. Use the following command to rebuild the trusted certificate store.

   ```
   update-ca-trust extract
   ```

3. If you will be using custom engine images that will be pulled from a Docker repository, add the proxy's root certificates to a directory under /etc/docker/certs.d.

   For example, if your Docker repository is at docker.repository.mycompany.com, create the following directory structure:

   ```
   /etc/docker/certs.d
   ```

```
|-- docker.repository.mycompany.com          # Directory named after Doc
ker repository
    |-- <proxy-root-certificate>.crt         # Docker-related root CA ce
rtificates
```

This step is not required with the standard engine images because they are included in the Cloudera Data Science Workbench RPM.

4. Re-initialize Cloudera Data Science Workbench to have this change go into effect.

```
cdsw start
```

Configuring Hostnames to be Skipped from the Proxy

Starting with version 1.4, if you have defined a proxy in the HTTP_PROXY(S) or ALL_PROXY properties, Cloudera Data Science Workbench automatically appends a list of IP addresses to the NO_PROXY configuration.

About this task

Starting with version 1.4, if you have defined a proxy in the HTTP_PROXY(S) or ALL_PROXY properties, Cloudera Data Science Workbench automatically appends the following list of IP addresses to the NO_PROXY configuration. Note that this is the minimum required configuration for this field.

```
"127.0.0.1,localhost,100.66.0.1,100.66.0.2,100.66.0.3,
100.66.0.4,100.66.0.5,100.66.0.6,100.66.0.7,100.66.0.8,100.66.0.9,
100.66.0.10,100.66.0.11,100.66.0.12,100.66.0.13,100.66.0.14,
100.66.0.15,100.66.0.16,100.66.0.17,100.66.0.18,100.66.0.19,
100.66.0.20,100.66.0.21,100.66.0.22,100.66.0.23,100.66.0.24,
100.66.0.25,100.66.0.26,100.66.0.27,100.66.0.28,100.66.0.29,
100.66.0.30,100.66.0.31,100.66.0.32,100.66.0.33,100.66.0.34,
100.66.0.35,100.66.0.36,100.66.0.37,100.66.0.38,100.66.0.39,
100.66.0.40,100.66.0.41,100.66.0.42,100.66.0.43,100.66.0.44,
100.66.0.45,100.66.0.46,100.66.0.47,100.66.0.48,100.66.0.49,
100.66.0.50,100.77.0.10,100.77.0.128,100.77.0.129,100.77.0.130,
100.77.0.131,100.77.0.132,100.77.0.133,100.77.0.134,100.77.0.135,
100.77.0.136,100.77.0.137,100.77.0.138,100.77.0.139"
```

This list includes 127.0.0.1, localhost, and any private Docker registries and HTTP services inside the firewall that Cloudera Data Science Workbench users might want to access from the engines.

Procedure

To configure any additional hostnames that should be skipped from the proxy, use one of the following methods depending on your deployment:

- On a CSD deployment, use the Cloudera Manager CDSW service's No Proxy property to specify a comma-separated list of hostnames.
- On an RPM deployment, configure the NO_PROXY field in cdsw.conf on all Cloudera Data Science Workbench hosts.